Soybean Genomics Research Program Accomplishments Report

Meeting Strategic Milestones for 2008 to 2012 in a Timely Manner

v 1.6 August 2010

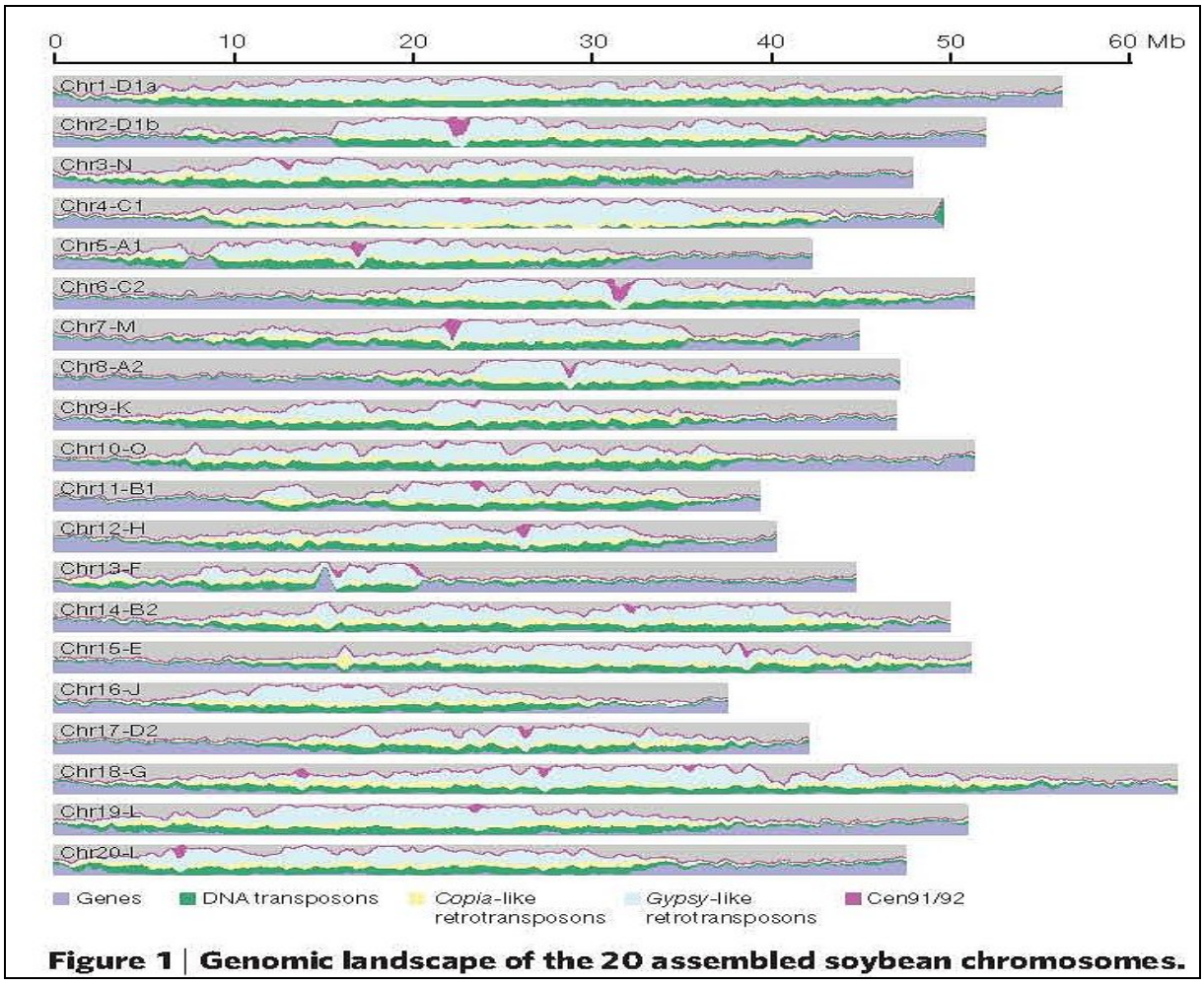

Figure 1 | Genomic landscape of the 20 assembled soybean chromosomes.

Schmutz, J., et.al. 2010. Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Table of Contents

Executive Summary

The soybean research community has engaged a transparent process for developing and implementing a strategic framework for a national program to unlock the secrets of the soybean genome. The success of this process is evidenced by achieving the soybean genome sequence in record time, compared to similar efforts in other major crops, and many useful tools to expedite elite variety development. This document provides an accounting of major accomplishments in relation to performance measures in the *Soybean Genome Research Strategic Plan, 2008-2012*. Many additional accomplishments and publications may be included in future versions of this report. However, the current intent is to provide an initial basis for assessing the progress made in this area of research, and to help scientists and stakeholders move soybean research forward with a new framework that is relevant to the needs of the soybean industry.

Strategic Plan Objectives 2008 - 2012	Major Milestone Accomplishment Timeline (2008 2009 2010 2011 2012)
A.1 – Genome Informatics	
A.1.a - Annotation Needs.	Integration of genomic sequence w/ phys & genetic maps
A.1.b - Merge Maps (Genetic, Physical and Sequence)	Concensus soybean linkage map 4.0
A.1.c - Integrate Genomic Data with Other Species	Syntenic relations across G max, Medicago, Arabidopsis
A.1.d - Genomic Database w/ Convenient Access	Soybase expanded to facilate genomic data retrival
A.1.e - Transposon and Repeat Sequence Databases	Expert curated transposon database
A.2 – Genome Finishing	
A.2.a - Initial Genome Assembly.	Size 975Mb, 66,153 protein-coding loci, 32317 EST genes, ...
A.2.b - Initial Annotation of Genome Sequence.	See A.1.d
A.2.c - Selective Re-Sequencing.	High thru-put SNP discovery by deep resequencing
A.4 – Genome Re-sequencing	
A.4.a - SNP Genotyping.	1536 Universal Soy Linkage Panel for QTL mapping
A.4.b - Resequencing for SNP Discovery.	High thru-put genotyping w/ GoldenGate assay
A.4.c - Simultaneous SNP Discovery and Genotyping	210,990 SSRs in BARCSOYSSR dbase
A.5 – Phaseoloid Genomics	
A.5.a - Genomic Research Goals for *Phaseolus*	High thru-put SNP discovery in common bean
A.5.b - Other Genera	Monosomic alien addition lines, interspecific crosses
B.1 – Gene Function Annotation/Informatics	
B.1.a - Gene Prediction and Confirmation	Identification of candidate genes for low stachyose trait
B.1.b - Resources Needed for Genome Annotation	Dbase for soybean transcription factors
B.1.c - Integration with Other Species	See B.1.a
B.2 – Discovery via Mutagenesis	
B.2.a - Reverse Genetics TILLING.	4 independent TILLING populations for mutation discovery
B.2.b - Fast-Neutron Mutagenesis	Mutant populations by fast-neutron or transposon tagging
B.2.c - Gene Function Studies w/Transposon Tagging	Established soybean transposon-based mutagenisis repository
B.3 – Functional Genomics Approaches	
B.3.a - Transcriptomic Approaches	High thru-put expression data & sRNA for disease resistance
B.3.b - Proteomic Approaches	Established protein reference map for soybean root hair cells
B.3.c - Metabolomic Approaches	Global metabolome of soy root hairs during symbiosis
B.5 – Breeder Perspectives on Gene Function	
B.5.a - Develop NIL pairs for QTL microarray analysis	Mappling populations for Aphid, BPMV, Cercosproa resistance
B.5.b - Random-populations for Bulk Seg Analysis	WO/2002/063020 Monsanto
C.1 – Association Mapping	Characterization of linkage disequilibruim in soybean
C.2 – Track Breeding-Induced Genomic Change	Bulked segregant analysis w/Golden Gate for Rpp3 loci
C.3 – Mining Yield QTLs in Exotic Germplasm	
C.3.a. - Analysis of 4 Yield QTL from Exotic Germplasm	Mapping populations w/exotic germplasm for yield QTL discovery
C.3.b. - Identify Yield QTL Alleles from Exotic Germplasm	Registration of LG04-6863 with diverse pedigree
C.3.c. - Evaluate Cultivars for Favorable Yield Genes	Nested Association Mapping for Yield QTL in diverse elite cvs
C.4 – Marker-Assisted Selection Resources	Soybase w/genetic, physical & genome maps for trait QTL
C.5 – Germplasm Genomics Informatics	
C.5.a - Development of HapMap Browser	GEYSIR, for SNP & chromosome scanning
C.5.b - Correlation of expression data with QTL (eQTL)	FastMap for SNP mapping & haplotype association
C.5.c - Tools for identifying candidate genes in QTL	VIGS applications for Rpp4 candidate gene identification
C.5.d - User interface for informatics tools	SSWP for bioinformatic resource integration
C.5.e - Large datasets	LIS for navigation of transcript data across plant genomes
C.5.f - Long-term curation of soybean genome sequence	See A.1.a
D.1 – Create a Transgenic Event Repository	Little progress
D.2 – Center for Transgenics/Transformation	NCBI dbase resources
D.3 – Soybean Regulatory Promoter Set	Soybean Upstream Regulatory Element dbase (SURE)
D.4 – Improve Soybean Transformation Efficiency	Satisfactory efficiences except for T-DNA mutagenesis

Introduction

The advent of a chromosome-scale draft assembly of the soybean (*Glycine max* L. Merr.) genome did not spring up overnight like some magic beanstalk. Rather, it is an outcome of a dynamic, technology driven, and timely strategic process whose origin may be traced formally to the *Soybean Genomics White Paper* [Boerma, H.R., D. Buxton, M. Kelly, K. Van Amburg, Soybean genomics white paper January 2000, http://129.186.26.94/Genomics/Soybean_Genomics.html (2000)]. That January 2000 document was a product of an October 21-22, 1999 meeting of seventeen experts in plant genomics, DNA markers, plant transformation, and bioinformatics. The workshop was planned by: Dr. Dwayne Buxton, National Program Leader for Oilseeds & Biosciences, USDA Agricultural Research Service; Dr. Roger Boerma, Research Professor and Coordinator of the Center for Soybean Improvement, University of Georgia; Maureen Kelly of AgSource, Inc., a subcontractor with the United Soybean Board focusing on Federal Research Coordination; and Kent Van Amburg, Production Committee Manager, United Soybean Board. Elizabeth Vasquez of MCA Consulting facilitated the workshop. A consensus was reached on research priorities in the area of soybean genomics. Milestones included: 1) doubling Simple Sequence Repeat (SSR) markers to 2000 within 3 years; 2) expansion of Single Nucleotide Polymorphism (SNP) markers to 10,000 within three to five years; 3) improving the efficiency of soybean transformation by five- to ten-fold in three years; 4) tagging 80% of the genes in the soybean genome within three to five years; 5) integration of genetic, physical and transcript maps of soybeans within three to five years; and 6) employing comparative genomics to define the structure and attributes of the soybean genome.

However, rising scientific enthusiasm for genomic investigations among living organisms also created an extraordinarily competitive environment for appropriations to finance these rather expensive ventures. Therefore it became mutually beneficial to establish research coalitions to improve the efficiency of genomic investigations among related species. For this reason, the U.S. Legume Crops Genomics Initiative (LCGI) was organized under the auspices of the American Soybean Association, United Soybean Board, National Peanut Foundation, USA Dry Pea and Lentil Council, the National Dry Bean Council, and the Alfalfa Council to facilitate communication and cooperation among scientists with an interest in genomic research on soybeans, peanuts, pea & lentils, common bean, alfalfa and model-legume crops. LCGI was founded on the premise that the development of an integrated legume genomics research system would enhance ability to leverage information across legume crops and model species. The first U.S. Legume Crops Genomics Workshop was convened on July 30-31, 2001 at Hunt Valley, Maryland by: H. Roger Boerma, University of Georgia; Judy St. John, Associate Deputy Administrator, USDA Agricultural Research Service; Jennifer Yezak Molen, AgSource, Inc., and was hosted by the USB, the National Peanut Foundation, the USA Dry Pea and Lentil Council, and the USDA-ARS. Twenty-six legume scientists skilled in relevant arts developed a white paper [Boerma, H.R., J. St. John, and J. Yezak Molen, U.S. legume crops genomics workshop white paper, http://www.legumes.org/ (2001)] that outlined high-priority research in the areas of: 1) genome sequencing of strategic legume species; 2) physical map development and refinement; 3) functional analysis: transcriptional and genetic; 4) development of DNA markers for comparative mapping and breeding; 5) characterization and utilization of legume biodiversity; and 6) development of a legume data resource. The nature of this cooperative interaction not only ensured timely research progress in all legume crops associated with the Initiative, but also enhanced the competitive position of the LCGI within the framework of the National Plant Genome Initiative, which is coordinated by the Interagency Working Group on Plant Genomics, Committee on Science, National Science & Technology Council.

Implementation of a coordinated effort for research & development of genomics across the legume family facilitated progress in the model species *Medicago truncatula* and *Lotus*

japonicus and in soybean (*Glycine max*); and accentuated the need to transfer genomic information from the model species to cool-season pulses [pea (*Pisum sativum*), lentil (*Lens culinaris*), chickpea (*Cicer arietinum*), field bean (*Vicia faba*)], and warm-season food legumes [peanut (*Arachis hypogea*), common bean (*Phaseolus vulgaris*)], and forage legumes [alfalfa (*Medicago sativa*), clover (*Trifolium* spp.)]. This mission was codified further by 1) a third white paper, entitled: *Legumes as a Model Plant Family: Genomics for Food and Feed*; and 2) by the publication of the monograph, *Legume Crop Genomics.*

The white paper was an outcome of the CATG (Cross-legume Advances Through Genomics) conference on December 14-15, 2004 in Sante Fe, NM (http://catg.ucdavis.edu) which was organized by the LCGI steering committee: Charlie Brummer (alfalfa), Paul Gepts (beans; chair), Randy Shoemaker (soybean), Tom Stalker (peanut), Norm Weeden (cool-season legumes) and Nevin Young (model legumes). In addition, Bill Beavis served as a bioinformatics resource and funding was provided by the National Science Foundation (Plant Genome Research Program) and the USDA (National Research Initiative). About 50 individuals in attendance represented the respective legume communities as well as various funding agencies. The objectives of the conference were to: (1) identify a unifying goal for an international cross-legume genome project; (2) identify cross-cutting themes to help integrate the different legume crop genomics programs, including a unified legume genomics information system, nutritional and health-related aspects of legumes, and detailed synteny and comparative genomics of legumes; and (3) outline specific components and milestones for the initiative. These deliberations identified four tiers of legume species, each with specific genomic resources to be developed. Based on phylogenetic arguments, and particularly the degree of synteny, two major foci of legumes were identified, the hologaleginoid clade or cool-season legumes and the phaseoloid/millettioid clade or warm-season legumes. In each of these two foci, one or two reference species were identified, *M. truncatula* and *L. japonicus* in the former and soybean in the latter. Development of a full range of genomics resources, including sequencing of the entire genome, was the highest priority for these reference species. For a second group, common bean and peanut, a broad range of genomic resources were recommended, including a physical map, BAC-end sequencing and marker development, anchoring of the genetic and physical map, ESTs of the major organs, chip resources, and sequencing of gene rich regions. A third group consisted of all other legume crops in the two foci, including pea, lentil, chickpea, field bean, clover, cowpea (*Vigna unguiculata*), and pigeon pea (*Cajanus cajan*). For these legumes, translational genomic tools were to be developed, principally for cross-legume markers, species-specific recombinant inbred lines, genetic maps, EST and BAC libraries. A fourth group included other legumes not in the two main foci, such as members of the basal legume clades. An abbreviated version of this white paper was published. (Gepts, P., W.D. Beavis, E.C. Brummer, R.C. Shoemaker, H.T Stalker, N.F. Weeden, and N.D. Young. 2005. Legumes as a model plant family. Genomics for food and feed Report of the Cross-Legume Advances through Genomics Conference. Plant Physiol. 137:1228-1235).

Publication of *Legume Crop Genomics*. (2004. eds. E.C. Brummer, H.T. Stalker and R.F. Wilson, AOCS Press, Champaign) helped the LCGI document in a unified manner the initial research strategies, the development of genomic tools and resources, and future direction of the legume research community. In addition, to establishing a benchmark for then state-of-art technology, this volume presented technical themes in a manner that helped many readers gain a more informed opinion of plant genomics. In that regard, the chapter by Gary Stacey presented an inventory of the available resources for soybean genomics: 1) Genetic map. The soybean composite genetic map was well developed. The classical map contained 63 loci on 19 linkage groups, while the molecular map encompassed 20 linkage groups and over 2300 cM based on 600 SSRs mapped in 6 populations (http://129.186.26.94/). In addition, the composite map contained >800 RFLP

markers, >600 AFLP markers, and hundreds of RAPDs; 2) BAC libraries. A number of public-sector BAC libraries were available providing >35-fold genome coverage. For example, a 4-6 X Forrest BAC library was used to construct a physical map. Similarly, a 5 X (~130 Kbp, HindIII inserts) library existed for the cultivar Williams 82; 3) Physical map. A double-digest-based map for the cultivar 'Forrest' was constructed (http://soybeangenome.siu.edu/), on which work continued to reduce the number of contigs (presently >3,000 contigs) and to add additional DNA markers (e.g., through BAC-end sequencing); 4) EST sequences. Over 300,000 soybean EST sequences were deposited in Genbank (http://129.186.26.94/soybeanest.html); 5) Functional Genomics. Lila Vodkin's laboratory (Univ. of Illinois) used ESTs to develop DNA microarrays for functional genomic analysis. Proteomic efforts were underway to apply translated EST sequences to protein identification (e.g., Wan et al., unpublished); 6) Soybean transformation and mutagenesis. Contrary to a NRC report (http://books.nap.edu/books/0309085292/html/R1.html), soybean transformation efficiencies were consistently >5% and, in some labs, efficiencies >12% were common. These improvements in transformation efficiencies led to development of transposon tagging projects for soybean, viral-induced gene silencing systems and TILLING populations for soybean. Soybean also had over 300 phenotypic mutants; some of which were leading to marketable traits (e.g. modified oils, low phytate); 7) Phenotype analysis. Soybean biochemistry, physiology and agronomic knowledge of soybean far exceeded that of any model legume. QTL discovery for production, protection and quality traits were integrated with the genetic and physical maps. These genomic resources firmly underpinned the status of the Soybean research community and helped broker international collaboration in soybean genomics with scientists in Japan, China, Korea, and various countries of South America and Europe. In addition, forums such as the biennial Cellular and Molecular Biology of Soybean Conferences helped improve communication on an interdisciplinary level.

As each of the legume crop communities began to amass critical genomic resources, it became necessary for each community to develop priorities that were crop specific. The strategic foundation for soybean genomics research was laid on May 20-21, 2003, when nineteen researchers participated in a workshop convened by Diane Bellis (AgSource Inc.), Roger Boerma (University of Georgia), Ed Ready (United Soybean Board), Richard F. Wilson (USDA, ARS National Program Staff), and hosted by the United Soybean Board Production Committee in St. Louis, MO. The scientists reviewed the current status of soybean genomic research and reached consensus on a strategic framework for outlining research priorities and significant near-term milestones. This input was captured in the *Strategic Plan for Soybean Genomics 2003-2007*. (http://soybase.org/SoyGenStrat2005/Soy_Genome_Strat_Plan_2005.html) Coordination of plan implementation was delegated through election of the Soybean Genetic Executive Committee (R.C. Shoemaker, P.B. Cregan, W. Parrott, H.R. Boerma, J. Orf). The SoyGEC took action to move forward aggressively on:1) fingerprinting Williams 82 BACs and physical map compilation (Jan Dvorak, Gary Stacey, Henry Ngyuen, David Lightfoot, David Grant, Randy Shoemaker); 2) anchoring ESTs and Wm82 BACs on the genetic map and enhancing physical map resolution (Perry Cregan, Jim Specht, Randy Shoemaker, Scott Jackson, Brian Scheffler); 3) manual annotation of Wm82 BAC fingerprints (Gary Stacey, Henry Ngyuen); 4) gene-specific oligonucleotide resources and 3' EST sequences for microarray and bead technology (Lila Vodkin); 5) development of a TILLInG resource (Niels Nielsen, Scott Jackson, David Lightfoot); and 6) RNAi technology for specific gene mutations (David Sommers, Wang, Tom Clemente, Wayne Parrott, John Finer, Zhang, Gary Stacey, Henry Ngyuen).

The SoyGEC (http://129.186.26.94/) also initiated strategies to proactively communicate soybean research community priorities to representatives of federal granting agencies. Without

encumbering individual initiatives, SoyGEC encouraged coordination of dedicated research teams to solve soybean problems of national and international importance. A National Science Foundation-sponsored workshop was held in St. Louis on October 21, 2003 to take an inventory of the current genomic resources in soybean, identify areas where more data were needed, and to set a research a strategy to advance soybean genomics research. Special attention was focused on research opportunities provided by unique aspects of soybean biology. The workshop included academic, governmental, and industrial scientists covering a wide variety of specialties related to both basic and applied research on soybean, along with scientists from outside the soybean field who provided general expertise in genomics and represented a wealth of experience garnered from other genomic projects. Representatives from federal funding agencies and soybean commodity groups observed the meeting. This workshop extended and further defined the findings from earlier workshops, which had surveyed the status and priority goals for soybean & legume genomics (http://129.186.26.94/Genetic_Resources/Soybean_Genetic_Resources.html; and http://129.186.26.94/Legume_Initiative/LegGenomicsPaper10Oct01.html). The proceedings for the October 2003 workshop *Genomic Perspectives of Soybean Biology* were published. (Stacey, G., L. Vodkin, W.A. Parrott, and R.C. Shoemaker. 2004. National Science Foundation-Sponsored Workshop Report. Draft Plan for Soybean Genomics. *Plant Physiology* 135:59-70).

Action to address the need for bioinformatic resources was taken by the National Center for Genome Research (NCGR) and USDA, ARS, NPS when the US Congress established the Model Plant Initiative (MPI) to use bioinformatics to leverage genomic information from model plant species. The Legume Information System (LIS), a joint NCGR & ARS effort, was one of the first projects in support of the MPI to demonstrate the integration of genomic information from *Arabidopsis thaliana*, *Medicago truncatula* and *Lotus japonicus* to important agronomic legumes, e.g., soybean, alfalfa, pea, dry beans. Since its inception, LIS development has been guided by ideas and suggestions from the legume research communities. Based on user needs, USDA-ARS and NCGR developed a plan to amplify the power of this information resource for identification of candidate genes, unique genes, and evolutionary relationships among genes for crop improvement. Bill Beavis, Randy Shoemaker and Rich Wilson presented details of the *Strategic Plan for Comparative Legume Biology 2005-2010* to NCGR and ARS leadership on January 12, 2004 in San Diego CA. It was agreed to utilize LIS as a foundation for a Comparative Legume Biology (CLB) Program. CLB expanded LIS from a passive data management system to a platform for novel data analysis and visualization tools. With the emergence of high throughput biotechnologies and bioinformatics, comparative biology enabled more sophisticated analyses of genomic sequences, genomic maps, micro-arrays, protein arrays, metabolic arrays, genetic regulatory networks, biochemical and whole organism phenotypes.

The SoyGEC convened the first assessment of research performance against the *Strategic Plan for Soybean Genomics 2003-2007* on July 19-20, 2005. Approximately 50 researchers and administrators participated in this USB supported workshop in St.Louis, MO. Presentations provided updates on the current status of soybean resources and related genomics technologies. Stakeholder input was generated in facilitated discussion groups to assess the status of soybean genomics, identify needs, and identify milestones to achieve objectives. The discussion groups included the general areas of Functional Genomics A (Transcriptome and Proteome), Functional Genomic B (Reverse Genetics), Physical and Genetic Maps, and Bioinformatics. Several topics received overwhelming support in all group reports. A high quality physical map in Williams 82 and integration with the physical map of Forrest was a very high priority. A whole-genome sequence was an expectation of the research community. The need for standardization of protocols, terminologies and ontologies was evident as interactions among groups and among research communities expand. Finally, there was urgent need to establish long-term facilities that

have the capability to archive, maintain, generate and provide biological resources. Overall, there was consensus that the ongoing research was on target, if not ahead of schedule, and still relevant to the goals and objectives of the *Strategic Plan for Soybean Genomics 2003-2007*.

On January 20, 2006 the USDA and DOE announced plans to share resources and coordinate the study of plant and microbial genomics. The soybean genome was first on the list for sequencing. Dr. Ari Patrinos, DOE Associate Director for SBER, directed the Joint Genome Institute (JGI) to carry out the work with the Stanford Human Genome Center. (http://www.energy.gov/news/2979.htm)

A Genome Strategic Planning Workshop to consider research goals and priorities for the next 5-year programmatic cycle was convened by the SoyGEC in St. Louis, MO on 30-31 May 2007. This workshop was attended by 48 experts in diverse disciplines. Stakeholders included the USB, the North Central Soybean Research Program, and one soybean producer. Whereas previous soybean strategic plans had dealt with preparation for the eventual sequencing of the genome, announcement that the soybean genome would be sequenced by DOE-JGI necessitated a reassessment of strategic objectives for the next research plan. Indeed, Jeremy Schmutz, Stanford, the leader of the DOE-JGI soybean sequence assembly effort reported that the work was proceeding exceptionally well, despite the ancient polyploidy of a now well-diploidized soybean. A 4X shotgun genome sequence of the genome had been developed and a draft assembly had been created. This assembly was being evaluated to determine the optimal means for obtaining the final goal of an 8X coverage. The 8X sequence was slated to be completed by the end of 2007, with a final assembly expected to be completed in mid-2008. Given this information, workshop participants adjourned to breakout sessions on: Soybean Genome Sequencing, Soybean Gene Function, or Soybean Germplasm Genomics. Group reports revealed that several objectives identified in the prior plan, such as delivery of large DNA constructs (BAC-sized) for genetic transformation, or a centralized tilling facility, were not deemed as high of a priority. Moreover, some 2005 Plan goals have not met milestones. This included the generation of large numbers of independent Tnt1 and Ds insertions for functional analysis of genes in soybean. Other 2005 Plan targets included methodical characterization of abiotic and biotic stresses using various expression, proteomic and metabolomic approaches. These approaches had apparently not achieved the desired momentum. (likely due to limited funding). In addition, a gratifying number of objectives and milestones identified in the 2005 Plan had been achieved on schedule. The number of SNPs and STSs proposed for discovery and development was exceeded. Inbred mapping resources (RILs) were developed. Transformation technologies have improved and various gene knock-out systems were working. USB and NSF funding enabled work on physical and transcript maps. Bioinformatic resources and staffing had more than doubled, in time to receive the whole genome shotgun sequence. These technological developments and the diminishing cost of many genomic-based technologies were taken into full account in the development of the *Strategic Plan for Soybean Genomics 2008-2012*.

The SoyGEC reached out to colleagues abroad at the first meeting of the International Soybean Genome Consortium at he NIAS in Tsukuba, Japan on April 20, 2007. This ongoing association has helped leverage resources for soybean genomic research in Japan, Korea, China and the US.

In February 2010, the SoyGEC and USB asked David Grant and Rich Wilson to prepare a report of program performance against objectives of the *Strategic Plan for Soybean Genomics 2008-2012*. The subject report, *Soybean Genomics Research Program Accomplishments Report: Meeting Strategic Milestones for 2008 to 2012 in a Timely Manner, v1.2 May 2010* documents accomplishments that are relevant to each performance measure of the plan. This information is intended to facilitate an assessment of program performance and help guide the SoyGEC and their designees in developing the next 5-year strategic plan for soybean genomic research.

A – Genome Sequence

A.1 – Genome Informatics

There is need to establish an integrated database, informatics tools for end users, and cyber-infrastructure resources to assist in: i) annotating genes and genomes, ii) merging maps, and iii) integrating the various soybean genomic resources along with those of other plant species, and iv) long-term support for genome databases and informatics.

- Version 1 of the genomic sequence of the soybean cultivar Wm82 has been completed.
- The Soybean (*Glycine max*) genome project was initiated through the DOE-JGI Community Sequencing Program (CSP) by a consortium led by Gary Stacey, Randy Shoemaker, Scott Jackson, Jeremy Schmutz, and Dan Rokhsar. Large-scale shotgun sequencing of soybean began in the middle of 2006 and was completed early in 2008. A total of ~13 million attempted Sanger shotgun reads were produced and deposited in the NCBI Trace Archive in accord with the Fort Lauderdale genome data release policy. The present assembly (Glyma1) is the first chromosome-scale assembly of the soybean genome.
- An integrated genome sequence with physical and genetics maps that facilitates the identification of candidate genes underlying QTLs and detailed plant trait phenotypic data.

  Schmutz, J., Steven B. Cannon, Jessica Schlueter, Jianxin Ma, Therese Mitros, William Nelson, David L. Hyten, Qijian Song, Jay J. Thelen, Jianlin Cheng, Dong Xu, Uffe Hellsten, Gregory D. May, Yeisoo Yu, Tetsuya Sakurai, Taishi Umezawa, Madan K. Bhattacharyya, Devinder Sandhu, Babu Valliyodan, Erika Lindquis, Myron Peto, David Grant, Shengqiang Shu, David Goodstein, Kerrie Barry, Montona Futrell-Griggs, Brian Abernathy, Jianchang Du, Zhixi Tian, Liucun Zhu, Navdeep Gill, Trupti Joshi, Marc Libault, Anand Sethuraman, Xue-Cheng Zhang, Kazuo Shinozaki, Henry T. Nguyen, Rod A. Wing, Perry Cregan, James Specht, Jane Grimwood, Dan Rokhsar, Gary Stacey, Randy C. Shoemaker and Scott A. Jackson Genome sequence of the palaeopolyploid soybean, 2010. Nature 463178-183.

A.1.a - Annotation Needs.

- The current gene set (Glyma1.0) integrates ~1.6 million ESTs with homology and *ab initio*-based gene predictions. Protein-coding genes have been given identifiers using the convention adopted by the Arabidopsis community. The identifiers are of the form Glyma%%g####, where %% is the chromosome number and #### is a numerical index that increases along each chromosome.
- The Glyma1 release produced by Jeremy Schmutz at JGI-Stanford Human Genome Center used the Arachne2 assembler in a mode tuned to the highly repetitive soybean genome. These sequence scaffolds were then integrated with soybean genetic and physical maps in collaboration with Steve Cannon (ARS, Ames IA). Comparison with the soybean EST set suggests that more than 98% of known soybean protein-coding genes are represented in the assembly. Thus, Glyma1 is largely complete with respect to "gene space." Vast tracts of repetitive sequence were also assembled into the pseudomolecules.

  SoyBase (http://soybase.org)
  Phytozome v5.0 http://www.phytozome.net)

A.1.b - Merge Maps (Genetic, Physical and Sequence).

- ARS scientists at Beltsville MD have made progress in the anchoring of the whole soybean genome sequence produced by the Department of Energy, Joint Genome Institute to the 20 chromosomes that make-up the soybean genome. The Consensus Soybean Linkage Map 4.0 contains more than 4,800 DNA sequence-based markers including 3792 SNP-containing sequence tagged sites and 1009 simple sequence repeat markers. An additional 1,240 SNP

markers were developed specifically to anchor segments of the genome that were not anchored by markers in Consensus Soybean Linkage Map 4.0. These markers were genetically mapped in a high resolution mapping population of 470 recombinant inbred lines developed from a cross of Williams 82 with the wild soybean PI 468916. As result of these efforts a total of 97% of the whole soybean genome sequence was anchored to the 20 soybean chromosomes

Hyten, D.L., I. Choi, Q. Song, J.E. Specht, T.E. Carter, R.C. Shoemaker, E. Hwang, L.K. Matukumalli, P.B. Cregan, P.B. 2010. A high density integrated genetic linkage map of soybean and the development of a 1536 Universal Soy Linkage Panel for quantitative trait locus mapping. Crop Science, in press.

- SoyBase curator David Grant and ARS scientists at Ames, IA have developed web-based displays for SoyBase that show integrated views of the genetic, physical and sequence maps of soybean.
- University of Missouri scientists are using a mapping population containing 1,025 recombinant inbred lines developed from Forrest and Williams 82 which represent US southern and northern germplasm for high resolution genetic mapping. Meanwhile, more than 30, 000 SNPs with 78% validation rate were identified by Solexa sequencing between Forrest and Williams 82. Mapping these SNPs is underway with customized Illumina SNP arrays. This map will serve as a reference map for the integration of Forrest and Williams 82 physical maps. They developed 6-dimensional BAC DNA pools and used this resource to anchor more than 2,500 genetic markers onto the physical map. In collaboration with Purdue University scientists a high quality of improved chromosome-based physical map combined with the genome sequence was constructed. This physical map will serve as the framework for ordering sequence fragments, comparative genomics, cloning genes and evolutionary analyses of legume genomes. The Stacey lab has completed the soybean karyotype and has integrated it with the sequence, genetic and physical maps.

  Wu X, Tri DV, Findley SD, Stacey G, Shannon JS, Sleper DAand Nguyen HT. (2010) Construction of a Reference Genetic Map for Soybean with a Large RIL Population. (in preparation)

  Xiaolei Wu, Guohua Zhong, Seth Findley, Perry Cregan, Gary Stacey, Henry T Nguyen (2008) Genetic marker anchoring by six-dimensional pools for development of a soybean physical map. BMC Genomics 9:28.

  Jungmin Ha, Brian Abernathy, David Grant, Xiaolei Wu, William Nelson, Gary Stacey, Henry T. Nguyen, Randy Shoemaker and Scott A. Jackson. Integration of draft sequence and physical map as framework for genomic research in Soybean (*Glycine max*). (in preparation)

  Findley SD, Cannon S, Varala K, Du J, Ma J, Hudson M, Birchler J, Stacey G. 2010. A fluorescence in situ hybridization system for karyotyping soybean. Genetics (in press)

A.1.c - Integrate Soybean Genomic Data with that of Related or Other Species.
Coordinate soybean genomic data with other species to identify and confirm orthologous genes. Generate syntenic comparisons to confirm gene predictions, models and to enable functional annotation of other non-coding sequences in: Model Species (*Arabidopsis, Medicago, Lotus*) and Poplar (*Populus trichocarpa* - Western Black Cottonwood, *Phaseolus vulgaris* - dry bean, *Vigna radiata* – mung bean, *Vigna unguicula* – cowpea.

- The genomes of most flowering plants have undergone polyploidization at some point in their evolution. How such polyploidization events have impacted the subsequent evolution of

genome structure is poorly understood. Two homoeologous regions in soybean (Glycine max), were sequenced which underwent a polyploidy event 10-14 million years ago (mya), and the orthologous/homoeologous regions in several related legumes (a second G. max genotype, G. tomentella, Phaseolus vulgaris and Medicago truncatula). The Phaseolus and Medicago genomes lack this polyploidy event, enabling us to determine the origin of differences between Glycine homologues. Comparison of the two G. max homoeologues revealed that the majority of low-copy genes (78%) have been retained in both homoeologues following polyploidization. In contrast, nucleotide binding-leucine rich repeat disease resistance gene clusters have undergone dramatic species/homoeologue-specific expansions, with some evidence for partitioning of subfamilies between homoeologues. The biggest difference between homoeologues, however, was in retroelement content, with homoeologue 2 (H2) having expanded to three times the size of homoeologue 1 (H1) due to retroelement insertions. Fluorescent in situ hybridization revealed that the H2 region was located near a centromere. The pericentromeric location correlated with a higher frequency of low copy gene loss from H2. However, low copy genes on H2 are under purifying selection and are evolving at the same rate as those on H1. One of the duplicated regions in soybean accumulated vast numbers of repetitive DNA elements known as retroelements, which can cause gene disruptions and gene silencing. Despite this accumulation of retroelements, over 75% of the low-copy duplicated genes in this region have been retained in the same order and continue to function. This finding contrasts with recent analyses of the maize genome, in which only about a third of duplicated genes appear to have been retained over a similar time period. The fate of the disease resistance genes was different, however. These have expanded in H1 but mostly been lost from H2. A proposed mechanism for this differential expansion and loss is that suppressed recombination in the pericentromere in H2 has suppressed gene "births" seen in the tandem clusters of resistance genes in H1.

Schlueter, J.A., Brian Scheffler, Scott Jackson, and Randy Shoemaker, 2008. Fractionation of synteny in a genomic region containing tandemly duplicated genes across Glycine max, Medicago truncatula and Arabidopsis thaliana. J. Hered. 99 (4): 390-395.

- The labs of Roger Innes, Nevin Young, Steven Cannon, Jeff Doyle, Valerie Geffroy, and Bruce Roe sequenced an approximately 1 Megabase region from two soybean cultivars (Williams 82 and PI96983), along with the homoeologous regions from the 13 mya duplication event and ~58 mya duplication event, and compared these regions to homologous regions in Phaseolus and Glycine tomentella. As part of these analyses, a large number of retrotransposon families were annotated and analyzed to develop insight on their evolutionary history.

  Innes, Roger W., C. Ameline-Torregrosa, T. Ashfield, T., E. Cannon, S.B. Cannon, B. Chacko, N.W. Chen, A. Couloux, A. Dalvani, R. Denny, S. Deshpade, J.J. Doyle, A. Egan, V. Geffroy, N. Glover, C.S. Hans, S. Howell, D. Ilut, S. Jackson, S., H. Lai, J. Mammadov, S. Martin Del Campo, M. Metcalf, A. Nguyen, M. O'Bleness, B. Pfeil, R. Podicheti, M. Ratnaparkhe, B. Roe, M.A. Saghai, S. Samain, I Sanders, B. Segurens, M. Sevigna, S. Sherman-Broyles, V. Thareau, D. Tucker, J. Walling, A. Wawrzynski, J. Yi, N.D. Young, 2008. Differential Accumulation of Retroelements and Diversification of NB-LRR Disease Resistance Genes in Duplicated Regions Following Polyploidy in the Ancestor of Soybean. Plant Physiology. 148:1740-1759.

  Wawrzynski, A. T. Ashfield, N.W. Chen, J. Mammadov, A. Nguyen, R. Podicheti, S.B. Cannon, V. Thareau, C. Ameline-Torregrosa, E. Cannon, B. Chacko, A. Couloux, A. Dalwani, R. Denny, S. Deshpande, A.N. Egan, N. Glover, S. Howell, D. Ilut, H. Lai, S.M. Del Campo, M. Metcalf, M. Bleness, B.E. Pfeil, M.B. Ratnaparkhe, S. Samain, I. Sanders, B. Segurens, M. Sevignac, S. Sherman-Broyles, D.M. Tucker, J. Yi , J.J. Doyle, V. Geffroy, B.A. Roe, M.A. Maroof , N.D.

Young, R.W. Innes. 2008. Replication of nonautonomous retroelements in soybean appears to be both recent and common. Plant Physiol 148: 1760-1771

A.1.d - Genomic Database Convenient to Access by ALL USERS.

Need an integrated database for use by all users, including breeders, geneticists, genomicists comparative biologists, molecular biologists, biochemists, etc.). This represents a long-term ongoing activity that will be necessary for the community to leverage the genome sequence data for use in all scientific disciplines.

- SoyBase has been expanded to completely integrate the newly developed BAC-based physical maps and the Wm82 whole genome sequence described above. This integration allows a user to rapidly retrieve all relevant information using almost any concept as an entry point. For example starting with a gene name or function a user can quickly find the position(s) of that gene on the genetic and sequence maps, information about that gene and links to the scientific literature for more details.

  SoyBase (http://soybase.org)

- See C.4.

A.1.e - Transposon and Repeat Sequence Databases.

Transposons – known as McClintocks's 'jumping genes' - are ubiquitous in plant genomes and confound the assembly and annotation of the genomes. Therefore, a comprehensive database is necessary.

- Using a semi-automated program developed by researchers at Purdue University identified 32,552 retrotransposons (Class I) and 6,029 DNA transposons (Class II) with clear insertion sites. These elements, together with numerous truncated elements/fragments, make up approximately 61% of the soybean genome. As in other crop species, soybean transposable elements are particularly enriched in centromere and pericentromeric regions. SoyTE, the soybean transposable element database, has been developed by ARS scientists at Ames, IA and is freely accessible for users to search, browse, visualize and download the transposable element (TE) sequences (http://soytedb.org). SoyTE was also integrated with the soybean physical and genetic maps, allowing users to better utilize soybean genetic resources. The 14,106 intact elements and 18,264 solo Long Terminal Repeats (LTRs) are classified into 510 distinct families. Among these, 353 families were categorized as Gypsy-like with the remaining 157 families categorized as Copia-like. The ratio of solo LTRs to intact elements (S/I) in recombination suppressed regions is significantly lower than in chromosome arms. Using the rice and maize centromeric retrotransposons for comparisons, two centromere-specific/enriched families in soybean were identified and characterized. This analysis supported the hypothesis that centromeric retrotransposon elements predate the divergence of dicots and monocots species. The age distribution showed that most LTR-retrotransposons (approximately 95%) in soybean were amplified in the past three million years (mys). Close examination of the largest family, Soybean Nonautonomous and Autonomous Retrotransposon Element (SNARE), in soybean revealed several new aspects of autonomous-nonautonomous retrotransposon interaction. Firstly, SNARE contains two autonomous and one nonautonomous subfamilies. Inter-element recombination mediated by reverse transcription during retrotransposition was frequently observed. Secondly, the bifurcation of non-autonomous elements risen recently from a single lineage into two distinct subgroups corresponds to two autonomous partners that appear to have been brought together in the present soybean genome by a polyploidization event about 15 million years ago. Thirdly, the "proliferation" of an unrelated "piggy-backing" retroelement was mediated by

rapid amplification of a single non-antonomous element, supporting the proposal that nested retrotransposons have the potential to amplify and transpose in the plant genome.

Du, J., D. Grant, Z. Tian, R.T. Nelson, L. Zhu, R.C. Shoemaker, J. Ma. 2010. SoyTEdb: a comprehensive database of transposable elements in the soybean genome. BMC Genomics.11:113.

Du, J., Tian, Z., Schmutz, J., Bowen, N.J., Shoemaker, R.C., and Ma, J. 2010. Bifurcation and enhancement of autonomous-nonautonomous retrotransposon partnership through LTR swapping in soybean. *The Plant Cell*. 22: 48-61

Du, J., Tian, Z., Christian, H., Laten, H., Jackson, S., Cannon, S., Shoemaker, R.C., and Ma, J. 2010. Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: new insights from genome-wide analysis and multi-specific comparison. *The Plant Journal* (in press)

- Knowledge of the structure and behavior of endogenous members of the CACTA transposable element family in soybean (similar to the maize Spm elements of McClintock) was expanded by the first description of the complete sequence of an autonomous CACTA element in soybean. Isolated from alleles of the soybean *T* locus (trichome color) by researchers at the University of Illinois, they are very large elements of 20.5 kilobases. In addition, the behavior of some CACTA elements to generate novel exon combinations by alternatively splicing was documented. This phenomenon has implications for the evolution of novel genes and the generation and mobilization of gene fragments in the soybean genome.

  Zabala, G. and Vodkin, L. O. 2007. Novel exon combinations generated by alternative splicing of gene fragments mobilized by a CACTA transposon in *Glycine max*. BMC Plant Biol 7:38.

  Zabala, G. and Vodkin, L. 2008. A putative autonomous 20.5 kb-CACTA transposon insertion in an *F3'H* allele identifies a new CACTA transposon subfamily in *Glycine max.* BMC Plant Biology 8: 124.

A.2 – Genome Finishing

There is need to fill gaps in highly repetitive areas, centromeres, and within many scaffolds of the draft soybean genome sequence.

A.2.a - Initial Genome Assembly.

- The size of the initial soybean genome assembly is estimated at approximately 975Mb, organized in 20 chromosomes. Small additional amounts of mostly repetitive sequence remain in unmapped scaffolds. About 66,153 protein-coding loci have been predicted. These include: 3305 full-length cDNA consistent genes, 32317 EST consistent genes, 7361 EST overlaps, 1832 genes generated from the longest ORF of EST evidence, 13704 homology and Solexa supported genes, and 7634 genes with homology to other plant peptides.

  SoyBase (http://soybase.org)
  Phytozome v5.0 http://www.phytozome.net)

A.2.b - Initial Annotation of Genome Sequence.

- See A.1.d

A.2.c - Selective Re-Sequencing.

- Jeremy Schmutz at JGI-Stanford Human Genome Center and Steve Cannon (ARS, Ames IA) led collaborations that determined a majority of *Glycine max* ESTs align to the genome at

nearly 100% identity, suggesting that Glyma1 is highly accurate in genic regions. The base-pair-level accuracy in repetitive regions was evaluated by comparing the assembly with BAC clones produced for the project. Discrepancies between the shotgun assembly and the independently obtained genetic and physical maps have been manually reviewed and corrected to minimize potential errors in the large-scale structure of the genome.

- A multi-agency effort, with DOE/JGI producing the sequence, and USDA-ARS "anchoring" the sequence to a genetic map, led to the integration of the soybean genetic map and soybean genomic sequence and produced the first high-quality draft assembly of the full soybean genome sequence. Cutting-edge sequencing and marker technologies, and novel bioinformatic methods, were used to develop 1536 new genetic SNP markers, designed specifically to genetically anchor soybean whole genome shotgun (WGS) scaffold sequences. Selected markers were mapped using high-throughput Solexa-generated sequences from *Glycine soja* (the wild progenitor of soybean) onto draft genomic sequences of *Glycine max* to identify regions with scorable differences. Mapping was carried out by ARS scientists at Beltsville MD. These new SNP map and existing maps were used to order and orient scaffold sequences into full chromosome assemblies. The chromosome assemblies captured more than 98% of the assembled WGS sequences. Sequence scaffold positions and orientations within the chromosome sequences were validated over more than 95% of their lengths. The chromosome assemblies are now the most highly-accessed portion of the SoyBase database, with the genome browser attracting more than 5,000 page views (>1,500 unique visitors) from more than 80 countries per month in November 2009.

  SoyBase (http://soybase.org)

  Shoemaker, R., D. Grant, T. Olson, W.C. Warren, R. Wing, Y. Yu, H-R. Kim, P. Cregan, B. Joseph, M. Futrell-Griggs, W. Nelson, J. Davito, J. Walker, J. Wallis, C. Kremitski, D. Scheer, S. Clifton, T. Graves, H. Nguyen, X. Wu, M. Luo, J. Dvorak, R. Nelson, S. Cannon, J. Tomkins, J. Schmutz, G. Stacey and S. Jackson. 2008. Microsatellite Discovery from BAC End Sequences and Genetic Mapping to Anchor the Soybean Physical and Genetic Maps. Genome 51(4):294-302.

  Hyten, D.L., S.B. Cannon, Q. Song, N.T. Weeks, E.W. Fickus, R.C. Shoemaker, J. Specht, A.D. Farmer, G.D. May, P.B. Cregan. High-Throughput SNP Discovery through Deep Resequencing of a Reduced Representation Library to Anchor and orient Scaffolds in the Soybean Whole Genome Sequence. (Accepted by BMC Genomics on November 25, 2009).

  Grant, D. M., Nelson, R. T., Cannon, S. B., Shoemaker, R. C. 2010. SoyBase, The USDA-ARS Soybean Genetics and Genomics Database. Nucleic Acids Research, Nucleic Acids Res. 38: D843 - D846.

A.4 – Genome Re-sequencing

Development of markers and mapping of traits to facilitate marker-assisted soybean breeding will require some limited 'resequencing' of related genomes.

A.4.a - SNP Genotyping.

- ARS scientists at Beltsville MD with the support of the United Soybean Board were the first to apply the high-throughput SNP genotyping technology known as the GoldenGate assay to soybean. Soybean has a highly duplicated genome and the ability to use SNP genotyping technology is often hampered by paralogous sequences that make it impossible to genotype SNPs accurately. This technology featured a custom 384 SNP assay in three mapping populations and 192 diverse soybean landrace accessions and elite cultivars; and found that known SNPs successfully converted to GoldenGate assays 91% of the time. This work also

was the first to demonstrate that the GoldenGate assay could be used to perform bulked segregant analysis (BSA). A 1,536 SNP GoldenGate assay was used to perform BSA to map the soybean rust resistant gene, Rpp3, to a single candidate region in the genome and confirmed the position with markers genotyped in the entire population. The ease of genotyping large numbers of lines with 1,536 SNP markers is leading to QTL mapping projects which have more power to detect QTL and to accurately estimate QTL effects. Collaborations are current with ARS scientists at Wooster, OH; Raleigh, NC; Urbana, IL, and university scientists at Ohio State University; the University of Nebraska; the University of Illinois; University of Georgia; North Dakota State University; Virginia PolyTech; University of Tennessee; University of Missouri; North Carolina State University in QTL mapping efforts and germplasm characterization. In these collaborations a total of over 16,000 DNA samples are being genotyped with a 1,536 Universal Soybean Linkage Panel (USLP 1.0) which consists of SNPs from previous GoldenGate assays that produce robust reliable assays. These SNPs are spread throughout the genome and have high minor allele frequencies in diverse germplasm. In addition, Pioneer Hybred International and Monsanto Inc. have also used the USLP 1.0 for genotyping efforts.

Hyten, D.L., I. Choi, Q. Song, J.E. Specht, T.E. Carter, R.C. Shoemaker, E. Hwang,, L.K. Matukumalli, P.B. Cregan, A high density integrated genetic linkage map of soybean and the development of a 1536 Universal Soy Linkage Panel for quantitative trait locus mapping. Crop Science, in press.

- Development and application of the Universal 1536 Soy Linkage Panel 1.0 (USLP 1.0). The analysis of hundreds of DNA markers on hundreds of soybean lines in populations developed by geneticists for the discovery of genetic factors that underlie traits of importance for soybean genetic improvement is laborious and time consuming. Single nucleotide polymorphisms (SNPs) are the DNA markers of choice for many researchers due to their abundance and the rapid high-throughput methods available for the characterization of many markers in one analysis. ARS scientists in the Soybean Genomics and Improvement Laboratory in Beltsville, MD with collaborators at the Univ. of Nebraska, the Univ. of Maryland, and with ARS scientists at Ames, IA and Raleigh, NC developed a selected set of 1536 SNP DNA markers that can be analyzed on 192 soybean DNA samples in a three day period using the Illumina GoldenGate assay. The set of 1536 markers is referred to as the Universal Soy Linkage Panel 1.0 (USLP 1.0) and was selected from more than 3000 markers, each of which was genetically mapped to identify markers with equidistant spacing across each of the 20 soybean consensus linkage groups that define the 20 soybean chromosome pairs. In addition, the 3000 markers were used to analyze 96 elite soybean varieties and 96 diverse accessions from the USDA Soybean Germplasm Collection in order to identify those markers that were maximally informative. The 1536 SNP markers in the USLP 1.0 will permit rapid genetic analysis of soybean populations that are segregating for important genetically controlled traits affecting disease and insect resistance, environmental stress such a drought, seed quality traits including soybean oil quality and protein concentration as well as seed yield and to discover the genome positions of genes controlling these traits.

A.4.b - Resequencing for SNP Discovery.

- ARS scientists at Beltsville MD and Urbana IL with the support of the United Soybean Board report that progress was made in discovery of SNPs for the genetic analysis of soybean germplasm with 37,000 SNP markers. Reduced representation DNA libraries were created from genomic DNA of the soybean cultivars Evans, Essex, Archer, Minsoy, Peking, Noir 1 as well as of the wild soybean PI 468916. In addition, similar libraries were created

with a mix of DNA of the genotypes Minsoy, Noir 1, Archer, Evans, and Peking. The genomic DNA libraries consisted of restriction digested fragments that were "size selected" to only include DNA fragments in the 100-150 basepair size range. Following numerous runs on the Solexa/Illumina Genome Analyzer more than 7 billion bases of DNA sequence was obtained from these various libraries. These were aligned with the whole genome sequence of soybean to facilitate the discovery of more than 60,000 putative SNPs. These will be used for the discovery and genotyping of SNPs in the entire soybean germplasm collection which consists of 19,000+ soybean accessions. The genotyping of the germplasm collection will facilitate more sophisticated association genetic analyses for traits of agronomic importance.

Hyten, D.L., Song, Q., Choi, I-Y., Yoon, M-S., Specht, J.E., Matukumalli, L.K., Nelson, R.L., Shoemaker, R.C., Young, N.D., and Cregan, P.B. High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor. Appl. Genet. 116:945-952. 2008.

A.4.c - Other Technologies – Simultaneous SNP Discovery and Genotyping.

- Soybean was the first plant species in which the Simple sequence repeat (SSR) genetic markers were demonstrated to function as useful markers. ARS scientists at Beltsville MD and Ames, IA with the support of the United Soybean Board used the newly completed soybean whole genome sequence to search for the presence of SSR sequences and found a total of 210,990 SSRs with di-, tri- and tetra-nucleotide repeats of five or more which included 61,458 SSRs consisting of repeat units of di-(≥10), tri-(≥8), and tetra-nucleotide (≥7). After screening for a number of factors including locus-specificity using electronic-PCR, a soybean SSR database **(BARCSOYSSR_1.0)** with the genome position and primer sequences for 33,065 SSRs was created. The average density of these SSR loci in the genome was 35/Mbp. To examine the likelihood that primers in the database would function to amplify locus-specific polymorphic products, 1034 primer sets were evaluated by amplifying DNAs of seven diverse *G. max* and one wild soybean genotypes. A total of 978 (94.6%) of the primer sets amplified a single discrete PCR product and 798 (77.2%) amplified polymorphic amplicons as determined by 4.5% agarose gel electrophoresis. Information on the BARCSOYSSR1.0 SSR markers can be found in SoyBase (http://soybase.org) the USDA, ARS, Soybean Genome Database. The inexpensive and easily used SSR markers in the BARCSOYSSR1.0 database can be used to identify markers at any position in the soybean genome for purposes of map-based gene cloning and marker assisted selection.

  Song, Q., G. Jia, Y. Zhu, D. Grant, R.T. Nelson, E. Hwang, D.L. Hyten, P.B. Cregan. 2010. Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. Crop Science, in press.

- Single-feature polymorphism (SFP) discovery is a rapid and cost-effective approach for plant genomic polymorphism studies. However, either a high false positive rate or low sensitivity was reported in previous SFP detection methods. ARS scientists at St Paul MN developed an alternative method for genome-wide SFP discovery by computing probe affinity differences and affinity shape powers formed between the neighboring probes of Affymetrix GeneChip and the expressed transcripts from different genotypes. A 2.05% false positive rate (FPR) and a sensitivity of 69.08% resulted from testing the published barley GeneChip data set with known SNPs and known non-polymorphic sequences. This method was used for SFP discovery in barley bowman and Bowman-uniculm2 mutant lines, and 365 SFPs were found. Fifteen SFPs out of 16 were confirmed by PCR sequencing (93.75%). Most of the SFP

located genes were functionally annotated in this study. This method can be applied to other organisms for genome-scope SFP discovery.

Xu, W., S. Cho, S.H.Yang, Y.E. Bolon, H. Jia, Y. Xiong, H. Bilgic, J. Boddu, G. Muehlbauer. 2009. Single-Feature Polymorphism Discovery by Computing Probe Affinity Shape Powers. BioMed Central (BMC) Genetics. 10:48

- The low phytate (LP) trait in plant seeds offers important nutritional and environmental benefits for food and feed uses. Mutants with reduced phytate content are commonly produced by chemical mutagenesis, as is the case for soybean [Glycine max (L.) Merr.] line CX1834. Quantitative trait loci (QTL) for LP in CX1834 have previously been reported, however the genetic basis for this trait has not been identified. In this study, we examined several possible chromosomal map locations for the LP mutation in CX1834. After eliminating the myo-inositol phosphate synthase (MIPS) gene family as the location of the LP mutations based on mapping studies, we focused on candidate genes in regions of known low phytate QTL on linkage groups (LGs) L and N. Using the soybean whole genome sequence, we identified genes encoding two putative multidrug resistance-associated proteins (MRPs). The sequence of a segment of the CX1834 MRP gene on LG N was compared with sequences from 16 other soybean lines with normal phytate content. A single nucleotide mutation from A to T, resulting in substitution of a stop codon for an Arg residue, was detected in the MRP gene on LG N of the low phytate line, but not in the normal counterparts. Further comparative sequence analysis of normal and LP progeny from a cross of CX1834 with V99-3337 (normal phytate) also indicated the A to T substitution in LP individuals, suggesting a mutation in an MRP gene as the possible cause of the low phytate phenotype.

  Maroof, Saghai, Natasha M. Glover, Ruslan M. Biyashev, Glenn R. Buss, Elizabeth A. Grabau. Genetic Basis of the Low Phytate Trait in Soybean Line CX1834. 2009. Crop Science 49:69-76

ARS scientists at Columbia, MO and the University of Tennessee defined the two mutations responsible for the low phytate trait and developed perfect molecular marker assays.This research was USB-funded.

Gillman JD, Pantalone VR, Bilyeu K: The low phytic acid phenotype in soybean line CX1834 is due to mutations in two homologs of the maize low phytic acid gene. *The Plant Genome* 2009, 2(2):179-190.

- University of Missouri scientists conducted genome-wide SFP detection in soybean using genomic DNA from Williams 82 and Forrest.. The validation rate for 200 SFP markers was 80%. This efficient and cost-effective method can be used to identify genome-wide markers in soybean.

  Xiaolei Wu, Jin Qiu, Henry Nguyen (2010) Genome-wide detection of soybean single-feature polymorphism by Affymatrix GeneChip using genomic DNA. (in preparation)

A.5 – Phaseoloid Genomics

Phaseolus vulgaris (dry bean) is a diploid species with sequence and genetic map synteny with soybean. *Phaseolus* diverged from *Glycine* about 19-23 mya, and may be very useful for the discovery of genes for protection against abiotic stresses in soybean. .

A.5.a - Genomic Research Goals for *Phaseolus*.

- ARS scientists at Beltsville MD initiated the identification of genetic variability for nutritional traits in common bean. Projects have been initiated for discovery of the genes or quantitative trait loci (QTL) with DNA markers that can be used by plant breeders to

incorporate QTL into new and more nutritional cultivars. Single Nucleotide Polymorphism (SNP) DNA markers will be discovered via the use of high throughput DNA sequence analysis using the Illumina/Solexa Genome Analyzer as well as via the analysis of common bean DNA gene fragments amplified using polymerase chain reaction (PCR) primers designed to orthologous soybean genes. Because common bean market classes fall into three distinct groups based on their geographic origins in Central or S. America, three sets of 768 SNP panels specific to the three groups will be developed. These three sets of SNPs will be used to characterize common bean populations segregating for nutritional traits for purposes of QTL discovery. The SNP analysis will be conducted using the Illumina GoldenGate assay which is analyzed on the Illumina BeadStation.

Hyten, D.L., Q. Song, E.W. Fickus, I. Choi, C.V. Quigley, E. Hwang, M. Pastor-Coralles, P.B. Cregan. High-throughput SNP discovery and assay development in Common Bean. BMC Genomics, submitted.

- A research team coordinated by Scott Jackson has developed a physical map for *Phaseolus vulgaris* funded by the USDA and is currently sequencing the common bean genome (also funded by USDA).

Schlueter, J.A., J.L. Goicoechea, K. Collura, N. Gill, J-Y Lin, Y.Yu, D. Kudrna, A. Zuccolo, C.E. Vallejos, M. Munoz-Torres, M.W. Blair, J. Tohme, J. Tomkins, P. McClean, R.A. Wing and S.A. Jackson 2008. BAC-end sequence analysis and a draft physical map of the common bean (Phaseolus vulgaris L.) genome. Trop. Plant Biol. DOI 10.1007/s12042-007-9003-9.

A.5.b - Other Genera.

The perennial *Glycine* species are of interest because of the nature of polyploidy in *Glycine* as a whole, and thus for understanding the duplicated nature of the soybean genome. These species diverged from soybean (and its progenitor, *G. soja*) around 5 mya .The perennial species constitute the secondary germplasm pool for soybean, and may provide traits such as drought tolerance and rust resistance.

- The 26 wild perennial species of the subgenus Glycine Willd. that are indigenous to Australia are an untapped source of genetic diversity for soybean breeding. ARS scientists at Urbana IL developed an efficient methodology by which G. tomentella Hayata (2n=78) can be hybridized with the soybean (2n=40) and self-fertile 2n=40, 41, and 42 progeny produced. PI 441001 (G. tomentella) was crossed to the soybean cultivar Dwight. F1 plants contained the expected 2n=59 chromosome complement and amphidiploid (2n=118) plants were produced by colchicine treatment. Dwight was used as a recurrent parent for producing BC1 (2n=79) to BC5 plants. BC2 plants contained 2n=55 to 59 chromosomes and BC3 plants contained 2n=41 to 50 chromosomes. Plants with 2n=41 (monosomic alien addition lines; MAALs) were usually self fertile whereas plants with greater than 42-chromosomes were male sterile and backcrossed to Dwight. At least one 2n=42 genetically and chromosomally stable line has been identified. The progenies of MAALs produce approximately 70% plants with 2n=40, 28% plants with 2n=41 and 2% plants with 2n=42 (disomic alien addition lines; DAALs). At least one 2n=42 genetically and chromosomally stable line has been identified. DAALs and diploid lines are being screened for resistant to pest and pathogens, and changes in seed composition.
- Shoemaker, Cregan, Ma, Schmutz, Schlueter and Jackson coordinate an $8 million NSF grant to develop tools for whole genomic analysis of wild Glycine relatives.

B – Gene Function

The soybean genome sequence may reveal as many as 60-65,000 genes, many with unknown function. Tools are needed to facilitate discovery of gene function. Annotation is obviously a critical need, but ability to target specific genes for mutation or for gene silencing is equally important. Functional genomic tools (e.g., transcriptomics, proteomics and metabolomics) will allow gene function to be framed within the context of soybean physiology. Use of these technologies (e.g., using DNA microarrays to map expression QTLs) will advance soybean improvement in the areas of: seed composition and quality; abiotic stresses (drought, plant iron chlorosis, temperature, flooding, ozone), and biotic stresses (soybean cyst nematode, fungal diseases including rust, aphids and foliar feeding insects, and viral diseases, BPMV, SMV)

B.1 – Gene Function Annotation/Informatics

Gene function annotation is necessary to provide breeders and other soybean geneticists with a sequence that will be usable for soybean improvement. Three key areas were discussed: i) genome annotation, ii) resources to improve gene modeling, and iii) informatic support for comparative genomics of gene function.

B.1.a - Gene Prediction and Confirmation in the Genome Sequence.

- Consensus EST sequences of *Glycine max*, *Medicago truncatula* and *Lotus japonicus* from the TIGR Plantta databasewere aligned to the soybean genome using Jim Kent's BLAT and filtered for best hit to the genome. A hit with 97% coverage was required to account for genome duplication. For final gene verification *G. max* ESTs were aligned using Brian Haas's PASA pipeline which aligns ESTs to the best place in the genome via gmap, then filters hits to ensure proper splice boundaries. The Arabidopsis and rice peptides from NCBI RefSeq were aligned to the (unmasked) genome by gapped BLASTX and high-scoring sequence pairs (HSP's) to increase sensitivity. The current "Glyma1.0" gene set is based on the homology-gene prediction program GenomeScan from Chris Burge, FgenesH predictions provided by Asaf Salamov at JGI and with PASA analysis to integrate over 1.6 million soybean ESTs. The gene set shown on the SoyBase sequence browser and at Phytozome was produced by Therese Mitros at UC Berkeley. Briefly, peptides from diverse angiosperms and TIGR legume EST assemblies were aligned to the genome, and their overlaps used to define putative protein-coding gene loci. The corresponding genomic region was submitted to GenomeScan and FgenesH, along with related angiosperm peptides and/or ORFs from overlapping EST assemblies. GenomeScan identifies likely protein coding exons, favoring regions that align well to the given homologous peptides. These homology-based predictions were integrated with expressed sequence information using PASA (Haas et al. 2003) using legume ESTs. The results were filtered to remove genes identified as transposon-related. Genes with apparently truncated ORFs may be prediction errors or pseudogenes.

  SoyBase (http://soybase.org)
  Phytozome v5.0 http://www.phytozome.net)

- Genetic factors underlying a major quantitative trait locus (QTL) contributing to low seed stachyose content were investigated in two separate populations derived from soybean line PI200508. The Williams 82 whole genome shotgun (WGS) sequence was exploited for candidate gene discovery. The physical interval containing a low stachyose QTL from PI200508 was identified in the WGS, and screened for areas that could be exploited for linkage mapping purposes. Microsatellite sequences designed in these areas were used to develop new markers, creating tighter linkage with the QTL. Examination of the Williams 82 sequence in this interval and the corresponding glyma0.1b gene model, available through the Phytozome website (http://www.phytozome.org), revealed one candidate gene with significant homology to previously characterized galactosyltransferase genes from

Arabidopsis, pea, and cucumber. Sequencing of the annotated coding region for this gene revealed a single unique sequence polymorphism between PI200508, and lines exhibiting wild type expression of stachyose content. The mutant phenotype appears to have arisen from a 3 bp deletion, and is easily distinguished from the wild type allele via a marker presented in this study. This mutation-specific marker explained 88% to 94% of the phenotypic variation for seed stachyose content, and 76% for seed sucrose content, traits that exhibited a strong negative correlation.

Skoneczka, J, M. A. Saghai Maroof, C. Shang, and G. R. Buss. 2009. Identification of Candidate Gene Mutation Associated With Low Stachyose Phenotype in Soybean Line PI200508 Crop Science 49:247-255

- ARS scientists in Columbia, MO and University of Missouri scientists defined the molecular genetic basis for the low raffinose and stachyose seed trait, developed perfect molecular marker assays, and demonstrated adequate germination for soybean lines with the low raffinose and stachyose seed trait.Part of this research was USB-funded.

  Dierking EC, Bilyeu KD: Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. *The Plant Genome* 2008, 1(2):135-145.

  Dierking EC, Bilyeu KD: Raffinose and stachyose metabolism are not required for efficient soybean seed germination. *Journal of Plant Physiology* 2009, 166(12):1329-1335.

B.1.b - Resources Needed for Genome Annotation.

- See A2a
- Scientists at the University of Missouri have developed an infrastructure for storing, managing, and analyzing soybean data. A gene expression data management tool BASE and physical map viewer WebFPC, populated with internal experimental data, have been installed. Computational analyses are performed for the ESTs from soybean root samples under drought stress. Out of a total 15,168 EST sequences collected by NCSB scientists, 13,430 EST sequences passed the quality control and have been submitted to GenBank. Unigene analysis identified 1331 novel genes in these ESTs. Gene Ontology classification and biological pathway projection for these ESTs have been added to the KEGG database.
- Bioinformatics faculty at the University of Missouri designed a comprehensive and searchable database to display a variety of data types for all of the annotated soybean transcription factor genes.

  http://casp.rnet.missouri.edu/soydb/

  Wang Z, Libault M, Joshi T, Valliyodan B, Nguyen HT, Xu D, Stacey G, Cheng J. 2010. SoyDB: A Knowledge Database of Soybean Transcription Factors. BMC Plant Biology 2010, 10:14.

- Bioinformatics faculty at the University of Missouri have completed the SoyKB (soybean knowledge base), which seeks to integrate all forms of soybean functional data with the genome sequence. This website is now populated by data obtained at the University of Missouri but is seeking inputs from other members of the community.

  http://soykb.org/

- ARS scientists at Beltsville bred a new soybean 'MiniMax' specifically for research and educational uses. MiniMax is much shorter that elite soybean varieties, so it takes up much

less space in the laboratory and classroom. It also has a shorter life cycle, which can be completed in 12 weeks. This will allow more rapid and less expensive genomic studies

Klink, V.P., M.H. MacDonald, V.E. Martins, S.C. Park, K.H. Kim, S.H. Baek, B.F. Matthews. 2008. *MiniMax,* a new diminutive *Glycine max* genotype with a rapid life cycle, embryonic potential and transformation capabilities. Plant Cell and Tissue and Organ Culture 92:183-195

B.1.c - Integration with Other Species.

- See B.1a

B.2 – Discovery via Mutagenesis

Although gene annotation may suggest a function, it is necessary to confirm this function through biochemical or genetic studies. It is also expected that the function of the majority of soybean genes will not be easily deduced simply by sequence comparison to other genomes. In these cases, the availability of mutations in each of the soybean genes will be extremely useful to decipher gene function and integrate this function into the context of soybean quality and agronomic performance. A variety of new technologies are available to generate mutations or to down-regulate gene expression, such as RNAi-mediated gene silencing, viral induced gene silencing (VIGS). A variety of other mutagenesis procedures also are needed to create robust platforms for the study of soybean gene function.

B.2.a - Reverse Genetics to Determine Gene Function: TILLING.

TILLING (Targeting Induced Local Lesions IN Genomes) is a PCR-based high-throughput mutation detection system that permits the identification of point mutations and small insertions and deletion "Indels" in pre-selected genes.

- Scientists at Southern Illinois University, ARS at W. Lafayette IN, and Columbia MO have played an integral role in the establishment of soybean TILLING libraries. The EMS Tilling library at SIU (from cvs. FORREST and WILLIAMS 82) contains 6150 M2 families. This service may be accessed via: http://www.soybeantilling.org/tilling.jsp. Using a complementary approach TILLING was applied to four mutagenized soybean populations, three of which were treated with ethyl methanesulfonate (EMS) and one with N-nitroso-N-methylurea (NMU). Seven targets were screened in each population and revealed 116 induced mutations. The NMU-treated population and one EMS mutagenized population had similar mutation density (~1/140 kb), while another EMS population had a mutation density of ~1/250 kb. The remaining population had a mutation density of ~1/550 kb. Because of soybean's polyploid history, PCR amplification of multiple targets could impede mutation discovery. Indeed, one set of primers tested in this study amplified more than a single target and produced low quality data. This was overcome by pretreating genomic DNA with a restriction enzyme. Digestion of the template eliminated amplification of the extraneous target and allowed the identification of four additional mutant alleles compared to untreated template. The development of four independent populations with considerable mutation density, together with an additional method for screening closely related targets, indicates that soybean is a suitable organism for high-throughput mutation discovery even with its extensively duplicated genome.

  Cooper, Jennifer L., Bradley J Till, Robert G Laport, Margaret C Darlow, Justin M Kleffner, Aziz Jamai, Tarik El-Mellouki, Shiming Liu, Rae Ritchie, Niels Nielsen, Kristin D Bilyeu, Khalid Meksem, Luca Comai and Steven Henikoff .2008. TILLING to detect induced mutations in soybean, BMC Plant Biology, 8:9

ARS scientists at Columbia, MO and University of Missouri scientists utilized a reverse genetics TILLiNG resource to identify the first induced mutations responsible for seed composition traits

in soybean: one for the low raffinose and stachyose trait in the *RS2* gene and one for the elevated oleic acid seed oil trait in the *FAD2-1A* gene. Molecular marker assays to directly select the desired alleles were also delivered

> Dierking E, Bilyeu K: New sources of soybean seed meal and oil composition traits identified through TILLING. *BMC Plant Biology* 2009, 9(1):89.

B.2.b - Forward/Reverse Genetic Approaches - Fast-Neutron Mutagenesis.
Fast neutron mutagenesis induces small deletions in the genome and is a very effective way to create mutations. Fast-neutron populations can be used for both forward and reverse genetic screens for useful mutations.

- Wayne Parrott, University of Georgia; Tom Clemente, University of Nebraska; Gary Stacey, University of Missouri, Columbia; Carroll Vance, USDA-ARS, St. Paul MN; and Zhanyuan J. Zhang, University of Missouri, Columbia collaborated under NSF grant #0820769 to develop forward and reverse genetics tools, and produce mutant soybean populations to link phenotypes to their underlying genomic sequence changes. These populations are being produced using fast neutron mutagenesis and three transposon tagging systems (Ac/Ds, Tnt1, and mPing). This work has identified lines that produce heritable mPing insertions and show increases in mPing copy number. Analysis of the mPing insertion sites indicate that it transposes to unlinked sites with a preference for gene-rich regions. Efforts are underway to screen for mPing-induced mutations and develop mPing-based activation tagging constructs. This project has produced a collection of Ds element containing lines, and the possibility that Tnt1 can be activated by tissue culture is being reexamined. Active transposable elements have the ability to generate novel single copy insertions, thus, reducing the number of transformations needed to produce a mutant population. This is especially crucial for soybean, where the transformation efficiency is relatively low. Activation tagging resources also are being developed that are designed to induce mutations that alter expression patterns in addition to disrupting coding regions. This project includes three transposon-based strategies: the tobacco retrotransposable element Tnt1, the maize Ac/Ds transposon, and the rice miniature inverted terminal repeat element (MITE) transposon mPing. As mutant lines are produced, information about the mutagen, transposon insertion sites, and associated phenotypes is deposited into a public database. These mutant lines will be available through a seed bank facility. More detailed information can be found at **http://soymutants.uga.edu**.

  > Hancock, C. Nathan., Kristen Chesser, Tom Clemente, Gary Stacey, Carroll P. Vance, Zhanyuan J. Zhang, and Wayne A. Parrott. 2010. Transposon Tagging And Fast Neutron Mutagenesis In Soybean, Plant & Animal Genomes XVIII Conference, January 9-13, 2010

B.2.c - Resources Needed for Gene Function Studies Using Transposon Tagging.
Transposon mutagenesis has a high likelihood of disrupting gene function. Usually, only a few transposon insertions occur in any given mutagenized line, making genetic analysis much easier. Recent work has demonstrated that both the maize Ac/Ds and tobacco Tnt1 retrotransposon are suitable for use in soybean. The Ac/Ds transposon has the drawback that transpositions are local and, therefore, one requires a starting population of well dispersed Ds insertions (perhaps 25,000) before any given soybean gene can be targeted. The Tnt1 retrotransposon is still being evaluated but promises to provide a large number of random mutations.

- Scientists at the University of Missouri and Purdue University with support from the United Soybean Board have used transposon tagging to define the function of the 30-50,000 genes present in soybean. NCSB scientists are using two transposons systems, the retrotransposon Tnt1 originally isolated from tobacco and the Ds element originally from maize. Twelve

activation-tagging Ds vectors were constructed, each containing various combinations of either the CaMV 35s promoter or individual seed-specific promoters (e.g., phaseolin, soybean seed lectin, or glycinin). These vectors were subsequently used to construct transgenic soybean plants by Agrobacterium-mediated transformation. The project is at an early stage. However, thus far, around 500 independent transgenic soybean lines have been developed using the different Tnt1 and Ds element constructs. In order to ascertain the insertion site, flanking DNA sequences were isolated by PCR walking and sequenced. Primers developed from these flanking sequences will be used to localize the insertion sites on the soybean physical map, which is anchored to the genetic map. A Soybean Transposon Insertion Mutant Database is under development to serve as a resource to the community. The goal of the project is the development of sufficient transposon insertions so as to target any soybean gene.

The soybean mutant database can be accessed at: **http://digbio.missouri.edu/gmgenedb/**;

Mathieu, Melanie, Elizabeth K. Winters, Fanming Kong, Jinrong Wan, Shaoxing Wang, Helene Eckert, Christopher Donovan, David Somers, Kan Wang, Gary Stacey and Tom Clemente. 2009. Establishment of a soybean (*Glycine max* Merr. L) transposon-based mutagenesis respository. Planta 229: 279-289.

B.3 – Functional Genomics Approaches

B.3.a - Transcriptomic Approaches.

Past work by the soybean community has provided a rich resource of DNA microarrays. Arrays are available covering roughly 36,000 genes identified through EST sequencing. Once the genome is completed an additional 20-30,000 genes will need to be added to existing arrays.

- Soybeans grown in the upper Midwest often suffer from iron deficiency chlorosis, which results in yield loss at the end of the season. To better understand the effect of iron availability on soybean yield, genes were identified in two near isogenic lines with changes in expression patterns when plants were grown in iron sufficient and iron deficient conditions. Transcriptional profiles of soybean (*Glycine max*, L. Merr) near isogenic lines Clark (PI548553, iron efficient) and IsoClark (PI547430, iron inefficient) grown under Fe-sufficient and Fe-limited conditions were analyzed and compared using the Affymetrix® GeneChip® Soybean Genome Array. There were 835 candidate genes in the Clark (PI548553) genotype and 200 candidate genes in the IsoClark (PI547430) genotype putatively involved in soybean's iron stress response. Of these candidate genes, fifty-eight genes in the Clark genotype were identified with a genetic location within known iron efficiency QTL and 21 in the IsoClark genotype. The arrays also identified 170 single feature polymorphisms (SFPs) specific to either Clark or IsoClark. A sliding window analysis of the microarray data and the 7X genome assembly coupled with an iterative model of the data showed the candidate genes are clustered in the genome. An analysis of 5' untranslated regions in the promoter of candidate genes identified 11 conserved motifs in 248 differentially expressed genes, all from the Clark genotype, representing 129 clusters identified earlier, confirming the cluster analysis results. These analyses identified the first genes with expression patterns that are affected by iron stress and are located within QTL specific to iron deficiency stress. The genetic location and promoter motif analysis results support the hypothesis that the differentially expressed genes are coregulated. The combined results of all analyses lead us to postulate iron inefficiency in soybean is a result of a mutation in a transcription factor(s), which controls the expression of genes required in inducing an iron stress response.

O'Rourke, Jamie A., Rex T Nelson, David Grant, Jeremy Schmutz, Jane Grimwood, Steven Cannon, Carroll P Vance, Michelle A Graham and Randy C Shoemaker. 2009. Integrating microarray analysis and the soybean genome to understand the soybeans iron deficiency response. *BMC Genomics* 2009, 10:376

- A strategy using near-isogenic lines (NILs) and Affymetrix Soybean GeneChip microarrays was employed to identify genetic markers closely linked to the soybean aphid [Aphis glycines Matsumura (Hemiptera: Aphididae)] resistance gene Rag1 in soybean [Glycine max (L.) Merr.]. Genomic DNA from the aphid-resistant cultivar Dowling and the aphid-susceptible cultivar Dwight was labeled and hybridized to arrays, identifying more than 1500 putative single feature polymorphisms (SFPs) between these genotypes. Polymorphisms closely linked to the Rag1 aphid-resistance locus were found in genomic DNA from two NILs developed through backcrossing Rag1 from Dowling four times to Dwight. Comparison of hybridization signals identified more than 70 SFPs in the NIL and the recurrent parent genotype. There were 22 SFPs shared by both NILs, representing molecular markers putatively linked to Rag1. Four selected SFPs were converted to SNP markers and confirmed by conventional genetic mapping to be closely linked to Rag1. This technique can identify polymorphisms in a genetic region and generate molecular markers closely linked to an agronomically important trait using a suitable oligonucleotide microarray.

  Kaczorowski, K.A., K.-S. Kim, B.W. Diers, and M.E. Hudson. 2008. Microarray-Based Genetic Mapping Using Soybean Near-Isogenic Lines and Generation of SNP Markers in the Rag1 Aphid-Resistance Interval, Plant Genome 1:89-98.

- Compatible virus infection induces and suppresses host gene expression at the global level. These gene-expression changes are the molecular basis of symptom development and general stress and defence-like responses of the host. To assess transcriptional changes in soybean plants infected with soybean mosaic virus (SMV), the first soybean trifoliate leaf, immediately above the SMV-inoculated unifoliate leaf, was sampled at 7, 14 and 21 days post-inoculation (p.i.) and subjected to microarray analysis. The identified changes in gene expression in soybean leaves with SMV infection at different time points were associated with the observed symptom development. By using stringent selection criteria (2- or –2-fold change and a Q value of 0.05), 273 (1.5 %) and 173 (0.9 %) transcripts were identified to be up- and downregulated, respectively, from 18 613 soybean cDNAs on the array. The expression levels of many transcripts encoding proteins for hormone metabolism, cell-wall biogenesis, chloroplast functions and photosynthesis were repressed at 14 days p.i. and were associated with the highest levels of viral RNA in the host cells. A number of transcripts corresponding to genes involved in defence were either down regulated or not affected at the early stages of infection, but upregulated at the late stages, indicating that the plant immune response is not activated until the late time points of infection. Such a delayed defense response may be critical for SMV to establish its systemic infection.

  Mohan, Babu, Alla G. Gagarinova, James E. Brandle and Aiming Wang. 2008. Association of the transcriptional response of soybean plants with soybean mosaic virus systemic infection, J Gen Virol 89 (2008), 1069-1080

- Scientists at the University of Tennessee and ARS at Jackson TN utilized molecular approaches to potentially enhance soybean resistance to SCN. The researchers employed microarrays to compare gene expression in a susceptible and resistant soybean to identify soybean genes involved in defense to SCN. Two Tennessee soybean lines TN02-226 (resistant) and TN02-275 (susceptible) against SCN race 2 were used. These two soybean lines are sisters from the cross Anand X Fowler soybean cultivars. A susceptible and

resistance response was established for SCN race 2 infection. Results indicated that soybean TN02-226 is a highly resistant line and TN02-275 is a moderate susceptible line. Moreover, the reaction data showed a 10-fold difference for these lines in response to SCN infection. Inoculation experiments were performed to obtain soybean tissues for microarrays. RNA was isolated from root samples, and screened with the GeneChip soybean genome microarray via Affymetrix Core Facility at the University of Tennessee. The objectives to be completed during the coming year include: • Complete microarray analysis to assess gene expression in a susceptible and a resistance response of soybean to SCN infection; • Detect molecular events of defense occurring during the SCN infection and identify soybean defense-related genes to SCN using time course assays on a defined set of genes as identified; and • Transfer these candidate genes in soybean by generating transgenic soybean hairy roots and assay them for conferring resistance to SCN. This project is funded by Tennessee Soybean Production Board

- Scientists at the University of Missouri, the National Center Genomic Research and ARS at Urbana IL evaluated nodulation between soybean and Bradyrhizobium japonicum initiated by the infection of plant root hair cells. Fewer than 20 plant genes involved in the nodulation process have been functionally characterized. The University of Missouri scientists, with collaborators at the University of Illinois, first utilized first generation soybean microarrays to characterize gene expression in roots of B. japonicum inoculated and mock inoculated wild type and supernodulation lines. However, this approach failed to identify a large number of regulated genes, likely due to tissue dilution due to the large size of soybean roots. Therefore, the transcriptional response of individual soybean root hair cells to B. japonicum inoculation was studied with three different technologies: microarray hybridization, Illumina sequencing, and quantitative real-time reverse transcription-polymerase chain reaction. A total of 1,973 soybean genes were differentially expressed with high significance during root hair infection, including orthologs of previously characterized root hair infection-related genes such as NFR5 and NIN. The regulation of 60 genes was confirmed by quantitative real-time reverse transcription-polymerase chain reaction. Analysis also highlighted changes in the expression pattern of some homeologous and tandemly duplicated soybean genes, supporting their rapid specialization. As an aid to the qRT-PCR experiments, they surveyed the broad array of soybean microarray and other expression data to identify four genes, whose constitutive expression can be used to normalize qRT-PCR expression data.

Laurent Brechenmacher, Moon-Young Kim, Jijun Zou, Marisol Benitez, Min Li, Crystal B. McAlvin, Trupti Joshi, Bernarda Calla, Mei Phing Lee, Reena Philip, Marc Libault, Lila O. Vodkin, Dong Xu, Suk-Ha Lee, Steven J. Clough, Gary Stacey. 2008. Transcription profiling of soybean supernodulation by *Bradyrhizobium japonicum.* Mol. Plant-Microbe Int. 21: 631-645

Libault M, Farmer A , Brechenmacher L, Drnevich J, Langley RJ, Bilgin DD, Radwan O, Neece DJ, Clough SJ, May GD , Stacey G. 2009. *Complete Transcriptome of the Soybean Root Hair Cell, a Single-Cell Model, and Its Alteration in Response to Bradyrhizobium japonicum Infection* Plant Physiol. 2010 Feb;152(2):541-52.

Bilgin, D.D., Delucia, E.H., Clough, S.J. 2009. A Robust Plant RNA Isolation Method for Affymetrix Genechip® Analysis and Quantitative Real-Time RT-PCR. Nature Protocols. 4:333-340.

Marc Libault, Sandra Thibivilliers, Osman Radman, Steven J. Clough and Gary Stacey. 2008. Identification of four soybean reference genes for gene expression normalization. Plant Genome. 1:44-54

- University of Missouri scientists conducted a large scale analysis of the transcriptional activity of over 5000 soybean transcription factors that they identified from the soybean genome sequence. This work involved a large library of qRT-PCR primers designed to the transcription factor genes, as well as high throughput Illumina Solexa sequencing. This paper identified a number of interesting transcription factors showing high tissue specificity.

  Marc Libault, Trupti Joshi, Kaori Takahashi, Andrea Hurley-Sommer, Kari Puricelli, Sean Blake, Dong Xu, Henry Nguyen, and Gary Stacey. 2009. Large scale analysis of soybean regulatory gene expression identifies a Myb gene involved in nodule development. Plant Physiol. 151: 1207-1220.

- University of Missouri scientist conducted a global survey of small RNAs (e.g., miRNAs) in soybean by using high throughput sequencing methods with small RNA libraries obtained from a number of different soybean tissues. This work identified a number of novel, soybean specific miRNAs.

  Trupti Joshi, Zhe Yan, Marc Libault, Hoon Jeong, Sunhee Park, Pamela J. Green, D Janine Sherrier, Andrew Farmer, Greg May, Blake Meyers, Dong Xu, Gary Stacey. 2009. Prediction of novel miRNAs and associated target genes in *Glycine max* BMC Bioinformatics **11**(Suppl 1):S14

- University of Missouri scientists conducted a global survey of the soybean transcriptome using high throughput sequencing methods to describe the transcriptome of 14 different soybean tissues and/or conditions. This large scale survey provides transcriptional support for over 55,000 genes in the soybean genome; thereby, expanding the number of strongly supported genes, as well as identifying genes current unannotated in the current release of the genome. This paper also provides a comparison of the soybean gene index to that of ***M. truncatula*** and ***L. japonicus***, as well as presenting the results in the context of genome comparisons among the three species.

  http://digbio.missouri.edu/soybean_atlas/

  Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, He J, Xu D, May G, Stacey G. 2010. An integrated transcriptome atlas of the crop model *Glycine max* and its use in comparative analysis in plants. Plant J. (in press).

- ARS scientists at St. Paul MN and Ames IA, Iowa State University, University of Illinois, NCGR, University of Missouri, and University of Nebraska explored molecular mechanisms that influence soybean seed composition. Candidate genes at the major soybean protein quantitative trait locus at Linkage Group I were identified using near-isogenic lines that differ in seed protein and oil. Genome-wide transcript profiles were obtained from these lines using GeneChip (R) and high-throughput sequencing technologies. Variations between the profiles presumably reflect changes related differences in seed protein and oil. These results demonstrate the power of transcriptome analyses to contrast near-isogenic lines and illustrate how gene profiling technologies may aid in the annotation of genes in the soybean genome. Further investigation may provide new insight into the genes and pathways involved in protein and oil accumulation in the soybean seed.

  Bolon, Y.E., B. Joseph, S.B. Cannon, B. Diers, A. Farmer, M.A. Graham, G. May, G. Muehlbauer, J. Specht, Z. Tu, N. Weeks, W. Xu, R.C. Shoemaker, C.P. Vance. 2008. Genomic Studies in Soybean: Toward Understanding Seed Oil and Protein Production [abstract]. IV International Conference on Legume Genomics and Genetics, December 7-12, 2008, Puerto Vallarta, Mexico. Abstract No. W28. p. 101 and Abstract No. P57, p. 63.

- ARS scientists at Urbana IL published a report on transcriptome changes in soybean stem tissue challenged with Sclerotinia sclerotiorum based on cDNA microarray analysis. This study examines the differential expression of small RNA (miRNAs and siRNAs) and gene transcripts using the Illumina sequence-by-synthesis `Genome Analyzer" technology (www.illumina.com). The sequencing yielded approximately 445,000 unique reads, being analyzed with soybean (www.phytozome.net/soybean) and S. sclerotiorum (www.broad.mit.edu/annotation/genome/sclerotinia_sclerotiorum/) genome sequence databases to determine plant and fungal origins and to narrow the lists of interest. In addition to complementing and expanding our previous microarray-based mRNA expression study, this new analysis also provides in-depth transcript data from the pathogen and insight into the possible role of small RNA as a mechanism by which host and pathogen genes are regulated during the infection process. The role of small RNA in the regulation of genes in plants under attack by diverse pathogens is beginning to be documented. Although fungi have been shown to have small RNA processing and RNA interfering abilities, little is known of the possible role of small RNA in plant pathogenic fungi during the infection process. These studies used the partially resistant plant introduction PI 194639 and S. sclerotiorum strain 105HT.

  Calla, B., K.K. Varala, H. Win, M.E. Hudson, L.O. Vodkin, S.J. Clough. 2009. Preliminary Analysis of High-Throughput Expression Data and Small RNA in Soybean Stem Tissue Infected with Sclerotinia sclerotiorum [abstract]. Proceedings of the International Sclerotinia Workshop. May 31-June 4, 2009, Wilmington, NC. P. 6.

- ARS scientists at Beltsville, MD with funding from United Soybean Board identified genes expressed in soybean during infection with soybean cyst nematodes. Using laser capture microdissection, the scientists collected soybean cells that were specific to the nematode feeding site in the root of the soybean cultivar 'Peking' that exhibited a resistant reaction with one nematode race and a susceptible reaction with a different nematode race. Genes were identified that were expressed by soybean roots in each of these samples from the different feeding sites using Affymetrix microarrays. Analysis of gene expression identified differentially expressed genes belonging to specific pathways related to plant defense, including a pathway leading to cell wall synthesis. Some of these genes may be targets for modification to control soybean cyst nematode development. These data are of interest to scientists seeking new methods to broaden resistance of soybean to soybean cyst nematodes.

  Klink, V.P., C.C. Overall, N.W. Alkharouf, M.H. MacDonald, B.F. Matthews. 2007. Laser capture microdissection (LCM) and comparative microarray expression analysis of syncytial cells isolated from incompatible and compatible soybean roots infected by soybean cyst nematode (*Heterodera glycines*). Planta 226:1389-1409.

  Klink, V.P., B.F. Matthews. 2008. The use of laser capture microdissection to study the infection of *Glycine max* (soybean) by *Heterodera glycines* (soybean cyst nematode). Plant Signaling and Behavior 3 (2):105-107.

  Rashed, N., M. MacDonald, B.F. Matthews. 2008. Protease inhibitor expression in soybean roots exhibiting susceptible and resistant interactions with soybean cyst nematode. Journal of Nematology 40:138-146.

  Klink, V.P., P. Hosseini, M.H. MacDonald, N.W. Alkharouf, V. Martins, B.F. Matthews. 2009. Population-specific gene expression in the pathogenic nematode *Hederodera glycines* exists prior to infection and during the onset of a resistant or susceptible reaction in the roots of *Glycine max.* BCM Genomics. http://www.biomedcentral.com/1471-2164/10/111.

  Klink V.P., K-H. Kim, V.E. Martins, M.H. MacDonald, H.S. Beard, .N.W. Alkharouf, s-C. Park, B.F. Matthews. 2009. A correlation between host-mediated expression of parasite genes as

tandem inverted repeats and abrogation of the formation of female *Heterodera glycines* cysts during infection of *Glycine max.* Planta 230:53-71.

Klink, V.P., B.F. Matthews. 2009. Emerging approaches to broaden resistance of soybean to soybean cyst nematode as supported by gene expression studies. Plant Physiology 151:1017-1022.

Klink V.P., P. Hosseine, P. Matsye, N.W. Alkharouf, B.F. Matthews. 2009. A gene expression analysis of syncytia laser microdissected from the roots of the *Glycine mzx* (soybean) genotype PI 548403 undergoing a resistant reaction after infection by *Heterodera glycines* (soybean cyst nematode). Plant Molecular Biology. 71:525-567.

Klink, V.P., P. Hosseine, P. Matsye, N.W. Alkharouf, B.F. Matthews. 2010. Syncytium gene expression in *Glycine max* [PI88788] roots undergoing a resistant reaction to the parasitic nematode *Heterodera glycines*. Plant Physiology and Biochemistry 48:176-193.

- ARS scientists at Beltsville, MD with funding from United Soybean Board examined the expression of soybean genes induced in the leaf upon infection by the fungus *Phakopsora pachyrhizi* Sydow, an exotic pathogen in the U.S that causes soybean rust. Using laser capture microdissection, the scientists isolated uredinia formed by *P. pachyrhizi* on the under side of leaves, extracted the RNA and constructed a cDNA library. A total of 925 cDNAs were identified in library using Illumina Solexa deep-sequencing techniques. DNA homology searches were conducted to determine the identity of the genes. Sixteen genes had similarities to known genes encoding proteins involved in different cellular functions such as energy production, cellular communication/signal transduction, and transcription. RT-PCR was used to determine the expression pattern of seventeen rust genes. Some of these genes may be of use to improve resistance of soybean to soybean rust.

  Choi, J.J., N.W. Alkharouf, K.T. Schneider, B.F. Matthews, R.D. Frederick. 2008. An enriched cDNA library and microarray analysis reveal expression patterns in soybean resistant to soybean rust. Functional and Integrative Genomics 8:341-351.

  Tremblay A., S. Li, B.E. Scheffler, B.F. Matthews. 2009. Laser capture microdissection (LCM) and expressed sequence tag analysis of uredinia formed by, *Phakopsora pachyrhizi,* the causal agent of. Asian soybean rust Physiological and Molecular Plant Pathology. 73:163-174.

  Tremblay, A, B.F. Matthews. 2009. Laser capture microdissection, a new technique to collect specific cells. Food and Agriculture Magazine 4:4.

- Scientists at the University of Illinois used high throughput Illumina sequencing of the small RNA transcriptome as well as sequencing of BACs to elucidate a novel naturally-occurring tissue-specific silencing mechanism by siRNAs (short interfering RNAs) that target members of a gene family, including those with diverged sequences located on other chromosomes. The phenomenon results in inhibition of a metabolic pathway in one tissue while the pathway is expressed in other organs. More specifically, the work documents a subset of small RNAs that target the chalcone synthase (CHS) gene family. The endogenous *I* locus that encodes CHS genes also spawns an aberrant double-stranded RNA that triggers the production of CHS siRNA. In turn, these CHS siRNAs target the CHS mRNAs (messenger RNAs) and causes their degradation, thus preventing the mRNAs from producing the CHS enzyme and subsequently the anthocyanin pigments, resulting in the colorless soybean seed. The dominant alleles specifying yellow seed coat have been incorporated by breeders into the germplasm of all modern cultivated soybean varieties long before the mechanism of the locus was understood to be mediated by tissue specific production of siRNAs.

Tuteja, J. and Vodkin, L.O. 2008. Structural features of the endogenous *CHS* silencing and target loci in the soybean genome. Plant Genome 48: 49-69.

Tuteja, J.H., Zabala, G., Varala, K., Hudson, M., and Vodkin, L.O. 2009. Endogenous, tissue-specific short interfering RNAs silence the chalcone synthase gene family in Glycine max seed coats. Plant Cell 21:3063-3077.

- To understand gene expression networks leading to functional properties of the soybean seed, researchers at the University of Illinois used microarrays to perform a detailed examination of soybean seed development during the stages of major accumulation of oils, proteins, and starches, as well as the desiccating and mature stages. Of particular interest are the genes found to peak in expression at the desiccating and dry seed stages, such as those annotated as transcription factors, which may indicate the preparation of pathways that will be needed later in the early stages of imbibition and germination. The researchers also examined the expression profiles found in the soybean cotyledon during seed germination when that organ transitions from accumulation of nutrient reserves to one whose main function is to serve as a nutrient reserve that supplies the needs of the young plant throughout seedling development. During this process the cotyledons experience a functional transition to a mainly photosynthetic tissue. Expression profiles of individual gene family members, enzymatic isoforms and protein subunits were classified according to their involvement in different functional activities relevant to seedling development and the functional transition of the cotyledon in soybean, especially the ones associated with the glyoxylate cycle.

Gonzalez, D. O. and Vodkin, L.O. 2007. Specific elements of the glyoxylate pathway play a significant role in the functional transition of the soybean cotyledon during seedling development. BMC Genomics 8: 468 (22 page online journal)

Jones, S.I., Gonzalez, D.O., and Vodkin, L.O. 2010. Flux of transcript patterns during soybean seed development. BMC Genomics 11:136.

- Scientists at the University of Illinois used microarrays and other genome resources to determine the global transcript profiles to elucidate plant defenses to the soybean aphid as well as response to the herbicide glyphosate and to identify the molecular basis of the soybean *W1*, purple flower locus.

Li, Y., Zou, J., Li, M., Bilgin, D., Vodkin, L.O., Hartman., G. L., and Clough, S. J. 2008. Soybean defenses against to the soybean aphid. New Phytologist 179: 185-195.

Zhu, J., Patzoldt, W., Shealy, R., Vodkin, L., Clough, S., Tranel, P. 2008. Transcriptome response to glyphosate in sensitive and resistant soybean. J. Agri Food Chem. 56:6355-6363.

Zabala, G. and Vodkin, L. O. 2007. A rearrangement resulting in small tandem repeats in the F3'5'H gene of white flower genotypes is associated with the soybean *W1* locus. Plant Genome 47: 113-124.

B.3.b - Proteomic Approaches

Transcriptomics involves analysis of gene transcription. There is need to develop developmental stage catalogs of mRNA expression for key plant organs and biological processes, and improved ways of relating proteomics, transcription products, to the annotated soybean genome sequence.

- ARS scientists at Beltsville MD with researchers at the University of Missouri identified plant genes and proteins associated with disease resistance to rust fungi. Discovery of proteins that are important for disease resistance to rust fungi may help improve cultivars through breeding or transgenic technology. Comparison between common bean plants that are naturally resistant and plants naturally susceptible revealed 1,500 proteins that contribute

to the resistance response. At least 90 of the common bean proteins can be found as versions in soybean and help provide soybeans with resistance to the soybean rust fungus. Similar types of genes and proteins have also been identified in the model plant Arabidopsis thaliana. This association demonstrates that the defense systems between plants are very similar. These results helped to identify disease resistance proteins that might eventually be used to protect susceptible plants.

Lee, J., Campbell, K., Scheffler, B.E., Feng, J., Naiman, D.Q., Garrett, W.M., Thibivilliers, S., Stacey, G., Tucker, M.L., Pastor Corrales, M.A., Cooper, B. 2008. Quantitative Proteomic Analysis of Bean Plants Infected by a Virulent and Avirulent Obligate Rust Fungus. Molecular and Cellular Proteomics. 8:19-31.

Tucker, M.L., Puthoff, D.P., Neelam, A., Ehrenfried, M.L., Scheffler, B.E., Ballard, L.L., Campbell, K.B., Cooper, B. 2008. Analysis of expressed sequence tags from Uromyces appendiculatus hyphae and haustoria and their comparison to sequences from other rust fungi. Phytopathology. 98:1126-1135.

Thibivilliers, S., Joshi, T., Campbell, K., Scheffler, B., Borerma, R., Xu, D., Cooper, B., Nguyen, H.T., Stacey, G. 2009. Generation of Phaseolus vulgaris ESTs and investigation of the regulation upon Uromyces appendiculatus infection. Biomed Central (BMC) Plant Biology. 9:46.

- Soybean roots have millions of single cell hairs on them that increase surface area and enhance water uptake. These root hairs are also the sites of Rhizobium bacterial infection that allows the plant to produce its own nitrogen and live without added nitrogen fertilizer. It is believed that a set of proteins makes root hairs unique from other parts of the plant such as a leaf, but it is unknown what these unique proteins are. Experiments identified proteins from single cell root hairs. Some of these proteins may serve as water channels and appear to be unique to root hairs. A much more comprehensive study to the one described above has now been completed and a manuscript is in preparation.

Brechenmacher, L., Lee, J., Sachdev, S., Song, Z., Nguyen, T., Joshi, T., Oehrle, N., Libault, M., Mooney, B., Xu, D., Cooper, B., Stacey, G. 2009. Establishment of a Protein Reference Map for Soybean Root Hair Cells. Plant Physiology. 149:670-68.

- Soybean cyst nematode causes approximately $1 billion in damages each year in the U.S. ARS scientists at Beltsville MD identified genes whose expression significantly increases in cells where the nematode is feeding and compared these genes in resistant and susceptible soybean varieties. Some of these genes may be useful to broaden resistance of soybean to the nematode in genetically transformed plants.

Nahed, R., Macdonald, M.H., Matthews, B.F. 2008. Protease inhibitor expression in soybean roots exhibiting susceptible and resistance reactions to soybean cyst nematode. Journal of Nematology. 40:138-146.

Klink, V.P., Hosseini, P., Macdonald, M.H., Alkharouf, N.W., Matthews, B.F. 2009. Population-specific gene expression in the pathogenic nematode Hederodera glycines exists prior to infection and during the onset of a resistant or susceptible reaction in the roots of Glycine max. Biomed Central (BMC) Genomics. 10:111. http://www.biomedcentral.com/1471-2164/10/111.

- Scientists at Southern Illinois University revealed the SCN pathosystem's proteome showed a very focused response among a small number of proteins.

Afzal AJ, A Natarajan, N Saini, M J Iqbal, MA Geisler, H El Shemy, R Mungur, L Willmitzer and DA Lightfoot. 2009. The nematode resistance allele at the rhg1 locus alters the proteome and metabolome of soybean roots. Plant Physiology 151: 1264-1280

B.3.c - Metabolomic Approaches

Soybean biochemical processes give rise to a plethora of metabolites, each with its own spectrum of activity. The biological regulation of a great majority of these compounds is unknown.

- Soybeans grown in the Upper Midwest typically have lower protein content than those grown in warmer regions of the U.S. Soybean near inbred-lines (NILs) were developed that differ at only a single quantitative locus (LGI). ARS scientists at St. Paul MN, in collaboration with several locations, did high-throughput RNA sequencing to determine genetic differences in the two lines. 100 million bp of cDNA were sequenced from seed of the NILs. In parallel, transcript profiles were obtained from multiple stages of seed fill in each NIL, and transcript accumulation changes were characterized. Using Affymetrix (R) Soy GeneChip microarray and high-throughput Illumina (R) transcriptome sequencing platforms, 13 genes displaying significant transcript accumulation differences between the NILs were found and mapped to the 8.4 Mbp LG I protein QTL region. This study demonstrates the power of complementary approaches to characterize contrasting NILs and provides genome-wide transcriptome data towards the improvement of gene annotation in the soybean genome. These genes may be useful in regulating soybean seed oil and protein

  Bolon, Y-T, Joseph, B, Cannon, S., Graham, M.; Diers, B.; Farmer, A.; May, G.; Muehlbauer, G.; Specht, J.; Tu, Z.-J., Weeks, N.; Xu, W.; Shoemaker, R.; and Vance, C. 2010. Complementary Approaches Characterize the Linkage Group I Seed Protein QTL in Soybean. Biomed Central (BMC) Plant Biology, Accepted: March 3, 2010

- USDA, ARS scientists at Beltsville MD with support from the United Soybean Board, used the Universal Soybean Linkage Panel 1.0 (USLP 1.0) to genotype 306 RILs, F1s and parents of the cross of Jake X PI 416908 which segregated for oleic acid. This population was received from a collaborator at the University of Missouri. A total of 515 of the 1536 SNPs were polymorphic in this population. DNA of a total of 576 RILs and parents from the cross of N98-4445A X N97-3225 which segregated for oleic and palmitic acid. from a collaborator at North Carolina State University, was analyzed with the USLP 1.0.

- University of Missouri scientists, in conjunction with collaborators at the Noble Foundation, have completed a global metabolome study of soybean root hairs uninoculated and inoculated with the nitrogen fixing symbiont, *Bradyrhizobium japonicum*.

  Laurent Brechenmacher, Zhentian Lei, Marc Libault, Seth Findley, Masayuki Sugawara, Michael J Sadowsky, Lloyd W Sumner, and Gary Stacey. 2010. Soybean metabolites regulated in root hairs in response to the symbiotic bacterium *Bradyrhizobium japonicum.* Plant Physiol. (submitted)

- Profiling soybean gene products will lay the foundation for a systems biology approach to key processes such as seed development, which will lead to the genetic improvement of yield and seed composition. Different biotic stresses such as fungal diseases and abiotic stresses such as drought and flooding also elicit phenotypic responses from the genome. Thus, by studying the gene products, a direct correlation between response and specific peptides, and/or metabolites can be made. This will lead to crop improvement either through breeding or transgenic efforts. Also, these approaches help in the search of novel bioactive compounds leading to the production of new generation soybean products for human health and nutrition. University of Missouri scientists are in the process of constructing soybean proteome and metabolome maps and there by contribute towards understanding of the biochemical network involved in seed development and stress responses. The University of Missouri scientists are in the process of completing the global metabolite profiling of soybean seeds and root tissues. A manuscript is in preparation describing the seed metabolites changes during seed

development is soybean. Also, the University of Missouri scientists, in conjunction with collaborators at the Noble Foundation, have completed a global metabolome study of soybean root hairs uninoculated and inoculated with the nitrogen fixing symbiont, *Bradyrhizobium japonicum*. A manuscript describing this work is now under review.

Laurent Brechenmacher, Zhentian Lei, Marc Libault, Seth Findley, Masayuki Sugawara, Michael J Sadowsky, Lloyd W Sumner, and Gary Stacey (2010) Soybean metabolites regulated in root hairs in response to the symbiotic bacterium *Bradyrhizobium japonicum.* Plant Physiol. (submitted)

Valliyodan B, Xu D, Stacey G, Nguyen HT (2010) LTQ-Orbitrap LC-MS analysis of metabolite changes during seed development in soybean. (Manuscript draft is in progress)

- The bioinformatics group at the University of Missouri has created a soybean metabolite data base SoyMetDB populated with metabolite information from various soybean soybean tissues.
- University of Kentucky scientists characterized the expression of triacylglycerol biosynthetic enzymes relative to vegetable oil accumulation in developing seeds of soybean, Arabidopsis and plant species with high epoxy and hyroxy fatty acids.

Li, R., K. Yu, T. Hatanaka, and D.F. Hildebrand. 2010. *Vernonia* DGATs increase accumulation of epoxy fatty acids in oil. Plant Biotechnology Journal 8:184-195.

Li, R., K. Yu, and D. Hildebrand. 2010. DGAT1, DGAT2 and PDAT Expression in Seeds and Other Tissues of Epoxy and Hydroxy Fatty Acid Accumulating Plants. Lipids 45:145-157. http://www.springerlink.com/content/0j8532qt663t6213/fulltext.html

B.5 – Breeder Perspectives on Gene Function

Breeders select for gene combinations through crossing and selection among resulting progeny. Functional genomics has the potential to provide breeders information on what genes control economically important traits and therefore should be selected. This information could increase the power of selection; however, identifying genes that control complex traits like yield will be especially difficult. Identifying these genes will require collaborative efforts between breeders and molecular geneticists. In this collaboration, the breeders will need to identify the map positions of target genes and develop the unique germplasm that is needed to identify the genes.

B.5.a - Develop recurrent parent/near-isogenic line pairs for important QTL to use in microarray analysis.

- ARS scientists at Wooster OH have released more than 30 high-yielding non-GMO soybean cultivars with improved insect and disease resistance over last 30 years and continue to develop top performing cultivars. Stressland, Apex, Croton3.9, Stalwart, Wooster, and Prohio are few of the latest releases from this program. We also released germplasm with multiple pest (e.g., Phytopthora rot and beetle) resistance to be used by public and private soybean breeders in USA. DNA technology is used to pyramid multiple disease and insect resistance in high-yielding and value added (e.g., high protein) elite soybean cultivars. On-going Research projects include: 1) Soybean Aphid resistant cultivars for Ohio were released in 2009. 2) Phyptophthora Stem and Root Rot. Collaboration with Dr. Anne Dorrance is producing (a) soybean cultivars/gemplasm with multiple P. sojae resistant genes, and (b) QTL maps for partial resistance to P. sojae. 3) Beetle resistant soybeans: Bean leaf beetle, Mexican bean beetle, and Japanese beetle resistant soybeans lines were released in 2009. 4) Bean Pod Mottle Virus (BPMV): A source of tolerance to BPMV was identified, QTL for tolerance were mapped, and markers were used to develop soybeans with BPMV tolerance.

5) Genes have been identified that increase yield and seed protein (>46%). 6) Frogeye leaf spot (FLS): marker assisted breeding is used to develop soybeans resistant to FLS.

Mian, R.M., Kang, S., Beil, S.E. 2008. Genetic Linkage Mapping of the Soybean Aphid Resistance Gene in PI 243540. Journal of Theoretical and Applied Genetics. 117:955-962.

Mian, R.M., Kang, S., Redinbaugh, M.G. 2009. Microsatellite Diversity of Soybean Genotypes Differing in Bean Pod Mottle Virus Leaf Symptom. Genetic Resources and Crop Evolution. 89:359-67.

Mian, R.M., Bond, J., Joobeur, T., Mengistu, A., Weibold, W., Grover, S., Wrather, A. 2009. Identification of Soybean Genotypes Resistant to Cercospora sojina by Field Screening and Molecular Markers. Plant Disease. 93:408-411.

- Soybean cyst nematode (SCN, *Heterodera glycines* Ichinohe) is the most destructive pest of soybean in the United States and worldwide. Host plant resistance is an effective approach to control this pest; however, the continuation of growing the same resistant cultivar(s) could result in SCN population shifts and loss of SCN resistance. University of Missouri soybean scientists have utilized exotic soybean accessions with broad-based resistance to soybean nematode (SCN), such as PI 437654, PI 438489B, and PI 567516C, to develop recombinant inbred lines (RIL) mapping populations. Using simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers (USLP 1.0), high resolution genetic linage maps, spanning over 95% of the soybean genome have been constructed. In addition to the confirmation of quantitative trait loci (QTL) previously reported and mapped on chromosomes (Chrs.) 8, 11, 18 (LGs A2, B1, G, respectively), QTL associated with resistance to multiple SCN HG types on Chrs. 10, 18, and 20 (LGs O, G 2nd locus, and I, respectively) were detected and mapped. Using the Solexa next-generation sequencing technology, four soybean genotypes have been sequenced yielding thousands of SNP and SSR markers. These markers were employed for genetic mapping. With molecular markers closely associated with and flanking to the target QTL regions, marker-assisted backcrossing (MAB) program was initiated to develop several near-isogenic lines (NILs). As of May 2010, BC2F1's have been produced. These NILs will be employed for fine-mapping and map-based cloning in near future. In addition, more than 30 RIL mapping populations have been developed with diverse PIs having broad-based resistance to SCN and other nematodes to identify resistant genes for further QTL analysis and NIL development.

Xiaolei Wu, Sean Blake, David A. Sleper, J. Grover Shannon, Perry Cregan, Henry T. Nguyen (2009) QTL, additive and epistatic effects for SCN resistance in PI 437654. *TAG* 118:1093

Vuong, T.D., D.A. Sleper, J.G. Shannon, and H.T. Nguyen. 2010. Novel quantitative trait loci for broad-based resistance to soybean cyst nematode (*Heterodera glycines* Ichinohe) in soybean PI 567516C. Theor Appl Genet (*in revision)*.

B.5.b - Develop random-mated population(s) utilizing a black-seeded male sterile system that will result in low levels of linkage disequilibrium for use in RNA-based Bulk Segregant Analysis (bulks are derived from phenotypes).

- QTL were associated with enhanced yield in *Glycine max.* Populations were created by crossing a black seed coat *Glycine max* PI290136 parent plant or progeny thereof with a yellow seed coat *Glycine max* parent plant; and screening the segregating population for the presence of a DNA molecular marker for the enhanced yield allele (SY5) derived from the Glycine max PI290136 plant on linkage group U03. Yellow seeded lines with enhanced yield relative to the yellow seed coat Glycine max parent plant were selected.

Concibidio, V.C., Delannay, X., Soybean Plants with Enhanced Yields and Methods for Breeding for and Screening of Soybean Plants with Enhanced Yields. WO/2002/063020

C – Germplasm Genomics

C.1 – Association Mapping

Association or linkage disequilibrium (LD) mapping has been proposed as a rapid method for the discovery of genes/quantitative trait loci (QTL) in germplasm with existing phenotypic data without the need to create and phenotype F_2 progeny, recombinant inbred lines or backcross populations. An Association Mapping Panel is needed to implement LD mapping in soybean

- ARS scientists at Beltsville MD reported the first characterization of the structure of linkage disequilibrium (LD) in soybean and determined that LD varied widely between three genome regions and between different soybean populations. LD is the non-random association of alleles in the genome and is the basis for using a new QTL mapping technique called association analysis. The estimates of LD have shown that the resolution of association analysis could be 10 times more precise than linkage analysis which is the current method used for mapping QTL. These estimates of LD suggest between 13,700 and 75,600 SNP DNA markers would be required to define haplotype variation in the euchromatic portion of the soybean genome via association analysis using whole genome scans. This work involved the discovery and genotyping of 738 SNPs located in three chromosomal regions in 120 soybean genotypes. These 120 soybeans included 26 accessions of the wild soybean as well as cultivated soybean germplasm and elite cultivars and thus included representatives of the major sources of germplasm for soybean improvement. LD in soybean has been used by geneticists to apply association analysis to find quantitative trait loci (QTL) for traits such as oil seed concentration, protein seed concentration and iron chlorosis deficiency.

  Hyten, D.L., Choi, I-Y., Song, Q., Shoemaker, R.C., Nelson, R.L., Costa, J.M., Specht, J.E., and Cregan, P.B. Highly variable patterns of linkage disequilibrium in multiple soybean populations. Genetics 175:1937-1944. 2007.

C.2 – Track Breeding-Induced Genomic Change

It is hypothesized that the documented increases in soybean yield from 1) the original introductions to 2) the publicly developed cultivars from hybridization and selection to 3) the elite cultivars developed after 1980 can be associated with SNP-based allele and haplotype variation in the three successive time-differentiated germplasm pools.

- ARS scientists at Beltsville MD provided the first global estimate of the frequency of single nucleotide polymorphisms (SNPs) in the available soybean germplasm collection. This assessment of DNA sequence variation was used to measure genetic variability in soybean and to characterize the effects of domestication and plant breeding on genome diversity thus dispelling the long held assumptions about the causes of limited genetic variation in North American soybean cultivars. This work revealed that DNA sequence diversity in the wild soybean was low as compared to other wild species and that 50% of the variation was lost during domestication. It was determined that the small group of Asian ancestors of North American soybeans contained 87% of the diversity of the larger pool of “Asian landraces” and that 60 years of breeding in North America has had only a minor impact on the genetic variability of the soybean crop. Sequence diversity was determined in 102 genes in 120 soybeans including the wild soybean from which soybean was domesticated as well as modern soybean cultivars.

Hyten D., Song Q., Zhu Y., Choi I-Y., Nelson R., Costa J., Specht J., Shoemaker R., and Cregan P. Impacts of genetic bottlenecks on soybean genome diversity. Proc. Natl. Acad. Sci. U.S.A. 103:16666-16671. 2006.

Yoon M-S., Song Q., Choi I-Y., Specht J., Hyten D., and Cregan P. BARCSoySNP23: A panel of 23 selected SNPs for soybean cultivar identification. Theor. Appl. Genet. 114:885-899. 2007.

Choi, I.-Y., Hyten, D.L., Matukumalli, L.K., Song, Q., Chaky, J.M., Quigley, C.V., Chase, K., Lark, K.G., Reiter, R.S., Yoon, M-S., Hwang, E-Y., Yi, S-In., Young, N.D., Shoemaker, R.C., Van Tassell, C.P., Specht, J.E., and Cregan, P.B. A soybean transcript map: gene distribution, haplotype and SNP analysis. Genetics 176:685-696. 2007.

Hyten, D.L., J.R. Smith, R.D. Frederick, M.L. Tucker, Q. Song, and P.B. Cregan. Bulked segregant analysis using the GoldenGate assay to locate the Rpp3 locus that confers resistance to Phakopsora pachyrhizi (soybean rust) in soybean. Crop Sci. 49: 265-271. 2009.

C.3 – Mining Yield QTLs in Exotic Germplasm

More than 85% of the genes present in the commodity type soybeans grown by U.S. and Canadian farmers can be traced to 17 soybean Ancestral Cultivars from Asia. The USDA Soybean Germplasm Collection consists of more than 19,000 soybean Plant Introductions collected from Asia and other countries which likely contain allelic variants that impact yield. Thus, methods must be developed to 1) identify germplasm lines that harbor unique genetic factors that can positively impact yield and 2) develop procedures to facilitate use of yield-enhancing genetic factors by U.S. soybean breeders.

C.3.a - Genomic Analyses of Four Yield QTLs Derived from Exotic Germplasm Four yield QTL, two in Northern germplasm and two in Southern germplasm, have been identified by soybean researchers. One approach to discovery of unique yield QTL will be to further characterize DNA sequence variation associated with these QTLs with the goal of identifying candidate genes for these QTLs. If successful, this research would likely facilitate subsequent discovery of unique yield genes/QTLs from exotic germplasm.

- ARS scientists at Urbana IL developed 23 heterogeneous inbred populations from exotic germplasm and selected for yield equivalence in maturity groups II, III, and IV. These experimental populations were derived from 31 different plant introductions that originated from 8 countries and were introduced into the U.S. between 1924 and 1994.The pedigrees of these lines ranged from 13 to 75% exotic germplasm and some lines had pedigrees that contained as many as 9 different exotic lines. Seeds of these populations were made available to all public soybean breeders for planting in 2009 to select inbred lines adapted to their specific locations. A population of 166 lines from the cross of NE 3303 x LG97-9301 (C3 population) was planted in yield tests at 4 locations to identify yield QTL. LG97-9301 has a pedigree that is 75% exotic germplasm including PI 253665D, PI 283331, and PI 391594. The Y9IVC population, which has BC2F3-derived lines in the background of the high yielding line LD00-3309 segregates for two yield QTL from PI 471938 and one yield QTL from Hutcheson. To confirm previously mapped QTL, three new confirmation populations were planted for their first year of yield testing. The first population was developed from an individual plant that segregated for a yield QTL previously mapped onto linkage group J in the A3 population. The A3 population was derived from the cross of IA3023 x LG99-11620, which is 50% exotic germplasm. The second and third populations were developed to confirm yield QTL on linkage groups C1 and A1 previously mapped from the B3 population from the cross of U98-311442 x LG00-8299, which is also 50% exotic germplasm. These three populations were each planted in two locations. Seven cooperative tests are conducted annually with soybean breeders in commercial companies to test experimental lines derived

from exotic germplasm in maturity groups II (65 entries); maturity group III (111 entries) and maturity group IV (75 entries).

C.3.b. - Identify/Confirm Additional Yield QTL Alleles from Exotic Germplasm. Continue the breeder-based "pre-breeding" efforts aimed at using exotic germplasm to detect favorable alleles that enhance yield in elite germplasm over a broad range of production environments.

- From 1970 to 2008 there were 2,242 soybean cultivars registered in North America through U.S. PVP, U.S. utility patent, and journal registration. Of these, 80% were developed through proprietary and 20% through public programs. The most frequently used germplasm for cultivar development were the cultivars `Williams" (parent used in last cross prior to inbreeding in 70 cultivars), `A3127" (63), `Essex" (45), `Amsoy" (38), `Corsoy" (33), `Wayne" (30), `Forrest" (27), `Hutcheson" (25), `MO13404" (23), and `Bedford" (23). Genetic Diversity (1 " coefficient of parentage), estimated from pedigree lineage, was 0.89 overall. Genetic diversity was the same within public (0.89) and proprietary (0.89) developed cultivars. The cultivar A3127 is a major progenitor of recently developed proprietary cultivars registered from 1999 to 2008. Of these 494 cultivars, half have a genetic contribution of at least 10% from A3127. New cultivars were predominately developed from the following crosses: 2-parent (70% of cultivars developed), complex (12%), 3-parent (5%), one backcross (5%), multiple (2, 3, or 4) backcrosses (3%), and backcross 5 or greater (2%). Approximately 1% of the cultivars were from selections within cultivars, broad base populations and mutation selections. In comparisons where both parent and progeny were evaluated in the same environments, seed yield increased 3.2% per breeding cycle. In these comparisons, seed yield had a correlation of 0.29 with parental diversity.

  Shannon, J.G., Nelson, R.L., Lee, J.D., Wrather, J.A. 2010. Registration of LG04-6863 Soybean Germplasm Line with Diverse Pedigree. Journal of Plant Registrations 4:70-72.

- Scientists at Pioneer Hi-Bred International developed and implemented a concept for identifying effective grain yield quantitative trait loci (QTL) for marker-assisted selection (MAS) across a wide range of genetic and/or environmental contexts. Preliminary yield trials were employed to model a target genotype within each context and immediately select the progeny that approach that target genotype in real time. Elite soybean cultivars with residual heterogeneity were leveraged as populations (the genetic context) to detect yield QTL within a limited set of environments (the environmental context), to model a target genotype, and to select subline haplotypes that comprised the target genotype. The yield potential of the selected subline haplotypes were then compared to their respective mother lines in highly replicated yield trials across multiple environments and years. Statistically significant yield gains of up to 5.8% were confirmed in some of the selected sublines, and two of the improved sublines were released as improved cultivars. This context-specific MAS (CSM) approach might also be applicable to the more typical biparental and backcross populations commonly used in plant breeding programs.

  Sebastian, S.A., L. G. Streit, P. A. Stephens, J. A. Thompson, B. R. Hedges, M. A. Fabrizius, J. F. Soper, D. H. Schmidt, R. L. Kallem, M. A. Hinds, L. Feng, and J. A. Hoeck. 2010. Context-Specific Marker-Assisted Selection for Improved Grain Yield in Elite Soybean Populations. Crop Sci. 50: doi: 10.2135/cropsci2009.02.0078

C.3.c. - Evaluate Recently Released Cultivars for Favorable Yield Genes.

- ARS scientists at Beltsville MD have developed about 1200 F2 plants from each of 40-50 matings between IA3023 and high yielding elite and exotic soybean lines. At least 1000 F2 plants from each of those matings with maturity as similar as possible to the common

Maturity Group III parent IA3023 were selected via single-seed descent (SSD) and the 1000 F3 seed of each mating were increased in a winter nursery to initiate a two-generation SSD advance to produce F4 and F5 seed. DNA was harvested from a minimum of 500 F5 plants per mating and increased in a winter nursery to produce sufficient seed of 250 RILs per mating for yield testing in replicated yield trials for characterization with 1536 SNP DNA markers using the Illumina GoldenGate assay. This research is supported with funds from United Soybean Board Project #9241 entitled "Nested Association Mapping to Identify Yield QTL in Diverse High Yielding Elite Soybean Lines". A total of 50 crosses of elite and exotic soybean lines with the elite variety IA3023 were grown in the greenhouse at University of Nebraska, Lincoln. DNA was extracted from all of the F1 plants and DNA marker tests were run to confirm that all the F1 plants were hybrids. The plants were harvested and threshed and anywhere from 100-600 F2 seeds were obtained per F1 plant. The resulting F2 seeds were planted in the field at Lincoln, NE for a one generation advance during the summer of 2009. For almost every cross combination, at least 1,000 F2 seeds were planted in the field.

- ARS scientists at Beltsville MD coordinated research with the high yield reference genotype, IA3023. This line was crossed to 25 diverse accessions to create 25 populations with 200 recombinant inbred lines (RILs) each. Thus, each population shares the reference genotype as one of the parents. The 5000 RILs will be phenotyped by collaborators and genotyped with SNPs as per a typical QTL population analysis. The 25 diverse accessions will be extensively genotyped or sequenced which allows for high marker density which can be extrapolated to all the RILs in the 25 populations, thus greatly increasing the power to detect and fine map QTL. Using this approach, essentially all SNPs present in all 5000 RILs can be determined and the most common QTL among the 25 diverse lines can be discovered. Funds were provided by the United Soybean Board, as part of Project #9241 entitled "Nested Association Mapping to Identify Yield QTL in Diverse High Yielding Elite Soybean Lines". During the winter of 2008-2009 a number of F1 plants from 50 crosses of elite and exotic soybean lines, each from crosses with the elite variety IA3023 were grown in the greenhouse at both the University of Illinois and the University of Nebraska, Lincoln. These 50 crosses with the common parent IA3023 were selected from a total of 120 crosses with IA3023 based upon an analysis of the elite and exotic parents with 1364 SNP DNA markers. This analysis was conducted in order to assure that a broad spectrum of genetic diversity was represented among the crosses that are used for the development of experimental lines to be used for the identification of genetic factors that positively impact soybean yield. The selected crosses were made by soybean breeders in a number of states including Arkansas, Illinois, Indiana, Ohio, Tennessee, Missouri, and Nebraska.

C.4 – Marker-Assisted Selection Resources

While a number of public soybean breeding programs have the equipment and personnel to conduct high throughput molecular marker analysis, there are many that do not have this capacity. One possible way to meet this need is by the development of genotyping centers that would provide marker analysis for important traits of broad interest as determined by U.S. breeders.

C.4 - SoyBase

- ARS scientists at Ames IA curate SoyBase. The recent decade has seen a huge increase in the acquisition of genetic and genomic data for many crop species. Managing and visualizing this vast array of information requires a specialized and manually curated database. SoyBase, the USDA-ARS soybean genetic database, is a comprehensive repository for professionally curated genetics, genomics and related data resources for soybean. SoyBase contains the

most current genetic, physical and genomic sequence maps integrated with qualitative and quantitative traits. The quantitative trait loci (QTL) represent more than 18 years of QTL mapping of more than 100 unique traits. SoyBase also contains the well-annotated "Williams 82" genomic sequence and associated data-mining tools. The genetic and sequence views of the soybean chromosomes and the extensive data on traits and phenotypes are extensively interlinked. This allows entry to the database using almost any kind of available information, such as genetic map symbols, soybean gene names or phenotypic traits. SoyBase is the repository for controlled vocabularies for soybean growth, development and trait terms, which are also linked to the more general plant ontologies. SoyBase can be accessed at http://soybase.org.

Grant, D.M., Nelson, R., Cannon, S.B., Shoemaker, R.C. 2009. SoyBase, The USDA-ARS Soybean Genetics and Genomics Database. Nucleic Acids Research. Doi: 10.1093/nar/gkp798.

- Scientists at Southern Illinois University developed SoyGD.

Lightfoot DA (2008) Soybean Genomics: Developments through the use of cultivar Forrest. Internat J of Plant Genome. 2008:1-22. doi:10.1155/2008/793158

C.5 – Germplasm Genomics Informatics

As SNPs are mapped on a multitude of cultivars and accessions from the germplasm collection, there is a need for a central data repository that is accessible, interpretable, and relatable to informatic needs.

C.5.a - Development of HapMap Browser

- Genomic Explorer y Survey of Immune Response (GEYSIR) (Faye Schilkey, PM) is an interactive, web-based genomic visualization tool developed as part of the NIAID/deCODE population genetics project. Web-based tools for exploring genomic data typically are statically rendered HTML pages, which lack live interactivity. With the exceptional amount of data and scales of size involved in working with genomic data, this lack of dynamic interaction usually becomes not only cumbersome for the user but inadequate for scientific discovery. To address this issue in the context of population genetics studies, GEYSIR was developed to enable exploration of a wide scale of genomic data, from single nucleotide polymorphisms to gene neighborhoods to marker sets and association data spanning all chromosomes. GEYSIR is designed to be a highly interactive, dynamic, and responsive web application. Additionally, it was designed up front for extensibility and reusability so that the code base and architecture can be reused for a wide variety of genomic data, organisms, research, and data models**. See: http://www.ncgr.org/**

C.5.b - Correlation of expression data with QTL (eQTL)

- Gene expression Quantitative Trait Locus (eQTL) mapping measures the association between transcript expression and genotype in order to find genomic locations likely to regulate transcript expression. The availability of both gene expression and high-density genotype data has improved ability to perform eQTL mapping in inbred and other homozygous populations. However, existing eQTL mapping software does not scale well when the number of transcripts and markers are high. Various methods have been developed, such as FastMap, for fast and efficient eQTL mapping in homozygous inbred populations with binary allele calls. FastMap exploits the discrete nature and structure of the measured single nucleotide polymorphisms (SNPs). In particular, SNPs are organized into a Hamming distance-based tree that minimizes the number of arithmetic operations required to calculate the association of a SNP by making use of the association of its parent SNP in the tree.

FastMap's tree can be used to perform both single marker mapping and haplotype association mapping over an m-SNP window. These performance enhancements also permit permutation-based significance testing. **See: http://cebc.unc.edu/fastmap.html**

- Scientists at Southern Illinois University mapped eQTL onto the physical map and sequence using BES bridges.

  Saini, N, J.L. Shultz and D.A. Lightfoot. 2008. Re-annotation of the physical map of Glycine max for ploidy by BAC end sequence driven whole genome shotgun read assembly. BMC Genomics 9:323-940

  Karangula UB, M.A. Kassem, L. Gupta, H.A. El-Shemy and D.A. Lightfoot. 2009. Locus interactions underlie seed yield in soybeans resistant to Heterodera glycines. Curr. Issues Mol. Biol. 11 (Suppl. 1): i73-84"

C.5.c - Develop tools for identifying candidate genes underlying QTL

- ARS scientists at Columbia, MO developed and published “breeder friendly” perfect molecular marker assays for direct selection of the desired alleles controlling a number of soybean seed composition traits such as low linolenic acid, elevated oleic acid, low P34 allergen, low raffinose/stachyose, low phytic acid, and null lipoxygenase. P35 under C.4-Marker Assisted Selection Resources.

  Gillman JD, Pantalone VR, Bilyeu K: The low phytic acid phenotype in soybean line CX1834 is due to mutations in two homologs of the maize low phytic acid gene. *The Plant Genome* 2009, 2(2):179-190.

  Dierking EC, Bilyeu KD: Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. *The Plant Genome* 2008, 1(2):135-145.

  Dierking E, Bilyeu K: New sources of soybean seed meal and oil composition traits identified through TILLING. *BMC Plant Biology* 2009, 9(1):89.

  Bilyeu KD, Palavalli L, Sleper DA, Beuselinck PR: Three microsomal omega-3 fatty-acid desaturase genes contribute to soybean linolenic acid levels. *Crop Sci* 2003, 43(5):1833-1838.

  Bilyeu K, Palavalli L, Sleper DA, Beuselinck P: Molecular genetic resources for development of 1% linolenic acid soybeans. *Crop Sci* 2006, 46(5):1913-1918.

  Chappell AS, Bilyeu KD: A *GmFAD3A* mutation in the low linolenic acid soybean mutant C1640. *Plant Breed* 2006, 125(5):535-536.

  Chappell AS, Bilyeu KD: The low linolenic acid soybean line PI 361088B contains a novel *GmFAD3A* mutation. *Crop Sci* 2007, 47(4):1705-1710.

  Bilyeu K, Ren C, Nguyen HT, Herman E, Sleper DA: Association of a four-basepair insertion in the P34 gene with the low-allergen trait in soybean. *The Plant Genome* 2009, 2(2):141-148.

  Lenis J, Gillman J, Lee J, Shannon J, Bilyeu K: Soybean seed lipoxygenase genes: molecular characterization and development of molecular marker assays. *TAG Theoretical and Applied Genetics* 2010, 120(6):1139-1149.

  USDA/ARS, Agresearch University of Tennessee: Notice of release of soybean germplasm line SB-01 with novel alleles for omega-3 fatty acid desaturase. In: *Germplasm Release.* Bilyeu K, Pantalone VR; 2010.

- The soybean aphid [Aphis glycines Matsumura] is an important soybean pest in North America. The dominant aphid resistance gene Rag1 was previously mapped from the cultivar `Dowling" to a 12 cM marker interval on soybean chromosome 7 [formerly linkage group (LG) M]. ARS scientists at W. Lafayette IN, Beltsville MD, and Urbana IL with researchers at the University of Illinois developed additional genetic markers more closely linked to Rag1 was needed to accurately position the gene to improve the effectiveness of marker-assisted selection (MAS) and to eventually clone it. Single nucleotide polymorphisms (SNPs) near Rag1 and were positioned relative to Rag1. 824 BC4F2 and 1,000 BC4F3 plants segregating for the gene were screened with markers flanking Rag1. Plants with recombination events close to the gene were tested with SNPs identified in previous studies along with new SNPs identified using the Williams 82 draft soybean genome sequence and gene-scanning melt-curve analysis. Progeny of these recombinant plants were evaluated for aphid resistance. These efforts resulted in the mapping of Rag1 between the two SNP markers 46169.7 and 21A, which corresponds to a physical distance on the Williams 82 draft genome of 115 kb. Markers identified in this study that are closely linked to Rag1 will be a useful resource in MAS for this gene. In addition, several candidate genes for Rag1 are present within the 115 kb interval.

  Kim, Ki-Seung, Bellendir, S., Hudson, K., Hill, C., Hartman, G., Hyten, D., Hudson, M., and Diers, B. 2010. Fine Mapping the Soybean Aphid Resistance Gene Rag1 in Soybean. Theoretical and Applied Genetics X:xxx-xxx.

- Scientists at Southern Illinois University demonstrated NIL confirmation of QTL for SCN SDS and seed yield, Al toxicity and insect resistance.

  Triwitayakorn K, Njiti VN, Iqbal MJ , Yaegashi S, Town C, and D. A. Lightfoot. 2005. Genomic analysis of a region encompassing QRfs1 and QRfs2: genes that underlie soybean resistance to sudden death syndrome. Genome/Génome 48: 125-138.

  Kazi S., Shultz J, Bashir, R., Afzal J, Njiti V, Lightfoot D.A. 2008. Separate loci underlie resistance to soybean sudden death syndrome in 'Hartwig' by 'Flyer'. Theoretical and Applied Genetics 116: 967-977

  Afzal AJ, Srour A, Saini N, Lightfoot DA, 2008. The multigeneic Rhg1 Locus: A model for the effects on root development, nematode resistance and recombination suppression. Nature Preceedings hdl:10101/npre.2008.2726.1

  Kazi S, Shultz J, Afzal J, Hashmi R, Jasim M, Bond J, Arelli PR, Lightfoot DA.2010. Iso-lines and inbred-lines confirmed loci that underlie resistance from cultivar 'Hartwig' to three soybean cyst nematode populations. Theor Appl Genet 120:633-640

  Yesudas C, Karzi S., Shultz J, Bashir, R., Woodrow L, Lightfoot D.A. 2010. Identification of loci underlying resistance to herbivory by Japanese Beetles in soybean. Theor Appl Genet (in press).

  Lightfoot DA, Njiti VN, Gibson PT, Kassem MA, Iqbal JM, and Meksem K. 2005. Registration of Essex x Forrest recombinant inbred line (RIL) mapping population. Crop Science 45(4): 1678-1681.

- Asian Soybean Rust (ASR) is a formidable threat to soybean production in many areas of the world including the US. Only five sources of resistance have been identified (Rpp1, Rpp2, Rpp3, Rpp4, and Rpp5). Rpp4 was previously identified in the resistant genotype PI459025B and it was mapped within 2 cM of Satt288 on soybean chromosome 18 (linkage group G). Using SSR markers, we developed a BAC contig for the Rpp4 locus in the susceptible cultivar Williams82 (Wm82). Sequencing within this region identified three Rpp4 candidate disease resistance genes (Rpp4C1-Rpp4C3 (Wm82)) belonging to a family of coiled coil-

nucleotide binding site-leucine rich repeat (CC-NBS-LRR) disease resistance genes with high similarity to the RCG2 disease resistance family from lettuce. Constructs developed from the Wm82 Rpp4 candidate genes were used for virus-induced gene silencing experiments to silence resistance in PI459025B, confirming that homologous genes confer resistance. Using primers developed from conserved sequences in the Wm82 Rpp4 candidate genes, we have identified five Rpp4 candidate genes (Rpp4C1-Rpp4C5 (PI459025B)) from the resistant genotype. Additional markers developed from the Rpp4 contig further defined the region containing Rpp4 to include Rpp4C2 (PI459025B), and likely Rpp4C4 (PI459025B) and Rpp4C5 (PI459025B). Sequencing of RT-PCR products revealed that Rpp4C4 (PI459025B) is highly expressed in the resistant genotype while expression of the other candidate genes is largely undetectable. These data support Rpp4C4 (PI459025B) as the single candidate gene for Rpp4-mediated resistance to ASR.

Meyer, J.D., Silva, D.C., Yang, C., Van De Mortel, M., Pedley, K.F., Hill, J.H., Shoemaker, R.C., Abdelnoor, R., Whitham, S.A., Graham, M.A. 2009. Identification and Analyses of Candidate Genes for Rpp4 Mediated Resistance to Asian Soybean Rust in Soybean (Glycine max). Plant Physiology. 150:295-307.

- Scientists at Southern Illinois University helped confirm the RLK at Rfs2/rhg1 as a major resistance gene for SDS and SCN.

  Afzal AJ, Srour A, Saini N, Lightfoot DA, 2008. The multigeneic Rhg1 Locus: A model for the effects on root development, nematode resistance and recombination suppression. Nature Proceedings hdl:10101/npre.2008.2726.1

- University of Missouri soybean scientists have developed genetic populations using an exotic soybean rust resistance germplasm, DT2000 (PI 635999). Among these, two recombinant inbred lines (RILs) populations derived from the Williams 82 x DT2000 and S01-8401 x DT2000 crosses were used for molecular characterization of resistance gene(s) in collaboration with soybean scientists in Vietnam and USDA-ARS, UIUC, Urbana, IL. Using bulked segregant analysis (BSA) and selective genotyping with the soybean universal single nucleotide polymorphism (SNP) panel (USLP 1.0), we have worked with USDA-ARS soybean scientists, Beltsville, MD and have detected resistance locus and mapped close to the known *Rpp3* location on Chr. 6 (LG C2). Based on chi-square test of segregation for lesion types, reddish brown (RB) and tan (TAN), it was speculated that resistance gene in DT2000 can be a new allele to *Rpp3* gene. Additional crosses have been developed for an allelism study to confirm the speculation. In addition to the *Rpp3*-related gene on Chr. 6, we also detected and mapped an additional rust resistance location to Chr. 18 (LG G). It suggested that besides single genes, quantitative trait loci (QTL) may play a significant role in controlling partial resistance in different soybean germplasm. We are currently continuing the development of additional populations of new resistance sources in attempts to dissect QTL for partial resistance to soybean rust.

  Pham, T.A., C.B. Hill, M.R. Miles, B.T. Nguyen, T.T. Vu, T.D. Vuong, T.T. VanToai, H.T. Nguyen, and G.L. Hartman. 2010. Evaluation of soybean for resistance to soybean rust in Vietnam. Field Crop Research (*in press*)

  Vuong, T.D, B.T. Nguyen, T.T. Vu, D.L. Hyten, P.B. Cregan, D.R. Walker, D.A. Sleper, J.G. Shannon, and H.T. Nguyen. 2009. Genetic mapping of resistance to soybean rust (*Phakopsora pachyrhizi* Syd.) in the soybean cultivar DT2000. The National Soybean Rust Symposium, December 9-11, New Orleans, LA. (http://www.apsnet.org/online/sbr).

C.5.d - User interface for informatics tools

- ARS scientists at Ames IA reported a large number of databases have been developed throughout the world. These databases house vast amounts of genetic and biological data and often provide unique analysis tools. Knowing where to find specific types of data or specific analysis tools is difficult. Staff with the biological databases SoyBase, Gramene and Legume Information System report the use of the semantic web protocol to make data and services more easily discoverable. This protocol will make data contained in each database more easily transferred to researchers. This protocol will also make discovering computational resources hosted by each database easier for public, private and university researchers. The use of published semantic protocol will facilitate the leveraging of resources built up by the participating databases for soybean and grass genetics to other important legume and grain species. The protocol will also make it easier to identify data that might be relevant to other non-legume or grass species. Scientific data integration and computational service discovery are made more difficult by the separate and independent construction of biological databases, which makes the exchange of scientific data between information resources difficult and labor intensive. A recently described semantic web protocol, the Simple Semantic Web Architecture and Protocol (SSWAP; pronounced "swap") offers the ability to describe data and services in a semantically meaningful way. Three major information resources (Gramene, SoyBase and the Legume Information System [LIS]) utilized SSWAP to semantically enable their data and web services. High priority Quantitative Trait Locus (QTL) data, genomic mapping data, trait and phenotypic data, and sequence data and associated services such as BLAST for publication, data retrieval, and service invocation via semantic web services were targeted. The data and services were mapped to concepts and categories as implemented in legacy and de novo community ontologies. Then SSWAP was used to express these offerings in OWL RDF/XML documents appropriate for their semantic discovery and retrieval. SSWAP services were implemented to respond to web queries and return data as requested. These services are registered with the SSWAP Discovery Server and are available for semantic discovery and engagement at http://sswap.info. A total of 10 services delivering QTL information from Gramene were created. Six services delivering information about soybean QTLs, and seven services delivering information from SoyBase were created. For LIS, we constructed three services, two of which allow the retrieval of DNA and RNA FASTA sequences, with the third service providing nucleic acid sequence comparison capability (BLAST). Our implementation of approximately two dozen such services means that biological data at three large information resources (Gramene, SoyBase, LIS) is available for programmatic access, semantic searching, and enhanced interaction between the separate missions of these three resources.

  Nelson, R., Avraham, S., Shoemaker, R.C., May, G., Ware, D., Gessler, D.D. 2009. Applications and Methods Utilizing the Simple Semantic Web Architecture and Protocol (SSWAP) for Bioinformatics Resource Discovery and Disparate Data and Service Integration. BioMed Central (BMC) BioData Mining. 10:309.

C.5.e - Large datasets

- Greg May serves as the PI on the Legume Information System (LIS) developed by the National Center for Genome Resources in cooperation with ARS. LIS is a comparative legume resource that integrates genetic and molecular data from multiple legume species enabling cross-species genomic and transcript comparisons. The LIS virtual plant interface allows simplified and intuitive navigation of transcript data from Medicago truncatula, Lotus japonicus, Glycine max and Arabidopsis thaliana. Transcript libraries are represented as images of plant organs in different developmental stages, which are selected to query the

analyzed and annotated data. Complex queries can be accomplished by adding modifiers, keywords and sequence names. LIS also contains annotated genomic data featuring transcript alignments to validate gene predictions as well as motif and similarity analyses. The genomic browser supports comparative analysis via novel dynamic functional annotation comparisons. In collaboration with ARS scientists at Ames, IA LIS is currently undergoing substantial redevelopment to improve the productivity of scientific research across legume species by providing a comprehensive, up-to-date information resource for legume data.

http://www.comparativelegumes.org

C.5.f - Long-term curation of soybean genome sequence

- See A1a.

D – Transformation/Trangenics

Advances in the utility of soybean transformation methods have resulted from the development of selectable marker-free transgenic soybean lines, multiple gene delivery systems, transformation and regeneration of elite cultivars, and tissue-specific and inducible promoters. The public sector benchmark for producing 500 plants per person per year was met in 2007. A key recommendation for 2007 that remains a major concern is the need for greater coordination and interaction among the existing soybean transformation laboratories. This coordination could lead to greater efficiency and capacity. A variety of transformation based methods for functional analysis of soybean genes have been tested. Among these are *Agrobacterium rhizogenes* mediated RNAi silencing, and viral induced gene silencing (VIGS).

D.1 – Create a Transgenic Event Repository.

- **Little progress has been made toward establishing a genetic stock repository for RILs, NILs, Tilling lines, Tnt1 or mPing lines, and fast neutron mutation lines.**

D.2 – Create a Virtual Center for Transgenics/Transformation.

- In addition to maintaining the GenBank® nucleic acid sequence database, the National Center for Biotechnology Information (NCBI) provides analysis and retrieval resources for the data in GenBank and other biological data made available through NCBI's Web site. NCBI resources include Entrez, the Entrez Programming Utilities, My NCBI, PubMed, PubMed Central, Entrez Gene, the NCBI Taxonomy Browser, BLAST, BLAST Link(BLink), Electronic PCR, OrfFinder, Spidey, Splign, RefSeq, UniGene, HomoloGene, ProtEST, dbMHC, dbSNP, Cancer Chromosomes, Entrez Genome, Genome Project and related tools, the Trace and Assembly Archives, the Map Viewer, Model Maker, Evidence Viewer, Clusters of Orthologous Groups (COGs), Viral Genotyping Tools, Influenza Viral Resources, HIV-1/Human Protein Interaction Database, Gene Expression Omnibus (GEO), Entrez Probe, GENSAT, Online Mendelian Inheritance in Man (OMIM), Online Mendelian Inheritance in Animals (OMIA), the Molecular Modeling Database (MMDB), the Conserved Domain Database (CDD), the Conserved Domain Architecture Retrieval Tool (CDART) and the PubChem suite of small molecule databases. Augmenting many of the Web applications are custom implementations of the BLAST program optimized to search specialized data sets. These resources can be accessed through the NCBI home page at www.ncbi.nlm.nih.gov.

  Wheeler, David L., Tanya Barrett, Dennis A. Benson, Stephen H. Bryant, Kathi Canese, Vyacheslav Chetvernin, Deanna M. Church, Michael DiCuccio, Ron Edgar, Scott Federhen, Lewis Y. Geer, Yuri Kapustin, Oleg Khovayko, David Landsman, David J. Lipman, Thomas L. Madden, Donna R. Maglott, James Ostell, Vadim Miller, Kim D. Pruitt, Gregory D. Schuler, Edwin Sequeira, Steven T. Sherry, Karl Sirotkin, Alexandre Souvorov, Grigory Starchenko, Roman L. Tatusov, Tatiana A. Tatusova, Lukas Wagner and Eugene Yaschenko. 2007. Database resources of the National Center for Biotechnology Information; Nucleic Acids Research, Vol. 35, Database issue **D5-D12**

D.3 – Establish A Soybean Regulatory Promoter Set.

Promoters (approx. 100) that induce correct spatial and developmental time expression in a common cassette to permit tissue-specific transgene expression are needed. In parallel, we need to create a reporter, for example GFP (green fluorescent protein), transgenic line for each cassette to serve as a control. This line would be provided with the cassette to the requesting investigator.

- The success of plant genetic transformation relies greatly on the strength and specificity of the promoters used to drive genes of interest. GFP gene expression mediated by a polyubiquitin promoter (Gmubi) from soybean (Glycine max) was analyzed in stably transformed soybean tissues. Strong

GFP expression was observed in stably transformed proliferative embryogenic tissues. In whole transgenic plants, GFP expression was observed in root tips, main and lateral roots, cotyledons and plumules in young plants as well as in leaf veins, petioles, flower petals, pollen, pods and developing seeds in mature plants. GFP expression was localized mainly in epidermal cells, leaf mesophyll, procambium and vascular tissues. Introduction of an intron-less version of the Gmubi promoter (Gmupri) displayed almost the same GFP expression pattern albeit at lower intensities. The Gmubi promoter showed high levels of constitutive expression and represents an alternative to viral promoters for driving gene expression in soybea**n.**

Hernandez-Garcia1, C.M., Adriana P. Martinelli, Robert A. Bouchard and John J. Finer. 2009. A soybean (Glycine max) polyubiquitin promoter gives strong constitutive expression in transgenic soybean. Journal Plant Cell Reports, 28 (5) 837-849

- John Finer, Paul Rauhton and Jeff Chen developed the Soybean Upstream Regulatory Element Database (SURE). This database provides validation tools for estimating the strength of promotors for soybean transformation.

  http://www/oardc.ohio-state.edu/SURE/

- Scientists at the University of Missouri have developed a soybean gene index resource that should be an excellent resource to identify promoters that show a high level of tissue specificity.

  http://digbio.missouri.edu/soybean_atlas/

D.4 – Improve Soybean Transformation Efficiency.

Current transformation efficiencies are approximately 3.5% using the organogenesis approach and approximately 25% using transformation of somatic embryos. A big limitation to public efforts to increase the number of transgenic events is the availability of greenhouse space. A variety of other technical challenges also limit the efficiency and utility of soybean transformation. The community decided on the following priorities to address these issues.

- Current efficiency is probably satisfactory for most applications except for T-DNA mutagenesis. The greatest need is probably to expand the range of genotypes amenable to transformation.

  Widholm, J., J. Finer, L. Vodkin, H. Trick, P. LaFayette, J. Li,and W.A. Parrott. 2009. Soybean. In: F. Kempken and C. Jung (eds.), Genetic Modification of Plants, Biotechnology in Agriculture and Forestry, Springer-Verlag:Berlin Heidelberg 64: 473-498.

- Scientists at the University of Kentucky have developed a soybean somatic embryo transformation protocol that is adapted from protocols published by scientists at the University of Georgia.

  http://www.uky.edu/Ag/Agronomy/PLBC/Soy_Tr/SOYBEAN%20TRANSFORMATION.htm

Cooperating Laboratories:

Arizona Genomics Computational Laboratory, BIO5 Institute. University of Arizona, Tucson, AZ.
Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, AZ
CENA, University of São Paulo, Brazil..
Center for Integrative Genomics, University of California, Berkeley, C
Center for Plant Science Innovation, University of Nebraska, Lincoln, NE
Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE
Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN
Department of Agronomy, Iowa State University, Ames, IA
Department of Agronomy, Iowa State University, Ames, IA
Department of Agronomy, Purdue University, West Lafayette, IN
Department of Bioinformatics and Genomics, University of North Carolina at Charlotte
Department of Biology, The University of Western Ontario, London, ON Canada
Department of Biology, University of Washington, Seattle, WA
Department of Biology, University of Wisconsin, Stevens Point, WI
Department of Computer Science, University of Missouri, Columbia, MO
Department of Crop and Soil Environmental Sciences, Virginia Tech, Blacksburg, VA
Department of Crop Sciences, Univ. of Illinois, Urbana, IL
Department of Genetics, Developmental and Cellular Biology, Iowa State University, Ames, IA
Department of Horticulture and Crop Science, OARDC/The Ohio State University, Wooster, OH
Department of Plant Biochemistry and Genetics, University of Kentucky, Lexington, KY
Department of Plant Biology and Genome Center, UC Davis, Davis, CA,
Department of Plant Microbiology and Pathology, University of Missouri, Columbia, MO
Department of Plant Pathology, Physiology and Weed Science, Virginia Tech, Blacksburg, VA
Department of Plant Soil and Agricultural Systems, Southern Illinois University, Carbondale, IL
Department of Plant, Soil and Agricultural Systems, Southern Illinois University, Carbondale IL
Department Plant Science and Landscape Architecture, University of Maryland, College Park, MD.
Division of Biochemistry & Interdisciplinary Plant Group, Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO
Fred Hutchinson Cancer Research Center, Seattle, WA
HudsonAlpha Genome Sequencing Center, Huntsville, AL
Institute of Plant Breeding, Genetics & Genomics, University of Georgia, Athens, GA
Joint Genome Institute – Stanford Human Genome Center, Department of Genetics, Stanford University School of Medicine, Palo Alto, CA
Joint Genome Institute, Walnut Creek, CA
National Center for Biotechnology Information, Rockville Pike, Bethesda, MD
National Center for Soybean Biotechnology, University of Missouri, Columbia, MO
Plant Transformation Core Facility, University of Missouri, Columbia, MO
RIKEN Plant Science Center, Yokohama, Japan.
Southern Crop Protection and Food Research Centre, Agriculture and Agri-Food Canada (AAFC), London, ON, Canada
The National Center for Genome Resources, Santa Fe, NM
USDA, ARS, Soybean Genomics and Improvement Laboratory, Beltsville, MD
USDA-ARS Corn Insects and Crop Genetics Research Unit, Ames, IA
USDA-ARS Crop Production and Pest Control Research Unit, West Lafayette, IN
USDA-ARS Plant Genetics Research Unit, Columbia, MO
USDA-ARS, Iowa State University, Ames, IA
USDA-ARS, Plant Science Research Unit, University of Minnesota, St. Paul, MN.

Most Significant Publications

1. Afzal AJ, A Natarajan, N Saini, M J Iqbal, MA Geisler, H El Shemy, R Mungur, L Willmitzer and DA Lightfoot. 2009. The nematode resistance allele at the rhg1 locus alters the proteome and metabolome of soybean roots. Plant Physiology 151: 1264-1280
2. Afzal AJ, Srour A, Saini N, Lightfoot DA, 2008. The multigeneic Rhg1 Locus: A model for the effects on root development, nematode resistance and recombination suppression. Nature Preceedings hdl:10101/npre.2008.2726.1
3. Bilgin, D.D., Delucia, E.H., Clough, S.J. 2009. A Robust Plant RNA Isolation Method for Affymetrix Genechip® Analysis and Quantitative Real-Time RT-PCR. Nature Protocols. 4:333-340.
4. Bilyeu K, Palavalli L, Sleper DA, Beuselinck P: Molecular genetic resources for development of 1% linolenic acid soybeans. *Crop Sci* 2006, 46(5):1913-1918.
5. Bilyeu K, Ren C, Nguyen HT, Herman E, Sleper DA: Association of a four-basepair insertion in the P34 gene with the low-allergen trait in soybean. *The Plant Genome* 2009, 2(2):141-148.
6. Bilyeu KD, Palavalli L, Sleper DA, Beuselinck PR: Three microsomal omega-3 fatty-acid desaturase genes contribute to soybean linolenic acid levels. *Crop Sci* 2003, 43(5):1833-1838.
7. Bolon, Y.E., B. Joseph, S.B. Cannon, B. Diers, A. Farmer, M.A. Graham, G. May, G. Muehlbauer, J. Specht, Z. Tu, N. Weeks, W. Xu, R.C. Shoemaker, C.P. Vance. 2008. Genomic Studies in Soybean: Toward Understanding Seed Oil and Protein Production [abstract]. IV International Conference on Legume Genomics and Genetics, December 7-12, 2008, Puerto Vallarta, Mexico. Abstract No. W28. p. 101 and Abstract No. P57, p. 63.
8. Bolon, Y-T, Joseph, B, Cannon, S., Graham, M.; Diers, B.; Farmer, A.; May, G.; Muehlbauer, G.; Specht, J.; Tu, Z.-J., Weeks, N.; Xu, W.; Shoemaker, R.; and Vance, C. 2010. Complementary Approaches Characterize the Linkage Group I Seed Protein QTL in Soybean. Biomed Central (BMC) Plant Biology, Accepted: March 3, 2010
9. Brechenmacher, L., Lee, J., Sachdev, S., Song, Z., Nguyen, T., Joshi, T., Oehrle, N., Libault, M., Mooney, B., Xu, D., Cooper, B., Stacey, G. 2009. Establishment of a Protein Reference Map for Soybean Root Hair Cells. Plant Physiology. 149:670-68.
10. Calla, B., K.K. Varala, H. Win, M.E. Hudson, L.O. Vodkin, S.J. Clough. 2009. Preliminary Analysis of High-Throughput Expression Data and Small RNA in Soybean Stem Tissue Infected with Sclerotinia sclerotiorum [abstract]. Proceedings of the International Sclerotinia Workshop. May 31-June 4, 2009, Wilmington, NC. P. 6.
11. Chappell AS, Bilyeu KD: A *GmFAD3A* mutation in the low linolenic acid soybean mutant C1640. *Plant Breed* 2006, 125(5):535-536.
12. Chappell AS, Bilyeu KD: The low linolenic acid soybean line PI 361088B contains a novel *GmFAD3A* mutation. *Crop Sci* 2007, 47(4):1705-1710.
13. Choi, I.-Y., Hyten, D.L., Matukumalli, L.K., Song, Q., Chaky, J.M., Quigley, C.V., Chase, K., Lark, K.G., Reiter, R.S., Yoon, M-S., Hwang, E-Y., Yi, S-In., Young, N.D., Shoemaker, R.C., Van Tassell, C.P., Specht, J.E., and Cregan, P.B. A soybean transcript map: gene distribution, haplotype and SNP analysis. Genetics 176:685-696. 2007.
14. Choi, J.J., N.W. Alkharouf, K.T. Schneider, B.F. Matthews, R.D. Frederick. 2008. An enriched cDNA library and microarray analysis reveal expression patterns in soybean resistant to soybean rust. Functional and Integrative Genomics 8:341-351.
15. Cooper, Jennifer L., Bradley J Till, Robert G Laport, Margaret C Darlow, Justin M Kleffner, Aziz Jamai, Tarik El-Mellouki, Shiming Liu, Rae Ritchie, Niels Nielsen, Kristin D Bilyeu, Khalid Meksem, Luca Comai and Steven Henikoff .2008. TILLING to detect induced mutations in soybean, BMC Plant Biology, 8:9

16. Dierking E, Bilyeu K: New sources of soybean seed meal and oil composition traits identified through TILLING. *BMC Plant Biology* 2009, 9(1):89.
17. Dierking EC, Bilyeu KD: Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. *The Plant Genome* 2008, 1(2):135-145.
18. Dierking EC, Bilyeu KD: Raffinose and stachyose metabolism are not required for efficient soybean seed germination. *Journal of Plant Physiology* 2009, 166(12):1329-1335.
19. Du, J., D. Grant, Z. Tian, R.T. Nelson, L. Zhu, R.C. Shoemaker, J. Ma. 2010. SoyTEdb: a comprehensive database of transposable elements in the soybean genome. BMC Genomics.11:113.
20. Du, J., Tian, Z., Schmutz, J., Bowen, N.J., Shoemaker, R.C., and Ma, J. 2010. Bifurcation and enhancement of autonomous-nonautonomous retrotransposon partnership through LTR swapping in soybean. *The Plant Cell*. 22: 48-61
21. Du, J., Tian, Z., Christian, H., Laten, H., Jackson, S., Cannon, S., Shoemaker, R.C., and Ma, J. 2010. Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: new insights from genome-wide analysis and multi-specific comparison. *The Plant Journal* (in press)
22. Findley SD, Cannon S, Varala K, Du J, Ma J, Hudson M, Birchler J, Stacey G. 2010. A fluorescence in situ hybridization system for karyotyping soybean. Genetics (in press)
23. Gillman JD, Pantalone VR, Bilyeu K: The low phytic acid phenotype in soybean line CX1834 is due to mutations in two homologs of the maize low phytic acid gene. *The Plant Genome* 2009, 2(2):179-190.
24. Gonzalez, D. O. and Vodkin, L.O. 2007. Specific elements of the glyoxylate pathway play a significant role in the functional transition of the soybean cotyledon during seedling development. BMC Genomics 8: 468 (22 page online journal)
25. Grant, D. M., Nelson, R. T., Cannon, S. B., Shoemaker, R. C. 2010. SoyBase, The USDA-ARS Soybean Genetics and Genomics Database. Nucleic Acids Research, Nucleic Acids Res. 38: D843 - D846.
26. Grant, D.M., Nelson, R., Cannon, S.B., Shoemaker, R.C. 2009. SoyBase, The USDA-ARS Soybean Genetics and Genomics Database. Nucleic Acids Research. Doi: 10.1093/nar/gkp798.
27. Hancock, C. Nathan., Kristen Chesser, Tom Clemente, Gary Stacey, Carroll P. Vance, Zhanyuan J. Zhang, and Wayne A. Parrott. 2010. Transposon Tagging And Fast Neutron Mutagenesis In Soybean, Plant & Animal Genomes XVIII Conference, January 9-13, 2010
28. Hernandez-Garcia1, C.M., Adriana P. Martinelli, Robert A. Bouchard and John J. Finer. 2009. A soybean (Glycine max) polyubiquitin promoter gives strong constitutive expression in transgenic soybean. Journal Plant Cell Reports, 28 (5) 837-849
29. Hyten D., Song Q., Zhu Y., Choi I-Y., Nelson R., Costa J., Specht J., Shoemaker R., and Cregan P. Impacts of genetic bottlenecks on soybean genome diversity. Proc. Natl. Acad. Sci. U.S.A. 103:16666-16671. 2006.
30. Hyten, D.L., Choi, I-Y., Song, Q., Shoemaker, R.C., Nelson, R.L., Costa, J.M., Specht, J.E., and Cregan, P.B. Highly variable patterns of linkage disequilibrium in multiple soybean populations. Genetics 175:1937-1944. 2007.
31. Hyten, D.L., I. Choi, Q. Song, J.E. Specht, T.E. Carter, R.C. Shoemaker, E. Hwang, L.K. Matukumalli, P.B. Cregan, P.B. 2010. A high density integrated genetic linkage map of soybean and the development of a 1536 Universal Soy Linkage Panel for quantitative trait locus mapping. Crop Science, in press.
32. Hyten, D.L., J.R. Smith, R.D. Frederick, M.L. Tucker, Q. Song, and P.B. Cregan. Bulked segregant analysis using the GoldenGate assay to locate the Rpp3 locus that confers resistance to Phakopsora pachyrhizi (soybean rust) in soybean. Crop Sci. 49: 265-271. 2009.
33. Hyten, D.L., Q. Song, E.W. Fickus, I. Choi, C.V. Quigley, E. Hwang, M. Pastor-Coralles, P.B. Cregan. High-throughput SNP discovery and assay development in Common Bean. BMC Genomics, submitted.

34. Hyten, D.L., S.B. Cannon, Q. Song, N.T. Weeks, E.W. Fickus, R.C. Shoemaker, J. Specht, A.D. Farmer, G.D. May, P.B. Cregan. High-Throughput SNP Discovery through Deep Resequencing of a Reduced Representation Library to Anchor and orient Scaffolds in the Soybean Whole Genome Sequence. (Accepted by BMC Genomics on November 25, 2009).
35. Hyten, D.L., Song, Q., Choi, I-Y., Yoon, M-S., Specht, J.E., Matukumalli, L.K., Nelson, R.L., Shoemaker, R.C., Young, N.D., and Cregan, P.B. High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor. Appl. Genet. 116:945-952. 2008.
36. Innes, Roger W., C. Ameline-Torregrosa, T. Ashfield, T., E. Cannon, S.B. Cannon, B. Chacko, N.W. Chen, A. Couloux, A. Dalvani, R. Denny, S. Deshpade, J.J. Doyle, A. Egan, V. Geffroy, N. Glover, C.S. Hans, S. Howell, D. Ilut, S. Jackson, S., H. Lai, J. Mammadov, S. Martin Del Campo, M. Metcalf, A. Nguyen, M. O'Bleness, B. Pfeil, R. Podicheti, M. Ratnaparkhe, B. Roe, M.A. Saghai, S. Samain, I Sanders, B. Segurens, M. Sevigna, S. Sherman-Broyles, V. Thareau, D. Tucker, J. Walling, A. Wawrzynski, J. Yi, N.D. Young, 2008. Differential Accumulation of Retroelements and Diversification of NB-LRR Disease Resistance Genes in Duplicated Regions Following Polyploidy in the Ancestor of Soybean. Plant Physiology. 148:1740-1759.
37. Jones, S.I., Gonzalez, D.O., and Vodkin, L.O. 2010. Flux of transcript patterns during soybean seed development. BMC Genomics 11:136.
38. Jungmin Ha, Brian Abernathy, David Grant, Xiaolei Wu, William Nelson, Gary Stacey, Henry T. Nguyen, Randy Shoemaker and Scott A. Jackson. Integration of draft sequence and physical map as framework for genomic research in Soybean (*Glycine max*). (in preparation)
39. Kaczorowski, K.A., K.-S. Kim, B.W. Diers, and M.E. Hudson. 2008. Microarray-Based Genetic Mapping Using Soybean Near-Isogenic Lines and Generation of SNP Markers in the Rag1 Aphid-Resistance Interval, Plant Genome 1:89-98.
40. Karangula UB, M.A. Kassem, L. Gupta, H.A. El-Shemy and D.A. Lightfoot. 2009. Locus interactions underlie seed yield in soybeans resistant to Heterodera glycines. Curr. Issues Mol. Biol. 11 (Suppl. 1): i73-84"
41. Kazi S, Shultz J, Afzal J, Hashmi R, Jasim M, Bond J, Arelli PR, Lightfoot DA.2010. Iso-lines and inbred-lines confirmed loci that underlie resistance from cultivar 'Hartwig' to three soybean cyst nematode populations. Theor Appl Genet 120:633-640
42. Kazi S., Shultz J, Bashir, R., Afzal J, Njiti V, Lightfoot D.A. 2008. Separate loci underlie resistance to soybean sudden death syndrome in 'Hartwig' by 'Flyer'. Theoretical and Applied Genetics 116: 967-977
43. Kim, Ki-Seung, Bellendir, S., Hudson, K., Hill, C., Hartman, G., Hyten, D., Hudson, M., and Diers, B. 2010. Fine Mapping the Soybean Aphid Resistance Gene Rag1 in Soybean. Theoretical and Applied Genetics X:xxx-xxx.
44. Klink V.P., K-H. Kim, V.E. Martins, M.H. MacDonald, H.S. Beard, .N.W. Alkharouf, s-C. Park, B.F. Matthews. 2009. A correlation between host-mediated expression of parasite genes as tandem inverted repeats and abrogation of the formation of female *Heterodera glycines* cysts during infection of *Glycine max*. Planta 230:53-71.
45. Klink V.P., P. Hosseine, P. Matsye, N.W. Alkharouf, B.F. Matthews. 2009. A gene expression analysis of syncytia laser microdissected from the roots of the *Glycine mzx* (soybean) genotype PI 548403 undergoing a resistant reaction after infection by *Heterodera glycines* (soybean cyst nematode). Plant Molecular Biology. 71:525-567.
46. Klink, V.P., B.F. Matthews. 2008. The use of laser capture microdissection to study the infection of *Glycine max* (soybean) by *Heterodera glycines* (soybean cyst nematode). Plant Signaling and Behavior 3 (2):105-107.
47. Klink, V.P., B.F. Matthews. 2009. Emerging approaches to broaden resistance of soybean to soybean cyst nematode as supported by gene expression studies. Plant Physiology 151:1017-1022.
48. Klink, V.P., C.C. Overall, N.W. Alkharouf, M.H. MacDonald, B.F. Matthews. 2007. Laser capture microdissection (LCM) and comparative microarray expression analysis of syncytial cells isolated from incompatible and compatible soybean roots infected by soybean cyst nematode (*Heterodera glycines*). Planta 226:1389-1409.

49. Klink, V.P., Hosseini, P., Macdonald, M.H., Alkharouf, N.W., Matthews, B.F. 2009. Population-specific gene expression in the pathogenic nematode Hederodera glycines exists prior to infection and during the onset of a resistant or susceptible reaction in the roots of Glycine max. Biomed Central (BMC) Genomics. 10:111. http://www.biomedcentral.com/1471-2164/10/111.
50. Klink, V.P., M.H. MacDonald, V.E. Martins, S.C. Park, K.H. Kim, S.H. Baek, B.F. Matthews. 2008. *MiniMax,* a new diminutive *Glycine max* genotype with a rapid life cycle, embryonic potential and transformation capabilities. Plant Cell and Tissue and Organ Culture 92:183-195
51. Klink, V.P., P. Hosseine, P. Matsye, N.W. Alkharouf, B.F. Matthews. 2010. Syncytium gene expression in *Glycine max* [PI88788] roots undergoing a resistant reaction to the parasitic nematode *Heterodera glycines*. Plant Physiology and Biochemistry 48:176-193.
52. Klink, V.P., P. Hosseini, M.H. MacDonald, N.W. Alkharouf, V. Martins, B.F. Matthews. 2009. Population-specific gene expression in the pathogenic nematode *Hederodera glycines* exists prior to infection and during the onset of a resistant or susceptible reaction in the roots of *Glycine max.* BCM Genomics. http://www.biomedcentral.com/1471-2164/10/111.
53. Laurent Brechenmacher, Moon-Young Kim, Jijun Zou, Marisol Benitez, Min Li, Crystal B. McAlvin, Trupti Joshi, Bernarda Calla, Mei Phing Lee, Reena Philip, Marc Libault, Lila O. Vodkin, Dong Xu, Suk-Ha Lee, Steven J. Clough, Gary Stacey. 2008. Transcription profiling of soybean supernodulation by *Bradyrhizobium japonicum.* Mol. Plant-Microbe Int. 21: 631-645
54. Laurent Brechenmacher, Zhentian Lei, Marc Libault, Seth Findley, Masayuki Sugawara, Michael J Sadowsky, Lloyd W Sumner, and Gary Stacey. 2010. Soybean metabolites regulated in root hairs in response to the symbiotic bacterium *Bradyrhizobium japonicum.* Plant Physiol. (submitted)
55. Lee, J., Campbell, K., Scheffler, B.E., Feng, J., Naiman, D.Q., Garrett, W.M., Thibivilliers, S., Stacey, G., Tucker, M.L., Pastor Corrales, M.A., Cooper, B. 2008. Quantitative Proteomic Analysis of Bean Plants Infected by a Virulent and Avirulent Obligate Rust Fungus. Molecular and Cellular Proteomics. 8:19-31.
56. Lenis J, Gillman J, Lee J, Shannon J, Bilyeu K: Soybean seed lipoxygenase genes: molecular characterization and development of molecular marker assays. *TAG Theoretical and Applied Genetics* 2010, 120(6):1139-1149.
57. Li, Y., Zou, J., Li, M., Bilgin, D., Vodkin, L.O., Hartman., G. L., and Clough, S. J. 2008. Soybean defenses against to the soybean aphid. New Phytologist 179: 185-195.
58. Li, R., K. Yu, T. Hatanaka, and D.F. Hildebrand. 2010. *Vernonia* DGATs increase accumulation of epoxy fatty acids in oil. Plant Biotechnology Journal 8:184-195.
59. Li, R., K. Yu, and D. Hildebrand. 2010. DGAT1, DGAT2 and PDAT Expression in Seeds and Other Tissues of Epoxy and Hydroxy Fatty Acid Accumulating Plants. Lipids 45:145-157. http://www.springerlink.com/content/0j8532qt663t6213/fulltext.html
60. Libault M, Farmer A , Brechenmacher L, Drnevich J, Langley RJ, Bilgin DD, Radwan O, Neece DJ, Clough SJ, May GD , Stacey G. 2009. *Complete Transcriptome of the Soybean Root Hair Cell, a Single-Cell Model, and Its Alteration in Response to Bradyrhizobium japonicum Infection* Plant Physiol. 2010 Feb;152(2):541-52.
61. Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, He J, Xu D, May G, Stacey G. 2010. An integrated transcriptome atlas of the crop model *Glycine max* and its use in comparative analysis in plants. Plant J. (in press).
62. Lightfoot DA (2008) Soybean Genomics: Developments through the use of cultivar Forrest. Internat J of Plant Genome. 2008:1-22. doi:10.1155/2008/793158
63. Lightfoot DA, Njiti VN, Gibson PT, Kassem MA, Iqbal JM, and Meksem K. 2005. Registration of Essex x Forrest recombinant inbred line (RIL) mapping population. Crop Science 45(4): 1678-1681.
64. Marc Libault, Sandra Thibivilliers, Osman Radman, Steven J. Clough and Gary Stacey. 2008. Identification of four soybean reference genes for gene expression normalization. Plant Genome. 1:44-54
65. Marc Libault, Trupti Joshi, Kaori Takahashi, Andrea Hurley-Sommer, Kari Puricelli, Sean Blake, Dong Xu, Henry Nguyen, and Gary Stacey. 2009. Large scale analysis of soybean regulatory

gene expression identifies a Myb gene involved in nodule development. Plant Physiol. 151: 1207-1220.

66. Maroof, Saghai, Natasha M. Glover, Ruslan M. Biyashev, Glenn R. Buss, Elizabeth A. Grabau. Genetic Basis of the Low Phytate Trait in Soybean Line CX1834. 2009. Crop Science 49:69-76
67. Mathieu, Melanie, Elizabeth K. Winters, Fanming Kong, Jinrong Wan, Shaoxing Wang, Helene Eckert, Christopher Donovan, David Somers, Kan Wang, Gary Stacey and Tom Clemente. 2009. Establishment of a soybean (*Glycine max* Merr. L) transposon-based mutagenesis respository. Planta 229: 279-289.
68. Meyer, J.D., Silva, D.C., Yang, C., Van De Mortel, M., Pedley, K.F., Hill, J.H., Shoemaker, R.C., Abdelnoor, R., Whitham, S.A., Graham, M.A. 2009. Identification and Analyses of Candidate Genes for Rpp4 Mediated Resistance to Asian Soybean Rust in Soybean (Glycine max). Plant Physiology. 150:295-307.
69. Mian, R.M., Bond, J., Joobeur, T., Mengistu, A., Weibold, W., Grover, S., Wrather, A. 2009. Identification of Soybean Genotypes Resistant to Cercospora sojina by Field Screening and Molecular Markers. Plant Disease. 93:408-411.
70. Mian, R.M., Kang, S., Beil, S.E. 2008. Genetic Linkage Mapping of the Soybean Aphid Resistance Gene in PI 243540. Journal of Theoretical and Applied Genetics. 117:955-962.
71. Mian, R.M., Kang, S., Redinbaugh, M.G. 2009. Microsatellite Diversity of Soybean Genotypes Differing in Bean Pod Mottle Virus Leaf Symptom. Genetic Resources and Crop Evolution. 89:359-67.
72. Mohan, Babu, Alla G. Gagarinova, James E. Brandle and Aiming Wang. 2008. Association of the transcriptional response of soybean plants with soybean mosaic virus systemic infection, J Gen Virol 89 (2008), 1069-1080
73. Nahed, R., Macdonald, M.H., Matthews, B.F. 2008. Protease inhibitor expression in soybean roots exhibiting susceptible and resistance reactions to soybean cyst nematode. Journal of Nematology. 40:138-146.
74. Nelson, R., Avraham, S., Shoemaker, R.C., May, G., Ware, D., Gessler, D.D. 2009. Applications and Methods Utilizing the Simple Semantic Web Architecture and Protocol (SSWAP) for Bioinformatics Resource Discovery and Disparate Data and Service Integration. BioMed Central (BMC) BioData Mining. 10:309.
75. O'Rourke, Jamie A., Rex T Nelson, David Grant, Jeremy Schmutz, Jane Grimwood, Steven Cannon, Carroll P Vance, Michelle A Graham and Randy C Shoemaker. 2009. Integrating microarray analysis and the soybean genome to understand the soybeans iron deficiency response. *BMC Genomics* 2009, 10:376
76. Pham, T.A., C.B. Hill, M.R. Miles, B.T. Nguyen, T.T. Vu, T.D. Vuong, T.T. VanToai, H.T. Nguyen, and G.L. Hartman. 2010. Evaluation of soybean for resistance to soybean rust in Vietnam. Field Crop Research (*in press*)
77. Phytozome v5.0 http://www.phytozome.net)
78. Rashed, N., M. MacDonald, B.F. Matthews. 2008. Protease inhibitor expression in soybean roots exhibiting susceptible and resistant interactions with soybean cyst nematode. Journal of Nematology 40:138-146.
79. Saini, N, J.L. Shultz and D.A. Lightfoot. 2008. Re-annotation of the physical map of Glycine max for ploidy by BAC end sequence driven whole genome shotgun read assembly. BMC Genomics 9:323-940
80. Schlueter, J.A., Brian Scheffler, Scott Jackson, and Randy Shoemaker, 2008. Fractionation of synteny in a genomic region containing tandemly duplicated genes across Glycine max, Medicago truncatula and Arabidopsis thaliana. J. Hered. 99 (4): 390-395.
81. Schlueter, J.A., J.L. Goicoechea, K. Collura, N. Gill, J-Y Lin, Y.Yu, D. Kudrna, A. Zuccolo, C.E. Vallejos, M. Munoz-Torres, M.W. Blair, J. Tohme, J. Tomkins, P. McClean, R.A. Wing and S.A. Jackson 2008. BAC-end sequence analysis and a draft physical map of the common bean (Phaseolus vulgaris L.) genome. Trop. Plant Biol. DOI 10.1007/s12042-007-9003-9.
82. Schmutz, J., Steven B. Cannon, Jessica Schlueter, Jianxin Ma, Therese Mitros, William Nelson, David L. Hyten, Qijian Song, Jay J. Thelen, Jianlin Cheng, Dong Xu, Uffe Hellsten, Gregory D.

May, Yeisoo Yu, Tetsuya Sakurai, Taishi Umezawa, Madan K. Bhattacharyya, Devinder Sandhu, Babu Valliyodan, Erika Lindquis, Myron Peto, David Grant, Shengqiang Shu, David Goodstein, Kerrie Barry, Montona Futrell-Griggs, Brian Abernathy, Jianchang Du, Zhixi Tian, Liucun Zhu, Navdeep Gill, Trupti Joshi, Marc Libault, Anand Sethuraman, Xue-Cheng Zhang, Kazuo Shinozaki, Henry T. Nguyen, Rod A. Wing, Perry Cregan, James Specht, Jane Grimwood, Dan Rokhsar, Gary Stacey, Randy C. Shoemaker and Scott A. Jackson Genome sequence of the palaeopolyploid soybean, 2010. Nature 463178-183.

83. Sebastian, S.A., L. G. Streit, P. A. Stephens, J. A. Thompson, B. R. Hedges, M. A. Fabrizius, J. F. Soper, D. H. Schmidt, R. L. Kallem, M. A. Hinds, L. Feng, and J. A. Hoeck. 2010. Context-Specific Marker-Assisted Selection for Improved Grain Yield in Elite Soybean Populations. Crop Sci. 50: doi: 10.2135/cropsci2009.02.0078
84. Shannon, J.G., Nelson, R.L., Lee, J.D., Wrather, J.A. 2010. Registration of LG04-6863 Soybean Germplasm Line with Diverse Pedigree. Journal of Plant Registrations 4:70-72.
85. Shoemaker, R., D. Grant, T. Olson, W.C. Warren, R. Wing, Y. Yu, H-R. Kim, P. Cregan, B. Joseph, M. Futrell-Griggs, W. Nelson, J. Davito, J. Walker, J. Wallis, C. Kremitski, D. Scheer, S. Clifton, T. Graves, H. Nguyen, X. Wu, M. Luo, J. Dvorak, R. Nelson, S. Cannon, J. Tomkins, J. Schmutz, G. Stacey and S. Jackson. 2008. Microsatellite Discovery from BAC End Sequences and Genetic Mapping to Anchor the Soybean Physical and Genetic Maps. Genome 51(4):294-302.
86. Skoneczka, J, M. A. Saghai Maroof, C. Shang, and G. R. Buss. 2009. Identification of Candidate Gene Mutation Associated With Low Stachyose Phenotype in Soybean Line PI200508 Crop Science 49:247-255
87. Song, Q., G. Jia, Y. Zhu, D. Grant, R.T. Nelson, E. Hwang, D.L. Hyten, P.B. Cregan. 2010. Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. Crop Science, in press.
88. SoyBase (http://soybase.org)
89. Thibivilliers, S., Joshi, T., Campbell, K., Scheffler, B., Borerma, R., Xu, D., Cooper, B., Nguyen, H.T., Stacey, G. 2009. Generation of Phaseolus vulgaris ESTs and investigation of the regulation upon Uromyces appendiculatus infection. Biomed Central (BMC) Plant Biology. 9:46.
90. Tremblay A., S. Li, B.E. Scheffler, B.F. Matthews. 2009. Laser capture microdissection (LCM) and expressed sequence tag analysis of uredinia formed by, *Phakopsora pachyrhizi,* the causal agent of. Asian soybean rust Physiological and Molecular Plant Pathology. 73:163-174.
91. Tremblay, A, B.F. Matthews. 2009. Laser capture microdissection, a new technique to collect specific cells. Food and Agriculture Magazine 4:4.
92. Triwitayakorn K, Njiti VN, Iqbal MJ , Yaegashi S, Town C, and D. A. Lightfoot. 2005. Genomic analysis of a region encompassing QRfs1 and QRfs2: genes that underlie soybean resistance to sudden death syndrome. Genome/Génome 48: 125-138.
93. Trupti Joshi, Zhe Yan, Marc Libault, Hoon Jeong, Sunhee Park, Pamela J. Green, D Janine Sherrier, Andrew Farmer, Greg May, Blake Meyers, Dong Xu, Gary Stacey. 2009. Prediction of novel miRNAs and associated target genes in *Glycine max* BMC Bioinformatics **11**(Suppl 1):S14
94. Tucker, M.L., Puthoff, D.P., Neelam, A., Ehrenfried, M.L., Scheffler, B.E., Ballard, L.L., Campbell, K.B., Cooper, B. 2008. Analysis of expressed sequence tags from Uromyces appendiculatus hyphae and haustoria and their comparison to sequences from other rust fungi. Phytopathology. 98:1126-1135.
95. Tuteja, J. and Vodkin, L.O. 2008. Structural features of the endogenous *CHS* silencing and target loci in the soybean genome. Plant Genome 48: 49-69.
96. Tuteja, J.H., Zabala, G., Varala, K., Hudson, M., and Vodkin, L.O. 2009. Endogenous, tissue-specific short interfering RNAs silence the chalcone synthase gene family in Glycine max seed coats. Plant Cell 21:3063-3077.
97. USDA/ARS, Agresearch University of Tennessee: Notice of release of soybean germplasm line SB-01 with novel alleles for omega-3 fatty acid desaturase. In: *Germplasm Release.* Bilyeu K, Pantalone VR; 2010.

98. Valliyodan B, Xu D, Stacey G, Nguyen HT (2010) LTQ-Orbitrap LC-MS analysis of metabolite changes during seed development in soybean. (Manuscript draft is in progress)
99. Vuong, T.D, B.T. Nguyen, T.T. Vu, D.L. Hyten, P.B. Cregan, D.R. Walker, D.A. Sleper, J.G. Shannon, and H.T. Nguyen. 2009. Genetic mapping of resistance to soybean rust (*Phakopsora pachyrhizi* Syd.) in the soybean cultivar DT2000. The National Soybean Rust Symposium, December 9-11, New Orleans, LA. (http://www.apsnet.org/online/sbr).
100. Vuong, T.D., D.A. Sleper, J.G. Shannon, and H.T. Nguyen. 2010. Novel quantitative trait loci for broad-based resistance to soybean cyst nematode (*Heterodera glycines* Ichinohe) in soybean PI 567516C. Theor Appl Genet (*in revision)*.
101. Wang Z, Libault M, Joshi T, Valliyodan B, Nguyen HT, Xu D, Stacey G, Cheng J. 2010. SoyDB: A Knowledge Database of Soybean Transcription Factors. BMC Plant Biology 2010, 10:14.
102. Wawrzynski, A. T. Ashfield, N.W. Chen, J. Mammadov, A. Nguyen, R. Podicheti, S.B. Cannon, V. Thareau, C. Ameline-Torregrosa, E. Cannon, B. Chacko, A. Couloux, A. Dalwani, R. Denny, S. Deshpande, A.N. Egan, N. Glover, S. Howell, D. Ilut, H. Lai, S.M. Del Campo, M. Metcalf, M. Bleness, B.E. Pfeil, M.B. Ratnaparkhe, S. Samain, I. Sanders, B. Segurens, M. Sevignac, S. Sherman-Broyles, D.M. Tucker, J. Yi , J.J. Doyle, V. Geffroy, B.A. Roe, M.A. Maroof , N.D. Young, R.W. Innes. 2008. Replication of nonautonomous retroelements in soybean appears to be both recent and common. Plant Physiol 148: 1760-1771
103. Wheeler, David L., Tanya Barrett, Dennis A. Benson, Stephen H. Bryant, Kathi Canese, Vyacheslav Chetvernin, Deanna M. Church, Michael DiCuccio, Ron Edgar, Scott Federhen, Lewis Y. Geer, Yuri Kapustin, Oleg Khovayko, David Landsman, David J. Lipman, Thomas L. Madden, Donna R. Maglott, James Ostell, Vadim Miller, Kim D. Pruitt, Gregory D. Schuler, Edwin Sequeira, Steven T. Sherry, Karl Sirotkin, Alexandre Souvorov, Grigory Starchenko, Roman L. Tatusov, Tatiana A. Tatusova, Lukas Wagner and Eugene Yaschenko. 2007. Database resources of the National Center for Biotechnology Information; Nucleic Acids Research, Vol. 35, Database issue **D5-D12**
104. Widholm, J., J. Finer, L. Vodkin, H. Trick, P. LaFayette, J. Li,and W.A. Parrott. 2009. Soybean. In: F. Kempken and C. Jung (eds.), Genetic Modification of Plants, Biotechnology in Agriculture and Forestry, Springer-Verlag:Berlin Heidelberg 64: 473-498.
105. Wu X, Tri DV, Findley SD, Stacey G, Shannon JS, Sleper DAand Nguyen HT. (2010) Construction of a Reference Genetic Map for Soybean with a Large RIL Population. (in preparation)
106. Xiaolei Wu, Guohua Zhong, Seth Findley, Perry Cregan, Gary Stacey, Henry T Nguyen (2008) Genetic marker anchoring by six-dimensional pools for development of a soybean physical map. BMC Genomics 9:28.
107. Xiaolei Wu, Jin Qiu, Henry Nguyen (2010) Genome-wide detection of soybean single-feature polymorphism by Affymatrix GeneChip using genomic DNA. (in preparation)
108. Xiaolei Wu, Sean Blake, David A. Sleper, J. Grover Shannon, Perry Cregan, Henry T. Nguyen (2009) QTL, additive and epistatic effects for SCN resistance in PI 437654. *TAG* 118:1093
109. Xu, W., S. Cho, S.H.Yang, Y.E. Bolon, H. Jia, Y. Xiong, H. Bilgic, J. Boddu, G. Muehlbauer. 2009. Single-Feature Polymorphism Discovery by Computing Probe Affinity Shape Powers. BioMed Central (BMC) Genetics. 10:48
110. Yesudas C, Karzi S., Shultz J, Bashir, R., Woodrow L, Lightfoot D.A. 2010. Identification of loci underlying resistance to herbivory by Japanese Beetles in soybean. Theor Appl Genet (in press).
111. Yoon M-S., Song Q., Choi I-Y., Specht J., Hyten D., and Cregan P. BARCSoySNP23: A panel of 23 selected SNPs for soybean cultivar identification. Theor. Appl. Genet. 114:885-899. 2007.
112. Zabala, G. and Vodkin, L. 2008. A putative autonomous 20.5 kb-CACTA transposon insertion in an *F3'H* allele identifies a new CACTA transposon subfamily in *Glycine max.* BMC Plant Biology 8: 124.
113. Zabala, G. and Vodkin, L. O. 2007. A rearrangement resulting in small tandem repeats in the F3'5'H gene of white flower genotypes is associated with the soybean *W1* locus. Plant Genome 47: 113-124.
114. Zabala, G. and Vodkin, L. O. 2007. Novel exon combinations generated by alternative splicing of gene fragments mobilized by a CACTA transposon in *Glycine max*. BMC Plant Biol 7:38.

115. Zhu, J., Patzoldt, W., Shealy, R., Vodkin, L., Clough, S., Tranel, P. 2008. Transcriptome response to glyphosate in sensitive and resistant soybean. J. Agri Food Chem. 56:6355-6363.

Report Writing Team

Richard F. Wilson, Oilseeds & Biosciences Consulting
David Grant, USDA ARS, Ames IA

Acknowledgements

This report was prepared with the support of the United Soybean Board projects 0276 and 0306.