

# **SOYBEAN GENOMICS RESEARCH**

## **A STRATEGIC PLAN**

**FOR 2008 – 2012**

**This Report Documents a 5-Year Strategic Plan for Soybean Genomics Research.  
The Plan was Co-Authored by a Representative Group of 45+ Scientists Who  
Attended a 30-31 May 2007 Planning Meeting Held in St. Louis, Missouri.  
(a list of all meeting participants can be found in the appendix)**

## Table of Contents

Introduction.....	3
The Soybean Sequencing Effort .....	5
Main Topic A – Genome Sequence .....	6
Sub-Topic A.1 – Genome Informatics.....	6
Sub-Topic A.2 – Genome Finishing.....	7
Sub-Topic A.3 – Transformation/Transgenics (moved to D).....	8
Sub-Topic A.5 – Phaseoloid Genomics.....	9
Sub-Topic A.6 – Breeder Needs as to Soybean Sequence.....	10
Main Topic B – Gene Function .....	11
Sub-Topic B.1 – Gene Function Annotation/Informatics.....	12
Sub-Topic B.2 – Discovery via Mutagenesis .....	12
Sub-Topic B.3 – Functional Genomics Approaches.....	14
Sub-Topic B.4 – Transformation/Transgenics (moved to D).....	16
Sub-Topic B.5 – Breeder Perspectives on Gene Function.....	16
Sub-Topic B.6 – Soybean Producer Expectations - Genomics.....	16
Main Topic C – Germplasm Genomics .....	17
Sub-Topic C.1 – Association Mapping.....	17
Sub-Topic C.2 – Track Breeding-Induced Genomic Change.....	18
Sub-Topic C.3 – Mining Yield QTLs in Exotic Germplasm.....	19
Sub-Topic C.4 – Marker-Assisted Selection Resources.....	20
Sub-Topic C.5 – Germplasm Genomics Informatics.....	21
Sub-Topic C.6 – Transformation/Transgenics (moved to D).....	22
Main Topic D – Transformation/Trangenics .....	22
Sub-Topic D.1 – Create a Transgenic Event Repository.....	23
Sub-Topic D.2 – Create a Virtual Center for Transgenics/Transformation.....	23
Sub-Topic D.3 – Establish A Soybean Regulatory Promoter Set.....	23
Sub-Topic D.4 – Improve Soybean Transformation Efficiency.....	23
Report Writing Team .....	24
Acknowledgements.....	24
Appendix – Meeting Participants.....	25

## **Introduction**

A Genome Strategic Planning Workshop was planned by the Soybean Genetics Executive Committee and convened in St. Louis, MO on 30-31 May 2007. This workshop was attended by 48 people that included many genomics, genetics, and breeding experts with expertise in a wide range of scientific disciplines. Also attending were representatives of the United Soybean Board, the North Central Soybean Research Program, and one Soybean Producer. Previous soybean strategic plans have dealt with identifying tools and resources needed to prepare for the eventual sequencing of the genome. This Workshop represented a paradigm shift for soybean genomics researchers. The joint announcement in January of 2006 by DOE-JGI and USDA-CSREES that sequencing the soybean genome was to be undertaken has led to an acceleration of the attainable goals and targets in many soybean research programs. This announcement has caused the soybean research community to rethink and reassess its strategic research objectives given the impact that the soon-to-be availability of genomic DNA sequence will have on genomic research. A report to the meeting participants by Jeremy Schmutz, Stanford, the leader of the DOE-JGI soybean sequence assembly effort (see next section), indicated 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 has been developed and a draft assembly has been created. This assembly is being evaluated to determine the optimal means for obtaining the final goal of an 8X coverage. The 8X sequence is slated to be complete by the end of 2007, with a final assembly expected to be completed in mid-2008.

Soybean genomic sequence brings a vast amount of data for use in optimizing the rate of scientific discovery and its translation into technological innovation in the production and use of soybean, so the primary objective of this Meeting was to assess the current status of soybean genomics, to identify the resources needed to take advantage of soybean sequence data, and to lay out a strategic plan for soybean genomics from 2008 to 2012.

The Meeting Agenda was split into four half-day sessions. Three major topic areas: A - Soybean Genome Sequencing, B - Soybean Gene Function, and C - Soybean Germplasm Genomics were covered in the first three sessions. In each, the topic area leader presented a brief report of “where-do-we-stand-now” followed by a charge to the participants to develop a “where-to-we-want-to-go” strategic plan for the given main topic. Thereafter, the participants split into sub-topic groups for roundtable discussions led by a person with expertise in the sub-topic. These discussions were intended to identify strategic needs of the research community and to identify milestones needed to achieve the objectives, with one person in each discussion asked to capture the “discussion bullets and desired milestone dates” on a flip chart. During the last half-day session, each sub-topic round-table leader/recorder provided an oral report to all meeting participants and provided a written electronic report to a 5-member writing team who stayed one additional day to assemble the sub-topic written reports into a single document that after review/revision became this Genomics Strategy Planning Document.

Meeting participants were requested to review the 2005 Soybean Strategic Plan prior to their arrival at this 2007 Meeting. The 2005 Plan can be found at this SoyBase web site: [http://soybase.org/SoyGenStrat2005/Soy\\_Genome\\_Strat\\_Plan\\_2005.html](http://soybase.org/SoyGenStrat2005/Soy_Genome_Strat_Plan_2005.html). As the writing team prepared the 2007 Strategic Plan while reviewing the 2005 Strategic Plan, it became apparent 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 in 2007 as they were in 2005. Moreover, some targeted 2005 Plan goals have not moved forward to the extent the 2005 strategic plan participants had originally proposed as completion dates. 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 have apparently not achieved the momentum that the 2005 Plan developers thought would occur (likely due to limited funding). Still, a gratifying number of objectives and milestones identified in the 2005 Plan have been achieved on schedule. The number of SNPs and STSs proposed for discovery and development was exceeded. Inbred mapping resources (RILs) have been developed. Transformation technologies have improved and various gene knock-out systems are working. Thanks to funding by the United Soybean Board and the National Science Foundation, physical and transcript maps are now in stages of completion. Bioinformatic resources and staffing have more than doubled, just in time to receive the whole genome shotgun sequence of soybean. Given these technological developments and the diminishing cost of many genomic-based technologies, the outlook for the next half-decade of soybean genomic research is quite optimistic.

Unless otherwise noted in this report, the year (e.g., 2008, 2009, etc.) associated with a goal or scheduled activity denotes that the goal or activity will be completed by December of the indicated year. For this report, participants were asked to project goals and activities over the next half-decade (2008 – 2012), recognizing of course, that near-term projections were likely to be more certain than long-term ones.

**The Soybean Sequencing Effort**  
**Department of Energy and Joint Genomic Institute.**  
[\(<http://www.jgi.doe.gov/sequencing/why/soybean.html>\)](http://www.jgi.doe.gov/sequencing/why/soybean.html)

Jeremy Schmutz (of JGI) provided the Meeting Participants with some statistics about the soybean sequencing effort. The ancient tetraploid nature of the soybean genome does not appear to be generating problems, at least based on the 4x data accumulated to date.

**Soybean Sequencing Effort Targets:**

**JGI Goal:** A near-complete, ordered and oriented genome sequence that covers at least 98% of the euchromatic soybean sequence.

**JGI Goal:** 80-100 Mb of the genome finished and included in the 8x release.

**Collaborative Goal:** High quality automated gene annotation and public genome browser.

**DOE-JGI Soybean Sequencing Schedule to Date and Forward (subject to revision)**

<b><u>Date</u></b>	<b><u>Activity</u></b>
<b>May 07</b>	Evaluate shotgun, set coverage and choose new clones
<b>Jun 07 - Sep 07</b>	QC new 8x library and sequence additional 4x
<b>Oct 07</b>	8x shotgun and BAC ends complete
<b>Dec 07</b>	Build shotgun assembly
<b>Jan 08 - Feb 08</b>	Order and orientate final assembly
<b>Mar 08 - Jun 08</b>	Final O & O assembly and Annotation
<b>Jul 08 - Oct 08</b>	Release Annotation Browser, manual annotation and analysis
<b>Nov 08</b>	Collate and work on publications
<b>Dec 08</b>	Submit publication(s)

## Main Topic A – Genome Sequence

### Sub-Topic A.1 – Genome Informatics

The Community goals of the Genome Informatics group are the establishment of integrated data, 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. One concern that arose repeatedly is the need for long-term support for genome databases and informatics, not only for soybean but also for other legumes and crop plants.

#### **A.1.a - Annotation Needs.**

- **2008** – Establish a soybean Informatics Steering Committee to address the community’s current and future informatics needs.
- **2008** – Establish the International Soybean Genome Annotation Group (ISGAG), which will serve as community body to interface with JGI for soybean genome annotation and the establishment of a controlled vocabulary nomenclature.
- **2008** – Establish community standards for expression, protein and metabolite profiling platforms and data.
- **2009** – Implement a HapMap browser that transcends linkage groups to ORF to SNP that will help connect the sequence to polymorphisms for breeders. Example: Genomic Explorer y Survey of Immune Response (GEYSIR) Software.
- **2010** – Broadly enable the ability to go from expression data to QTL data. Example: Provide users with an informatic means of mapping microarray expression data onto the genetic QTL data present in SoyBase.
- **2012** – Integrate genome sequence with physical and genetics maps with the goal of integration of functional and phenotypic data.
  - Establish tools for the identification of candidate genes underlying QTLs.
  - Integrate plant traits and phenotypes (e.g., digital image or measurement data) with genetic maps and other genetic data.

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

- **2008** – Merge cytological, genetic and physical maps with the draft whole genome sequence (WGS) of the soybean genome. The goal is to make the sequence as useful as possible by integrating with existing data sets.
- **2008** – Convert linkage group and chromosome names to common number (1-20).
- **2009 to 2010** – Populate database to overlay physical, genetic, cytological maps onto draft genome sequence to a level of “75% consistency”.

#### **A.1.c - Integrate Soybean Genomic Data with that of Related or Other Species.**

Coordinate soybean genomic data with data now available in other species to identify and confirm orthologous genes. Here are some specific goals and datelines:

- **2008** – After the release of the soybean genome sequence – generate syntenic comparisons with the sequences of the below species to confirm gene predictions and models and to enable functional annotation of other non-coding sequences.
  - Model Species (*Arabidopsis*, *Medicago*, *Lotus*).
  - Poplar (*Populus trichocarpa* – *Western Black Cottonwood*).

- **2010** – Obtain genome sequence of *Phaseolus vulgaris* – dry bean.
  - Use the soybean and dry bean sequence to enable sequence transfer to the pulse crops, e.g., *V. radiata* – mung bean and *Vigna unguicula* – cowpea.
- **2012** – Begin finishing draft sequences of many other legumes to:
  - Enable orthologous comparisons: make functional inferences between related genomes.
  - Enable ortholog comparisons between Galegoid and Phaseoloid legumes.

#### **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. The database should include:

- **2008 to 2012** – Develop a soybean genomic database that has:
  - The ability to navigate from maps to genes to traits.
  - Different entry portals at a unified web site providing scientists of various backgrounds a user-friendly interface that enables direct access to relevant data (i.e., multiple ways to access and manipulate the data well).
  - Expansion of the Soybean Breeders Toolbox.
  - Augmented databases with phenotypic data.
  - Expression data QTLs.
  - Transcend data and/or data types (trait to gene or gene to trait).
- **2010** – Integrate all databases, including the “Seed and Population” databases that exist for available genetic stocks, mutants and germplasm collection, into the soybean genomic data base.

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

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

- **2008** – Establish support system to expedite the creation of the transposon database. This is an immediate need.
- **2009** – Release of an expert-curated transposon database for manual-based genome annotation.

**Note:** Additional annotation/informatics bullets were developed in the B-1 and C-5 sub-topic sessions, so go to the Main Topic B and C sections to view these bullets.

#### **Sub-Topic A.2 – Genome Finishing**

What is meant by a “finished genome”? By the end of 2008, the genome sequence will not be “finished” *per se* (i.e., as an end-to-end sequence), but it should be of high enough quality to be sufficient for most research purposes. Still, many gaps will remain in highly repetitive areas, centromeres, and within many scaffolds. The following are steps we feel necessary to make the sequence as usable as possible for soybean researchers.

**A.2.a - Initial Genome Assembly.**

- **2008** – 99% of all genes will be sequenced with high accuracy.
- **2008** – 20,000 full length cDNAs sequenced (will support annotation).
- **2008** – 100% of scaffolds > 100 kb ordered and oriented within pseudomolecules.
  - each scaffold > 100 kb will have at least two map-consistent markers.
- **2008** – Pseudomolecule assemblies will be publicly available.

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

- **2009** – Annotations will be available for download and via a browser at JGI
- **2009 to 2012** – Gene expression support will accompany annotations where possible (cDNA, EST, homology to transcripts from other species).

**A.2.c - Selective Re-Sequencing.**

- **2010** – BAC libraries will be created and BAC-end sequenced to low coverage (~5x) for ~20 diverse accessions.
- **2010 to 2012** – Targeted re-sequencing will be carried out from these accessions for regions of interest.
- **2009 to 2012** – A deep transcript sequencing project for gene discovery and gene-model validation (i.e. 454, Solexa, etc.) was suggested (but due to evolving technologies there was less consensus on this goal than on the above two).

**Sub-Topic A.3 – Transformation/Transgenics (moved to D)**

Note: Bullets in this discussion section were moved to a “new” section - Main Topic D.

**Sub-Topic A.4 – Genome Re-sequencing**

Once the genome sequence is available, the first goal will be using it for the improvement of soybean via development of markers and mapping of traits in order to develop the most efficient tools for the application of marker-assisted selection soybean breeding. To do this, some limited ‘resequencing’ of related genomes will be necessary for marker development and trait mapping.

**A.4.a - SNP Genotyping.**

- **2007** – The best current platform is the Illumina BeadStation using the Golden Gate Assay – now used by Beltsville ARS group (Cregan and Hyten). Expectation of significantly lower costs relative to high-throughput mapping — if other genomicists and breeders adopt the technology.
- **2007** – A set of 1536 loci with high minor allele frequencies will be genotyped across core germplasm collection, and in 500 RILs of the inter-specific mating of Williams 82 (*G. max*) x PI 468.916 (*G. soja*).
- **2008** – SNPs will be mapped using one or more of these mapping populations:
  - Beltsville: RIL populations: 500 Williams 82 x PI 468.916 (*G. soja*) , 300 Harosoy x Clark, 233 Minsoy x Noir, 233 Minsoy x Archer.
  - Missouri: 1,300 Forrest x Williams 82, 600 *G. soja* x *G. max*
  - Virginia Tech: 800 PI96.983 (*G. max*) x Lee68 , 300 PI407.162 (*G. soja*) x V71-370 (*G. max*).
  - SIU: RIL populations: 975 Resnik x Hartwig, 500 Essex x Forrest.

- **2008** – The Illumina Infinium assay for genotyping 25,000 SNPs will be ready for association mapping, using a core collection of genotypes.

#### **A.4.b - Resequencing for SNP Discovery.**

- **2008** – Discover a minimum of 15,000 SNPs (one every ~50kb) – or as many as 25,000 if costs fall and technology improves.
- **2010 to 2012** – Discover 120,000 SNPs located across genome to permit a successful haplotyping of the entire germplasm collection of 18,000+ accessions.

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

- **2008** – Current re-sequencing projects will automatically identify new SSR loci, which would allow 2,000 more SSRs to be placed on map along with SNPs.
- **2008** – Several groups are currently working with Single Feature Polymorphisms (SFPs), using Affymetrix data, and will use these to genotype RILs and NILs.
- **2008** – Sequenom SNP genotyping now available for about 1000 existing SNPs, but for any newly discovered SNPs, this platform requires the redesign of the primers for the SNP-containing amplicons.
- **2008** – Re-sequencing via the Sanger / Solexa / 454 / ABI SOLiD platforms will be necessary for SNP discovery in specific targeted genotypes (2-5).
- **2008** – BAC-based re-sequencing should be investigated as it provides positional information across larger segments of the genome.

### **Sub-Topic A.5 – Phaseoloid Genomics**

The group expressed primary interest in *Phaseolus vulgaris* (bean) as a diploid model species for syntenic sequence comparisons with soybean. *Phaseolus* and its allies (e.g., *Vigna* species) are  $2n = 22$ , with relatively small genomes (roughly half that of soybean), and both *Phaseolus* and *Vigna* diverged from *Glycine* about 19-23 mya. *Phaseolus* has much in common with *Glycine* (e.g., determinate nodules), and may be very useful for the syntenic discovery of genes involved in abiotic stresses (e.g., phosphorus, drought). These expectations for *Phaseolus* genomics are conditioned on its status as a sequencing target; it is currently being considered by JGI, and is being sequenced at a more limited level by other groups (e.g., S. Jackson lab – USA, and the V. Geffroy lab – INRA, in collaboration with the R. Innes lab – PGRP R-gene project).

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

Genetic redundancy, so prevalent in soybean due to genome duplication(s), is not so abundant in *Phaseolus*; therefore, genetic dissection of difficult agronomic traits may be more efficient in *Phaseolus*, and the results transferred rapidly back to soybean (e.g., drought, rust resistance, etc.). In most cases, the future role of the soybean research community will be indirect, primarily one of voicing support for genomics initiatives by the bean research community for the following goals:

- **2008** – Targeted sequencing of orthologous regions involved in stress and resistance responses; primarily sequencing of BAC libraries by various groups.
- **2008 to 2009** – Deep sampling of ESTs for key traits such as root and various nodule developmental stages, including late stages of such development.
- **2010** – Production of a draft sequence of *Phaseolus vulgaris*.

- **2008 to 2012** – Integration of *Phaseolus* data in databases with soybean (e.g., through the Legume Information System).

#### **A.5.b - Other Genera.**

Other genera were discussed more briefly since these can fill the temporal gap between *Phaseolus* and *Glycine*. A BAC library exists for one of the closest generic relatives of *Glycine*, which is *Teramnus* (diverged 10-12 mya, close to the divergence of homoeologous genomes in *Glycine*). Unfortunately, *Teramnus* species have relatively large genomes (nearly the size of soybean) despite a relatively low chromosome number ( $2n = 28$ ) and are not of economic interest. *Pachyrhizus* (jicama) diverged from *Glycine* 15-18 mya and is of some economic importance in the developing world. *Pueraria lobata* (kudzu) is a weed that is closer to *Glycine* (13-15 mya) than *Pachyrhizus*. 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. Because these species diverged from soybean (and its progenitor, *G. soja*) around 5 mya, they afford the opportunity to localize changes in homoeologous regions to events shared among species and thus potentially due to the polyploid event vs. later changes that have occurred since separation from their common ancestor. Moreover, the perennial species, constituting the secondary germplasm pool for soybean, represent an untapped resource for a wide range of agronomically important traits such as drought tolerance and rust resistance. Some of these have been studied (e.g., rust resistance in *G. canescens*); crosses have been made between soybean and one of the perennial species, *G. tomentella*.

- **2010-2012** – Library construction and targeted sequencing of one or more perennial *Glycine* species.

#### **Sub-Topic A.6 – Breeder Needs as to Soybean Sequence**

In this session, breeders addressed many items, such as, genes (traits –phenotypes) in the same linkage blocks, the need for additional genic markers, specific genes controlling traits of interest, and alleles currently available in germplasm. Also discussed was the need for genomics information to be integrated in a way that breeders can query it quickly and thus better use markers to select desirable lines and cross combinations. Additional needs were: precise map positions for the E1 thru E8 (maturity genes); ways to push phenotypic information onto genomic data bases; the need to map the genes for deleterious traits (to better select for the non-deleterious alleles); a breeder-friendly SNP detection system; and breeder useful “quick” assays for SNP markers.

- **2008 to 2009** – Construct a 1536-SNP Oligo Pool Assay (OPA) to provide to breeders (on a cost-recovery basis) for use in all aspects of breeding.
- **2008 to 2009** – Re-sequence 17 well-chosen genotypes for thousands of SNPs to evaluate sequence-based genic and other diversity in soybeans.
- **2008 to 2009** – Assay 1000 elite lines for allelic composition at 10,000 SNP loci.
- **2009 to 2012** – Design highly polymorphic breeder friendly marker assays for those SNP loci linked to key genes/QTLs, to be used for formal MAS, or for allele-specific frequency enrichment in progenies or populations.
- **2009 to 2012** – Begin the development of an inexpensive yet convenient 10,000-SNP assay to allow the soybean breeder to routinely examine the diversity of each year’s selected parents and progeny.

- **2009 to 2012** – Continue developing and refining breeder-friendly databases and software for relating phenotypes/markers/germplasm to DNA sequence.
  - Integrate phenotypic data into the genomic databases.
  - Develop breeder outreach relative to SNP assays and genomic databases.

### **Main Topic B – Gene Function**

The soybean genome sequence, once in hand, may reveal as many as 60-65,000 genes involved in soybean plant biology from germination to maturity. Many of these genes will control important agronomic and quality traits. We also anticipate that the soybean sequence will become THE model for basic studies of legume crop biology with much specificity for the soybean itself. Yet, tools must be put in place to allow the analysis of soybean gene function. Annotation is obviously a critical need. Another key element will be the ability to target specific genes for mutation or for gene silencing. The phenotypes of such mutants will be revealing with regard to the functional role of each gene. Functional genomic tools (e.g., transcriptomics, proteomics and metabolomics) will fully develop once the genome sequence is available, allowing gene function to be framed within the context of soybean physiology. Innovative uses of these technologies (e.g., using DNA microarrays to map expression QTLs) also promise advances in identifying important genes for soybean improvement. The available tools for the study of gene function are generic and, therefore, some priorities are needed to identify those aspects of soybean biology that offer the greater opportunities to impact soybean production and value. The community felt that the following topical areas, in rough order of priority, are those that should receive immediate attention as **key soybean processes to explore using functional genomic tools:**

- **Seed composition and quality; including genotype by environment effects**
  - **Leaf and root development**
  - **Need standardization of developmental stages**
  - **Profile developmental stages**
- **Abiotic stresses**
  - **Drought**
  - **Soil pH (plant iron chlorosis)**
  - **Temperature**
  - **Flooding**
  - **Ozone**
- **Biotic stresses**
  - **Soybean cyst nematode**
  - **Fungal diseases including rust**
  - **Aphids and foliar feeding insects**
  - **Viral diseases, BPMV, SMV**

### **Sub-Topic B.1 – Gene Function Annotation/Informatics**

This bioinformatics session in Main Topic B dealt with the needs of genome annotation and determining gene function (for other informatic needs, see prior Sub-Topic A.1- Genome Informatics and subsequent Sub-topic B.5 - Germplasm Genomics 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.**

- **2008** – Expressed sequence confirmation of 35% of the initial predicted genes
- **2009** – Create an International Soybean Genome Annotation Group (ISGAG) to coordinate core annotation and data maintenance responsibilities.
- **2009** – Establish standard curatorial practices for community-driven literature and gene annotation updates.
- **2009** – Organize workshops/jamborees for community-driven annotation updates at the gene family level.
- **2009** – Integration of mutagenesis, knockout, expression data for gene function annotation.
- **2010** – Confirmation of 75% of predicted transcriptome (high throughput technology, EST, sequence-based confirmations, etc.).

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

- **2008** – Approximately 20,000 full length-cDNA .
- **2008 to 2012** – Establish / extend resources for annotation of transposons, repeats, small RNAs, literature, and conserved non-coding elements.

#### **B.1.c - Integration with Other Species.**

- **2009** – Integrate KEGG, GO, and Plant Ontology annotation with other plant genome resources.
- **2008 to 2012** – Establish / integrate evidence of synteny, orthologous genes, expression/co-expression levels, and regulatory networks in a comparative context.

**Note:** Additional annotation/informatics bullets were developed in the A-1 and C-5 sub-topic sessions, so go to the Main Topic A and C sections to view these bullets.

### **Sub-Topic 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. For example, RNAi-mediated gene silencing is now a routine method to study soybean gene function.

Root traits can be rapidly analyzed by *Agrobacterium rhizogenes*-mediated delivery of RNAi constructs through hairy root transformation. A new method for studying above ground traits is transient silencing using viral-induced gene silencing (VIGS). These are new and exciting developments for soybean. A variety of other mutagenesis procedures are now either established or being evaluated. These technologies are complementary and each has its strengths and weaknesses. Therefore, the community recommended that efforts continue to build upon these technologies 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. Given a sufficiently large, highly mutated soybean population, mutations in any gene can be identified. Previous strategic plans recommended that TILLING populations and libraries should be developed as a public genetic resource. On-going TILLING projects can be found at Southern Illinois University, Purdue University, in conjunction with USDA-ARS, and the University of Missouri. Individually, these facilities provide a service that can identify approximately 6 point mutations in any given gene.

- **2008** – Develop a community supported TILLING service facility that can efficiently service the soybean community. Ideally, this facility should have external support to reduce the cost below the current price of ca. \$3,000 per gene.
- **2010** – Resolve any issues that would prevent the close collaboration between the existing TILLING projects to allow greater capacity and efficiency.

### **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. The mutant populations that arise can be screened for useful mutations. In addition, PCR can be used to rapidly screen pools of mutagenized seeds for a deletion in a given target gene. Therefore, fast-neutron populations can be used for both forward and reverse genetic screens for useful mutations.

- **2008** – Complete screening of existing populations of fast-neutron mutagenized seed to evaluate the usefulness of this method for both forward and reverse genetics.
- **2010** – Assuming that this initial screening proves promising, develop a community service resource for the soybean community to screen for mutations in target genes using high-through-put PCR methods.
- **2012** – Develop the capacity to target at least 1000 genes per year via PCR-based screening of fast-neutron populations.

### **B.2.c - Resources Needed for Gene Function Studies Using Transposon Tagging.**

Both TILLING and fast-neutron mutagenesis have the advantage that they do not involve the construction of transgenic plants. Therefore, mutants derived by these methods can be directly used in soybean breeding. However, in both cases, it is necessary to use heavily mutagenized populations requiring extensive back-crossing before the mutant lines can actually be used. TILLING also has the drawback that only a small percentage of the

point mutations result in a nonsense mutation that would be expected to disrupt function. Otherwise, one has to be lucky that a missense mutation will give a measurable phenotype. In contrast, 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. The various methods for soybean mutagenesis are complementary and the community felt that all should be pursued. Recent work has demonstrated that both the maize Ac/Ds and tobacco Tnt1 retrotransposon do transpose in soybean and, therefore, are suitable for mutagenesis. 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.

- **2008** – Generate 1600 independent Ds transformation events.
- **2010** – Generate 3200 independent Ds events.
- **2010** – Make a decision whether to continue the use of the Ac/Ds transposon system. If appropriate, then increase the number of independent Ds transformants to a total of 5,000 independent events.
- **2010** – Generate at least 500 independent Tnt1 transformation events.
- **2010** – Compare the cot node approach to the somatic embryogenesis approach for use with the Tnt1 transposon system.
- **2008** – Demonstrate transposition of the rice ping pong retrotransposon in soybean. Compare this system to that of Ac/Ds and Tnt1.
- **2012** – Generate 200,000 independent Tnt1 insertions.

### **Sub-Topic 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. However, new technology (e.g., Solexa high-throughput-sequencing platform) is predicted to supplant DNA microarrays sometime in the future. The exact timing of this is uncertain. The community felt that DNA microarrays will continue to be useful in the near term, especially if they can be made available at low cost.

- **2009** – Develop a full list of soybean open-reading-frame (ORF) (i.e., gene-coding) sequences derivable from the manually annotated soybean genome sequence.
- **2009** – Use this list of soybean ORFs to create a new generation soybean oligonucleotide array representing the entire soybean genome. Note that this objective may be unnecessary if other high throughput technologies have made DNA microarray technology obsolete.
- **2010** – Develop a unified database for housing and analysis of all soybean transcriptomics data. For maximum utility, this database should be linked to other soybean genomic information (e.g., genome sequence and genetic maps; see bioinformatics section of this document).

- **2008 to 2009** – Use high-through-put sequencing platforms (e.g., Solexa) to survey the presence of small, non-coding RNA in the soybean genome.
- **2010** – Conduct experiments to map expression QTLs for key soybean processes. The current research being conducted at Virginia Tech University on Phytophthora resistance will provide an evaluation of the utility of this approach.
- **2010** – Use laser-capture-microdissection to conduct tissue specific transcriptome surveys. For example, this technology should be applied to flowers, seeds, leaves, roots and nodules, as well as key tissues responding to biotic and abiotic stresses.
- **2010** – Initiate the development of a clone library representing the entire soybean ORFeome. For example, this effort could begin by targeting all soybean transcription factors. These clones should be made in Gateway or comparable vectors that will allow rapid cloning into other compatible vectors for construction of gene fusions, yeast two-hybrid libraries and other uses.
- **2010** – Utilize transcriptomic approaches to analyze the near isogenic lines and variety of soybean mutants that will arise from the completion of other priorities outlined in this document.

### **B.3.b - Proteomic Approaches.**

Transcriptomics provides for an analysis of gene transcription. However, mRNA levels do not correlate well with the level of the corresponding proteins. It is the proteins, such as enzymes, which confer the phenotype that is of primary importance for understanding soybean gene function. The technology for large scale proteomic studies is continuing to improve. Therefore, the community felt that a greater focus on soybean proteomics is warranted.

- **2010** – Develop a proteome catalog of key organs and processes (see list above). Provide analysis of this catalog in the context of soybean physiology and biochemistry.
- **2010** – Produce antibodies to a subset of the ORFeome for use in protein arrays and cellular localization studies. Focus these resources on key processes involved in seed oil and composition, disease resistance, and soybean specific genes.
- **2009** – Utilize the annotated soybean genome sequence to improve the current database for soybean proteomics. This will need to be continued as knowledge of the full coding capacity of the soybean genome increases.
- **2010** – Proteomics is more than just the identification of proteins. Beginning in 2010, if not earlier, efforts should be made to elucidate the key interactions of proteins and protein complexes involved in important soybean processes.
- **2010** – Protein modification can profoundly affect function. A variety of modifications are possible. In 2010, if not earlier, research should focus on understanding the role of these modifications in key soybean processes. An important initial focus should be on identifying the soybean phosphoproteome.

### **B.3.c - Metabolomic Approaches.**

Soybean biochemical processes give rise to a plethora of metabolites, each with its own spectrum of activity. Currently, the great majority of these chemicals are unidentified. The challenge of metabolomics is to identify these chemicals, study their dynamics and explain their cellular function. This is still a developing area with rapid technological

advances taking place. This research area is very important for soybean where many metabolites are known to be involved in determining the quality of soybean as a functional food.

- **2008** – Continue on-going efforts to expand the catalog of chemical standards relevant to soybean metabolomics. This will be a continuing process that will extend beyond 2012.
- **2009** – Construct and continue to improve the metabolome catalog of key soybean organs, and processes. These experiments need to be done cognizant of the effects of photoperiod, nutrition and stresses on modulating metabolite levels. Again, this will need to be a continuing activity that will be accelerated by the adoption of new technology and methods, as well as increased access to pertinent chemical standards.
- **2008** – Continue development and improvement of protocols for metabolite extraction and analysis.
- **2010** – Develop and continue to improve a soybean metabolome database. For maximum utility, this database should be functionally linked to other soybean resources (e.g., transcriptome and proteome database and genome sequence).

#### **Sub-Topic B.4 – Transformation/Transgenics (moved to D)**

Note: The bullets in this section were moved to a “new” section - Main Topic D.

#### **Sub-Topic 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 the 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.**

- **2010** – Develop and analyze for a starting set of 6 to 10 QTL.
- **2012** – Develop and analyze for an additional set of 10 to 15 QTL.

##### **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).**

- **2010** – Develop and analyze for a starting set of 6 to 10 traits.
- **2012** – Develop and analyze for an additional set of 10 to 15 traits.

#### **Sub-Topic B.6 – Soybean Producer Expectations - Genomics**

During this session, Ed Ready (representing the USB), David Wright (representing the NCSRP), and Jim Sallstrom (USA-Minnesota Producer and USB Director) provided the discussants with information about what producers expect from genomics research.

USB provides financial support for public research in three main genomics / genetics / breeding areas: namely, research to improve genetic yield potential, greater protection of existing yield potential against abiotic and biotic stress, and improved soybean seed composition. The USB supports public research aimed at improving producer-desired traits in those three areas. USB does not concentrate on increasing genetic yield potential, since the seed companies spend enormous effort doing so, but USB does follow up on promising yield results seen in research on other problems. The USB supports research identifying new germplasm sources to diversify the current USA germplasm, to discover novel alleles for disease resistance, drought resistance or improved seed composition, and to develop new molecular approaches or techniques that enhance the ability of both public and private breeders. This ultimately puts ever higher-yielding, more pest-resistant cultivars into the hands of producers, while simultaneously moving towards a seed composition ultimately desired by soybean seed processors and consumers.

The NCSRP Board has decided to focus its public research support on existing and emerging pest problems (insects, nematodes, fungi, bacteria, viruses) that reduce yields of soybeans. Of particular interest in the North Central region are Aphids, Soybean Cyst Nematodes (SCN), Asian Soybean Rust, Phytophthora, and a number of other diseases.

Producers are interested in ensuring that soybean seed composition meets end-users' needs now and in the future. Soybean seed protein and oil content needs may change, given the increasing use of soybean oil as a source for the rapidly increasing USA market for biofuels. Until now, the soybean market has been driven by the need for protein, with oil as a valuable co-product. However, increased demand for oil may shift the market to an oil-driven market. This could create needs for varieties with higher oil and/or create needs to develop new markets for soybean meal.

## **Main Topic C – Germplasm Genomics**

### **Sub-Topic 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<sub>2</sub>, recombinant inbred lines or backcross populations. As the first step towards the implementation of LD mapping in soybean, an Association Mapping Panel will be developed and genotyped. The Association Mapping Panel will serve as the “control” population with which to compare additional sets of cultivars or germplasm accessions. A mechanism must be developed whereby genotypic (SNP genotypes) and phenotypic information can be collected on large sets of genotypes with various biotic and abiotic stress resistances and quality traits. The resulting phenotypic and genotypic data will be compared with similar data on the Association Mapping Panel for the purpose of gene/QTL discovery.

**C.1.a - Goals:**

- **2007** – Identify the members of the Soybean Association Mapping Panel which will include the following soybean germplasm resources:
  - The 1690 accessions of the USDA Soybean Core Collection identified by Dr. Randall Nelson, USDA, ARS, Urbana, IL.
  - 100 “elite” cultivars.
  - 300 diverse *Glycine soja* accessions
  - 500 additional accessions including:
    - Germplasm accessions with SCN resistance.
    - Germplasm accessions with Phytophthora resistance.
    - Germplasm with other disease resistances.
- **2008** – Identify breeders and other researchers who:
  - Have phenotypic data for specific traits including SCN, Phytophthora, seed size, protein, sugars, fatty acids, amino acids, iron-deficiency chlorosis (IDC), drought, etc.
  - Nominate additional traits for phenotyping and identify genotypes to be phenotyped.
  - Screen the “core” collection with 12,000 SNPs.
- **2009** – Produce seed of “core” collection for a collaborative phenotyping effort:
  - Phenotype nominated lines and the Association Mapping Panel for two nominated traits.
  - Nominate additional traits for screening.
- **2010** – Complete the core collection genotyping
  - Screen “core” collection with 25,000 SNPs
  - Phenotype nominated lines and the Association Mapping Panel for two additional traits

**Sub-Topic C.2 – Track Breeding-Induced Genomic Change**

Soybean introductions and selections from them were grown by U.S. soybean farmers until the early 1940’s when new cultivars were released that relied upon hybridization and selection as the means for new cultivar development. These cultivars demonstrated significant yield increases over the original introductions and selections. A sustained period of yield increases occurred during the period from 1940-1980 and these increases continued to the present as a result of the combined efforts of both public and private breeding programs. 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 time-differentiated successive germplasm pools. It is plausible to suggest that yield improvements are not identical in all cultivars and that specific genetic improvements can be identified that can subsequently be combined into new cultivars with even greater yield potential.

## C.2 - Goals:

- **2007** – Identify genotypes from three time-differentiated germplasm pools for whole-genome SNP analysis. Genotypes will be selected as follows:
  - Commodity type cultivars grown pre-1940 by U.S. producers that were mainly introductions from Asia or selections from introductions.
  - Cultivars released (yield basis) by U.S. and Canadian public breeders from 1940 to 1980 and which were selections from hybridized cultivars.
  - Post-1980 public cultivars.
  - Post-1980 proprietary cultivars (Monsanto, Pioneer, Syngenta, Dairyland, Soy Genetics, and others). Requires development of protocols to protect intellectual property.
  - Consultation with quantitative geneticists/genomicists as to an analysis aimed at detecting a “signature” in the soybean genome that could be ascribed to intense breeder selection for ever-greater grain yield.
- **2008** – Develop the protocols for
  - Genotyping the lines in the above three cultivar groups with 6,000 SNPs.
  - Undertake preliminary QTL analysis to determine the feasibility of yield QTL discovery using this retrospective approach to detect recent selection in the soybean genome.
- **2009** – If the results look promising, then:
  - Genotype the lines with 12,000 SNPs to extend the analysis.
- **2010** – If further discrimination is needed to evaluate individual pedigrees, then
  - Genotype the lines with additional SNPs to bring total to 25,000 SNPs, and thereby complete in detail this retrospective QTL analysis.

### **Sub-Topic 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 17,000 soybean Plant Introductions collected from Asia and other countries which likely contain allelic variants that impact yield *per se* that were not present in the Ancestral Cultivars. Thus, methods must be developed to 1) identify germplasm lines that harbor unique genetic factors that can positively impact the yield of currently grown cultivars and 2) develop procedures to define the yield-enhancing genetic factors so that they can be used 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.

- **2008** – Complete the phenotypic and QTL analyses to define the genome positions of four previously identified yield QTLs.
- **2008** – Initiate the process to move the four yield QTLs to different genetic backgrounds and develop near isogenic lines (NILs) that carry combinations of the four previously identified yield QTLs.
- **2009** – Conduct haplotype analysis of the genome regions associated with the four exotic yield QTLs in a wide spectrum of adapted and exotic germplasm to determine the uniqueness of the newly discovered yield QTLs.
- **2011** – Complete introgression of the favorable alleles at each of the four yield QTLs into different genetic backgrounds.
- **2012** – Collect extensive yield data to assess the effects of the four yield QTLs in different genetic backgrounds including modern elite cultivars.
- **2012** – Collect physiological, EST, and proteomic data on the NILs that carry combinations of confirmed high versus low yield “alleles” at the four QTLs.

#### **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 in exotic germplasm that will enhance yield in elite germplasm. Apply the same foregoing approaches (e.g., those listed above in the C.3.a goals) to these evolving pre-breeding mapping populations, but now use the Illumina 1536-SNP genotyping approach after these populations have been well-phenotyped for yield performance in a broad array of production environments.

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

- **2008** – Negotiate with Chinese officials to obtain seed of *ca.* 400 Chinese cultivars that were released since 1996 and add to our USA germplasm banks.
- **2009 to 2010** – Initiate seed increases and subsequent yield performance trials of those cultivars that have little or no USA components in their pedigrees.
- **2011** – Generate allelic profiles of the selected Chinese cultivars using 1536 SNP markers to conduct a “diversity contrast” with U.S. ancestral cultivars. Select the most promising Chinese varieties to begin pre-breeding efforts.
- **2012** – Initiate crosses between the “best” Chinese cultivars the “best” U.S. cultivars.

#### **Sub-Topic 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. One model that might be pursued is that used by U.S. small grain breeders who have four federally-funded regional genotyping centers that conduct marker analyses for a number of genes that provide disease resistance. Another approach would be to identify one or a few laboratories with experience in marker assisted selection (MAS) to serve as genotyping centers. Additional needs include the development of “breeder friendly” MAS assays that can be easily used by soybean

breeders in their laboratories, and the training of breeders and their laboratory assistants in the use of MAS.

#### **C.4 - Goals:**

- **2007 to 2012** – Continue outreach efforts to inform the breeding community of newly available effective and economical MAS and genotyping platforms.
- **2008** – Complete these activities:
  - Conduct a survey to determine the needs for MAS and genotyping in the public sector and small seed companies.
  - Develop workshops for MAS training.
  - Begin to develop breeder friendly SNP marker and genotyping assays for genes/QTL: 10 assays - place in Soybean Breeders Toolbox.
- **2010** – Continue development of breeder friendly SNP marker and genotyping assays for genes/QTL: 50 assays - place in Soybean Breeders Toolbox.
- **2012** – Continue development of breeder friendly SNP markers and genotyping assays for genes/QTL: 100 assays - place in Soybean Breeders Toolbox.

#### **Sub-Topic C.5 – Germplasm Genomics Informatics**

Soybean has a rich set of germplasm consisting of landraces and other related species. One outcome of the genome sequencing project is that we will be able to more efficiently and precisely tap some of the rich genetic resources to address key issues and constraints in soybean production. Moreover, as SNP markers are mapped on a multitude of cultivars and accessions from the germplasm collection, there is a need to collate this information into a central repository that is both accessible, interpretable, and relatable to the informatic needs described earlier in Sub-Topics A.1 and B.1 (see above).

##### **C.5.a - Development of HapMap Browser**

- **2008** – Create a beta version of HapMap Browser:
  - To permit viewing of haplotypes across multiple genotypes.
  - To provide information about assay.
- **2008** – Establish a pipeline for moving data to relevant databases (details developed by committee – see section A.1).
- **2009** – Put the HapMap Browser on-line, containing ~6 k SNP containing fragments and ~ 9 k SNPs.
  - Integrate with SoyBase and Soybean Breeders Toolbox.
- **2010 to 2012** – Continue to populate the browser.

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

- **2009** – Develop a whole-genome array or equivalent technology to measure both copy number variation and expression data.
- **2010** – Create a color-coded expression ‘heat map’ that can be overlain onto HapMap and the Genetic Map.

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

- **2008** – Establish a strategy and develop/acquire software tools to accomplish this.

#### **C.5.d - User interface for informatics tools**

- **2007** – There was a great deal of discussion regarding user interface, querying abilities, etc. It was suggested that database developers and soybean breeders have a future meeting to establish priorities for interface development.

#### **C.5.e - Large datasets**

- **2009 to 2010** – Begin to include relevant germplasm data from other species, beginning with other *Glycine* species and then *Phaseolus vulgaris*. This could include cross-species markers and physical maps. A long-term plan for inclusion of trait data in those dataset should be implemented.
- **2009** – Develop tools for sequence-based cross-species searching, not synteny-based only.

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

- **2008** – Begin developing a plan for the long-term curation of the genome sequence of soybean. Should include plans for updates on annotation, correction of assembly errors and incorporation of other relevant data.

**Note:** Additional annotation/informatics bullets were developed in the A-1 and B-1 sub-topic sessions, so go to the Main Topic A and B sections to view these bullets.

#### **Sub-Topic C.6 – Transformation/Transgenics (moved to D)**

Note: The bullets in this section were moved to a “new” section - Main Topic D

#### **Main Topic D – Transformation/Trangenics**

Soybean transformation has shown significant improvement and enabled public and private sector production of commercial cultivars with transgenic traits. 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 has met the 2007 benchmark of being able to produce 500 plants per person per year. 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) using bean mottle mosaic virus (see subtopic B-2).

**Sub-Topic D.1 – Create a Transgenic Event Repository.**

- **2008** – Identify a repository site by 2008 and responsible director to provide public access to transgenic events including insertional mutant collections and released transgenics. Seek seed funding for the initial planning.
- **2009** – Seek financial support for the establishment of a stock center/ repository with continuing support from USDA-ARS or other sources. Community consensus was that a cost-recovery model is not sustainable and, therefore, continuing external support will be essential. One possibility is to expand the services offered through the University of Illinois stock center.

**Sub-Topic D.2 – Create a Virtual Center for Transgenics/Transformation.**

- **2008** – Host a planning meeting for the creation of virtual center to interconnect transformational programs in the soybean community. Do this in conjunction with the Soy2008 meeting that will be held in Indianapolis, July 20-23, 2008.
- **2009** – Establish virtual center. One charge of the center will be to serve as a repository for vectors and promoters. It will also provide guidance on shipping permits and other regulatory matters. The Center will be available to the community to support a variety of projects. The Center will provide greater efficiency and capacity enabling larger scale projects.

**Sub-Topic D.3 – Establish A Soybean Regulatory Promoter Set.**

Promoter (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.

- **2008** – Identify 100 promoters for seed (maturation and germination), leaf, root, nodule, biotic and abiotic induced stress.
- **2009** – Identify the ‘unique’ soybean genes from the completed genome. From this gene list, chose 1,000 genes, including transcription factors and other proteins of interest.
- **2012** – Using the above set of 1000 genes, create a set of RNAi, insertion, forward and reporter constructs. Provide these to the soybean stock center for distribution.
- **2010** – Complete construction of cassettes and accompanying reporter cassettes. These resources would be distributed by the soybean stock center.

**Sub-Topic 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 on 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.

- **2008** – Need to re-evaluate growth of the mini-max soybean genotype as a possible solution to space limitation. Current data suggests that mini-max is poor for embryogenesis. However, it may be suitable for the cot node approach. There is some confusion about the uniformity of the small growth habit of mini-max and this should be resolved.
- **2008** – Establish soybean community standards for an excellent transgenic field facility. These standards can then be used by the researchers to establish local field facilities for trait evaluation.
- **2010** – Increase the efficiency of the cot node soybean transformation procedure to 6-10%.
- **2010** – Expand the genotype range for somatic embryogenesis to at least 10 genotypes with a minimum of 5% efficiency.
- **2010** – Improve somatic embryogenesis by standardizing the protocols so that phenotyping can be done on well-matched material.
- **2010** – Develop a system for rapid analysis of transgenes in somatic embryogenesis to modify seed composition (e.g., seed oil, amino acid content, etc).
- **2010** – Develop high-through-put screening methods for soybean seed traits that can be used to evaluate transgenes.
- **2010** – Develop simple assays for seed traits that can be used to evaluate transgenes.
- **2010** – Attempt to target one biochemical pathway and evaluate schemes for maximum expression of the product of at least a 4-gene pathway.
- **2010** – Need to develop new selectable markers for soybean transformation.

---

#### **Report Writing Team**

**Assembler - James Specht, University of Nebraska, Chair, SoyGEC**

**Introduction - Randy Shoemaker, USDA-ARS, Ames, IA, past member SoyGEC**

**Topic A Write-Up - Scott Jackson, Purdue University, SoyGEC**

**Topic B Write-Up - Gary Stacey, University of Missouri, SoyGEC**

**Topic C Write-Up - Perry Cregan, USDA-ARS, Beltsville, MD, (past) SoyGEC**

---

#### **Acknowledgements**

The Meeting Participants thank the United Soybean Board for its generous funding support for the Hotel Meeting Room and Equipment, for the morning and afternoon refreshment breaks, for the two lunches, and for the one evening dinner meal.

The Meeting Planners and Organizers also thank Ann Chase (the Smith-Bucklin representative for the USB) for the assistance she provided relative to the negotiations and finalization of the hotel contract.

We also thank Ed Ready (Smith-Bucklin, representing the USB), David Wright (representing the North Central Soybean Research Project), and Jim Sallstrom (Minnesota soybean producer and USB member) for providing their perspectives with regard to what traits should be targeted in genomics research.

### Appendix – Meeting Participants

First Name	Last Name	E-mail Address	Institution
James	*Specht	<jspecht1@unl.edu>	Univ. of Nebraska
Randy	*Shoemaker	<rcsshoe@iastate.edu>	ARS-USDA, Ames, IA
Scott	*Jackson	<sjackson@purdue.edu>	Purdue University
Gary	*Stacey	<staceyg@missouri.edu>	Univ. of Missouri
Perry	*Cregan	<creganp@ars.usda.gov>	ARS-USDA, Beltsville, MD
Diane	Bellis	<dbellis@agsourceinc.com>	Information Contractor with USB
Kristin	Bilyeu	<bilyeuk@missouri.edu>	ARS-USDA, Columbia, MO
Roger	Boerma	<rboerma@uga.edu>	Univ. of Georgia
Yung-Tsi	Bolon	<hsie0024@umn.edu>	ARS-USDA, St. Paul, MN
Glenn	Bowers	<glenn.bowers@syngenta.com>	Syngenta
Ed	Cahoon	<ecahoon@danforthcenter.org>	ARS-USDA, Danforth Center, St. Louis, MO
Steve	Cannon	<scannon@iastate.edu>	ARS-USDA, Ames, IA
Tommy	Carter	<tommy_carter@ncsu.edu>	ARS-USDA, Raleigh, NC
Tom	Clemente	<tclemente1@unl.edu>	Univ. of Nebraska
Steve	Clough	<sjclough@uiuc.edu>	ARS-USDA, Urbana, IL
Brian	Diers	<bdiers@uiuc.edu>	Univ. of Illinois
Anne	Dorrance	<dorrance.1@osu.edu>	Ohio State Univ.
Jeff	Doyle	<jjd5@cornell.edu>	Cornell Univ.
Michelle	Graham	<magraham@iastate.edu>	ARS-USDA, Ames, IA
David	Grant	<dgrant@iastate.edu>	ARS-USDA, Ames, IA
Eliot	Herman	<eherman@danforthcenter.org>	ARS-USDA, Danforth Center, St. Louis, MO
Matthew	Hudson	<mhudson@uiuc.edu>	Univ. of Illinois
David	Hytten	<hytten@ars.usda.gov>	ARS-USDA, Beltsville, MD
Karen	Kaczorowski	<kkaczoro@purdue.edu>	ARS-USDA, W. Lafayette, IN
Warren	Kruger	<warren.m.kruger@monsanto.com>	Monsanto
David	Lightfoot	<ga4082@siu.edu>	Southern Illinois Univ.
Jianxin	Ma	<maj@purdue.edu>	Purdue University
Saghai	Marroof	<smarroof@vt.edu>	Virginia Tech
Greg	May	<gdm@ncgr.org>	NCGR
Khalid	Meksem	<meksemk@siu.edu>	Southern Illinois Univ.
Henry	Nguyen	<nguyenhenry@missouri.edu>	Univ. of Missouri
Paula	Ohloft	<paula.olhoft@basf.com>	BASF
James	Orf	<orffx001@umn.edu>	Univ. of Minnesota
Wayne	Parrott	<wparrott@uga.edu>	Univ. of Georgia
Ed	Ready	<eready@smithbucklin.com>	Smith Bucklin (represents USB)
Jim	Sallstrom	<jimsall@means.net>	MN Producer, USB Member
Shannon	Schlueter	<acissej@purdue.edu>	Purdue University
Jessica	Schlueter	<sds@purdue.edu>	Purdue University
Monica	Schmidt	<mschmidt@danforthcenter.org>	ARS-USDA, Danforth Center, St. Louis, MO
Daria	Schmidt	<daria.schmidt@pioneer.com>	Pioneer
Jeremy	Schmutz	<jeremy.schmutz@stanford.edu>	Stanford Univ.
Sam	Sparace	<smsprc@clemson.edu>	Clemson Univ.
Robert	Stupar	<stup0004@umn.edu>	Univ. of Minnesota
Jay	Thelen	<thelenj@missouri.edu>	Univ. of Missouri
Chris	Towne	<cdtowne@tigr.org>	TIGR
Carroll	Vance	<vance004@umn.edu>	ARS-USDA, St. Paul, MN
Lila	Vodkin	<l-vodkin@uiuc.edu>	Univ. of Illinois
David	Wright	<dwright@iasoybeans.com>	NCSRP representative, Ames, IA