SOYBEAN GENOMICS
WHITE PAPER
JANUARY, 2000

Table of contents

I. EXECUTIVE SUMMARY

On October 21 and 22, 1999, seventeen expert researchers with knowledge of genomics, DNA markers, plant transformation, and bioinformatics participated in a workshop hosted by the United Soybean Board Production Committee. Over the course of the two days, the scientists reached consensus on research priorities and on time required to reach benchmarks if the work is funded.

A. Molecular markers
1. Develop 1000 Simple Sequence Repeat (SSR) markers (bringing to 2000 the number of SSRs in the public domain) within 3 years.
2. Expand the number of Single Nucleotide Polymorphism (SNP) markers to 10,000 within three to five years.
3. Characterize allelic variation in major candidate genes in less than three years and determine variation for all significant genes in five to ten years.

B. Transformation
1. Improve the efficiency of soybean transformation by five- to ten-fold in three years.
2. Develop technology to precisely deliver DNA within three to five years.
3. Develop transgenic screens to elucidate gene function within three years or less.

C. Structural and Functional Genomics
1. Tag 80% of the genes in soybeans within three to five years.
2. Develop and integrate the genetic, physical and transcript maps of soybeans within three to five years.
3. Assign biological function to identified soybean genes in five or more years.
4. Use comparative genomics to understand soybean interaction with pathogens and symbionts in five or more years.
5. Further identify detailed bioinformatics needs of soybean genomics researchers as a starting point for development of software tools to support the soybean genomics program in less than one year.

II. IMPROVEMENTS IN SOYBEANS COME FROM SCIENTIFIC DISCOVERIES

Scientific discoveries provide the foundation for all technological innovation. Technological advances have enabled U.S. soybean producers to achieve unparalleled production efficiency. During the past 75 years average soybean yields have more than tripled from 12 bushels per acre in 1924 to nearly 40 bushels per acre in recent years. Although the annual rate of yield increase for this entire period has averaged 0.34 bushels per acre, it has accelerated to an average of 0.5 bushels per acre per year since 1972. At least half of this is attributable solely to genetic improvement through breeding. These increases followed rapid producer adoption of new technologies that have emerged from soybean research and development by both public institutions and private industry. Often, a 5- to 10-year lag occurs between a scientific discovery and its development into a technology usable by producers; the amount of research, however, directly determines the speed at which new technologies are put into the hands of soybean producers.

As successful as past soybean improvement efforts have been, the pace of innovation must accelerate to keep U.S. soybean production globally competitive and to meet the demands of an increasing world population. One of the rapidly developing, new technologies with potential to greatly accelerate soybean improvement is the broad area of genomics. Genes control all metabolic processes required to sustain life. Until recently, soybeans could only be crossed with closely related plants to obtain new genes. Wild relatives and other species are also sources of useful genes, but in many cases, the reproductive specificity of soybeans prevent effective utilization of other plant genes. This document contains consensus recommendations of experts on research needs for technological advances in five areas: soybean molecular markers, transformation, structural genomics, functional genomics, and bioinformatics. For the purposes of this paper, these five areas are collectively called genomics.

Molecular markers, like sign posts and road maps, and plant breeders in incorporating new genes into improved soybean cultivars. Development of additional robust markers will accelerate this process. Some limitations can be addressed by genetic transformation and by the transfer of new genes from any source into plants. Because genome characterization is now a high throughput process, the next few years will bring a rapid increase in the number of new gene sequences identified through structural genomics. However, the functions for most of these genes will not be known. The successful characterization of plant genetic systems and functions will yield technologies with better access to the genes that control agronomic traits of economic importance, such as general disease and insect resistance, maturity, and soybean quality. These technologies will enable researchers to modify plant characteristics to provide specific oil and protein composition, enhanced disease resistance, and more consistent plant responses to variations in environmental conditions. New bioinformatic tools will be necessary to analyze these new types of data as they are collected.

III. PRIORITY RESEARCH FOR MOLECULAR MARKERS

A. Development of Simple Sequence Repeat (SSR) Markers

SSR markers are single locus markers with multiple alleles. They are presently the DNA marker of choice in soybeans because of their simplicity and their many alleles, which enables detection of polymorphism among elite cultivars and breeding lines. SSRs are well distributed throughout the soybean genome. These markers have broad application and can be used in a range of laboratories for both sophisticated and unsophisticated analyses. About 1000 SSR markers were available to the public at the end of 1999. To obtain adequate coverage for quantitative trait loci (QTL) analysis and marker assisted breeding applications, development of an additional 1000 SSRs is required (total of 2000 SSRs). **TIME LINE: 3 years**

B. Develop Single Nucleotide Polymorphism (SNP) Markers

SNPs are mutations that occur throughout the soybean genome and can serve as molecular markers. Researchers have made rapid progress in the development of SNP markers in the human genome and the cost of using these markers continues to decline. SNPs will become the ultimate low-cost, high throughput molecular marker systems in the future. SNPs are clearly the most abundant marker but they will be available in soybeans. Currently there are few SNPs in the public domain. The number of SNP loci available to public researchers must be expanded to 10,000. **TIME LINE: 3-5 years**

C. Characterizing Allelic Sequence for Important Genes

An important application of the "uniqenes" (a set of important soybean genes in which each gene is present only once) discovered in the soybean genomics project is the determination of the scope of allelic variation for genes conditioning economically important soybean traits. This variation should be characterized by the development of gene specific SNPs. These SNPs will be use by plant breeders for parent identification and selection of progeny with superior phenotypes. **TIME LINE: Allelic variation in major "candidate" genes can be characterized in less than 3 years while the determination of variation for all significant genes will require 5 to 10 years.**

IV. PRIORITY RESEARCH FOR TRANSFORMATION

A. Improve the Efficiency of Soybean Transformation by 5- to 10-fold

To fully take advantage of genomic knowledge and the production of transgenic cultivars, soybean transformation systems must become more efficient. At the present time, soybean transformation lags behind that of other major crops. Efforts should therefore be increased to develop improved transformation methodologies for soybeans. The goal is to increase the efficiency of soybean transformation 5- to10-fold (one person producing 300 transgenic lines per year). Substantial improvements can be made by further modifications to the existing systems as well as the development of novel approaches. For now as well as existing systems, there is a need to better understand the factors that influence the induction and regeneration of soybean tissue cultures. In addition, testing of new gene promoters, selectable markers, and gene coding terminators can lead to increases in transformation rates. The availability of tissue specific gene promoters also will increase the range of traits that can be improved by genetic engineering. **TIME LINE: 3 years**

B. Technology to Precisely Deliver New DNA

Current transformation methods deliver DNA randomly and imprecisely into the soybean genome. Due to these unpredictable insertion patterns, many transgenic plants do not express the phenotype in a predictable and stable manner. In addition, it would be useful to remove selectable markers and other "carry-along" DNA after they are no longer needed. The goal of this research is to develop technology for delivering DNA precisely into the soybean genome. This precision will include site-specific insertion, single copy insertion, and "clean delivery" in terms of providing the option to remove unneeded "carry-along" DNA. This technology will furthermore lead to the precise integration of long DNA inserts, such as a bacterial artificial chromosome (BAC) clone, and the opportunity to conduct multiple gene integration into the same location (i.e., directed "gene stacking"). Precise integration technology should be developed to be adaptable for a wide range of transformation methods, including direct DNA delivery and Agrobacterium-based transformation. The immediate outcome will be to recover a much higher percentage of transgenic plants where stable expression is found. Additionally, a future outcome is that the site-specific integration process by itself will enhance the frequency of stable DNA incorporation. **TIME LINE: 3-5 years**

C.   Develop Transgenic Screens to Elucidate Gene Function

As the term implies, assignment of function to genes is the ultimate goal of functional genomics, and for some gene categories, such assignment can only be definitively confirmed by testing the function of cloned genes in a plant itself. Strategies to elucidate gene functionality in plant systems that should be pursued include:
i)   the ability to routinely engineer plants with bacterial artificial chromosome (BAC) clones;
ii)   the ability to elucidate gene functionality by using T-DNA tagging systems to disrupt genes and study the subsequent up- and down-regulation of gene expression;
iii) the development of a viral-based transformation system, which would allow the screening of large numbers of genes via their transient expression; and
iv)   the development of a heterologous transposon tagging system for soybeans.
TIME LINE: 3 years or less

V. PRIORITY RESEARCH IN STRUCTURAL AND FUNCTIONAL GENOMICS

A. Tag 80% of the Genes in Soybeans

Gene discovery is a primary research priority in the field of genomics. It is the foundation of all functional analyses and is the ultimate target of most structural and physical genetic analysis. A number of strategies can be used to accomplish this goal. Currently, the most straightforward approach is to clone and partially sequence the messages from expressed genes. This type of project is called an expressed sequence tag (EST) project and is presently underway. The advantage of this approach is that each sequence represents a gene product. Continuation of this effort is a high priority. Other strategies such as:
i)   bacterial artificial chromosome (BAC)-end sequencing;
ii)   genomic sequencing; and
iii)   evaluation of proteins (proteomics);
may become viable or more efficient methods to accomplish this research objective. The advantage of these approaches is that they may identify genes not discovered in an EST project. The relative efficiency of each approach must be continually evaluated so that gene discovery proceeds at an optimal rate. **TIME LINE: 3-5 years**

B. Develop and Integrate the Genetic, Physical, and Transcript Maps of Soybeans

An integrated soybean genome map would increase the efficiency of crop improvement through application in functional genomics, maker assisted breeding, and transformation. The goal is to enhance the existing physical map of the soybean genome until it is more than 95 percent complete by incorporating genetic markers and the majority of identified genes, ESTs, and open reading frames (ORFs). Specific needs include:
i)   improvements in BAC contig generation and reliability,
ii)   integration of the physical map with existing genetic markers and newly developed SNP markers, and
iii) integration of transcripts, cDNAs, ESTs and BAC-derived ORFs with contig, genetic, and physical maps.
TIME LINE: 3-5 years

C. Assign Biological Function to Identified Soybean Genes

The purpose of assigning function is to discover the genes of agronomic importance. The assignment of function to genes proceeds at several levels:
i)   Determine the expression patterns of genes in tissues and organ systems of the plant by measuring the expression of thousands of genes at a time (i.e., "global" expression patterns). Expression comparisons under conditions including pathogen challenge, heat, cold, and drought stresses, and nutrient limitations will yield classes of genes involved in these critical processes. Expression profiles of many agronomically important genotypes containing traits of economic importance and QTL will also aid in assigning function. Expression profiling will also yield the information needed to select promoters useful for plant transformation.
ii)   Compare the soybean coding regions to the vast amount of sequence data from other organisms (especially plants) so as to determine possible functions. Metabolic reconstruction of complete pathways, especially those unique to plants, is a goal.
iii) As the tools of proteomics continue to be developed, these should be employed as needed to complete the functional assignments to expressed genes; and
iv)   Information from gene transfer to test function in transformed plants. The use of this method for soybeans will depend on whether more efficient transformation of soybeans and other plant systems is available.
TIME LINE: 5 or more years

D. Soybean Interaction with Pathogens and Symbionts

Using an understanding of the complexity of the soybean genome to begin to unravel the functional integration of component genes and gene products will enable research on the signaling and interaction between genomes. Soybean interactions with other organisms, e.g., *Bradyrhizobium japonicum, Mycorrhizae*, soybean cyst nematode (SCN) greatly affects performance. Understanding the soybean genomic components influencing and influenced by that interaction will provide a view into the genomes of the interacting partners. In some cases, whole genomic analysis will be more feasible as in *B. japonicum*. In others, comparative genomics, e.g., between SCN and *C. elegans*, will identify potential pathogenicity targets for SCN control. Single resistance and virulence genes operate in a matrix of integrated gene expression. Comparative genomics will help us understand gene relationships within organisms and the genomic control of inter-organism interactions. **TIME LINE: 5 or more years**

VI. PRIORITY RESEARCH IN BIOINFORMATICS

Genomics projects, by their nature, require the collection, storage, and analysis of many data points (i.e., sequences, expression levels, map positions). Much of this can realistically be accomplished only through the use of computers. Informatics components can be separated into the development of infrastructure and tools, and the application of those tools to synthesize information into useable results. Infrastructure needs include the development of relational database management systems, visualization tools, algorithm development, distributed computing, storage systems, and networking. Information integration is a biological problem, which includes pathway reconstruction, understanding of developmental processes, and integration of multiple phenotypic information. Databases and analysis programs are not ends in themselves but are rather essential tools for accomplishing the research goals identified above. Thus, bioinformatics must be considered as an integral part of all genomics projects. With this in mind, the expert panel proposed that a workshop to identify *detailed* bioinformatics needs of soybean genomics researchers should occur within the next several months. Bioinformatics experts will need the benefit of that workshop's discussions to start the development of software tools to support the soybean genomics program.

Appendix A

On October 21 and 22, 1999, seventeen expert researchers with knowledge of genomics, DNA markers, plant transformation, and bioinformatics participated in a workshop hosted by the United Soybean Board Production Committee. Over the course of the two days, the scientists reached consensus on a list of research priorities in the area of soybean genomics.

The workshop was planned by: Dr. Dwayne Buxton, National Program Leader for the Agricultural Research Service in Oilseeds and Bioscience; Dr. Roger Boerma, Research Professor and Coordinator of the University of Georgia Center for Soybean Improvement; Maureen Kelly of AgSource, Inc., a subcontractor with the United Soybean Board focusing on Federal Research Coordination; and Kent Van Amburg, Production Committee Manager for the United Soybean Board of Smith, Bucklin and Associates. Elizabeth Vasquez of MCA Consulting facilitated the workshop.

Appendix B

Soybean Genomics Workshop Participants

Tom Clemente
University of Nebraska – Lincoln
E324 Beadle Center
Lincoln, NE 68588-0665
Telephone: (402) 472-1428
Fax: (402) 472-3139
E-mail: tclemente@unl.edu

John J. Finer
OARDC
Ohio State University
1680 Madison Avenue
Wooster, OH 44691
Telephone: (330) 263-3887
Fax: (330) 263-3887
E-mail: finer.1@osu.edu
David A. Lightfoot
Southern Illinois University
Department of Plant & Soil Sciences
Carbondale, IL 62901-4415
Telephone: (618) 453-1797
Fax: (618) 453-7457
E-mail: ga4082@siu.edu

Joe Polacco
University of Missouri
17 Schweit Hall
Columbia, MO 65211
Telephone: (573) 882-4789
Fax: (573) 882-5635
E-mail: polaccoj@missouri.edu

Randy C. Shoemaker
USDA-ARS
Department of Agronomy
Iowa State University
Ames, IA 50011
Telephone: (515) 294-6233
Fax: (515) 294-2299
E-mail: rcsshoe@iastate.edu
Lila Vodkton
University of Illinois
Department of Crop Sciences
384 ERML
1201 W. Gregory Drive
Urbana, IL 61801
Telephone: (217) 244-6141
Fax: (217) 333-4582
E-mail: l-vodkin@uiuc.edu

Vergel C. Concibido
Monsanto Company
379 Birchwood Crossing Lane
Maryland Heights, MO 63043
Telephone: (314) 694-1231
Fax: (314) 694-3914
E-mail:
vergel.C.Concibido@monsanto.com
David Grant
USDA-ARS & Iowa State
University
G104 Agronomy Hall
Ames, IA 50011
Telephone: (515) 294-1205
Fax: (515) 294-2299
E-mail: dgrant@iastate.edu
David Ow
ARS-- USDA
Plant Gene Expression Center
800 Buchanan Street
Albany, CA 94710
Telephone: (510) 559-5909
Fax: (510) 559-5678
E-mail: ow@pgec.ars.usda.gov

Bob Reiter
Monsanto Company
3302 SE Convenience Blvd.
Ankeny, IA 50021-9424
Telephone: (515) 963-4211
Fax: (515) 963-4242
E-mail: robert.s.reiter@monsanto.com

Jeff Skolnick
Danforth Center
7425 Forsyth
Box 1099
St. Louis, MO 63105
Telephone: 314-615-6931
Fax: (314) 615-6931
E-mail: skolnick@danforthcenter.org
David Webb
SOYGENETICS LLC
7552 NW 28th Street
Ankeny, IA 50021-9424
Telephone: (515) 289-0262
Fax:
E-mail: webbgmax@mindspring.com

Perry Cregan
USDA-ARS
Building 006, Room 100
BARC-West
Beltsville, MD 20705
Telephone: (301) 504-5070
Fax: (301) 504-5718
E-mail: pcregan@asrr.arsusda.gov
Ted Klein
DuPont
Stine Haskell 614
P.O. Box 30
Newark, DE 19714-0030
Telephone: (302) 283-2403
Fax: (302) 283-2449
Ted.m.klein@usa.dupont.com
Wayne Parrott
University of Georgia
3111 Plant Science Building
Athens, GA 30602
Telephone: (706) 542-0928
Fax: (706) 542-0914
E-mail: wparrott@arches.uga.edu

Ernest F. Retzel
Director, Academic Computing &
Bioinformatics
Academic Health Center
University of Minnesota
650 Children's Rehabilitation Center
UMHC Box 43
426 Church Street S.E.
Minneapolis, MN 55455-0312
Telephone: (612) 626-0495
Fax: (612) 626-6069
E-mail: ernest@ahc.umn.edu
James E. Specht
University of Nebraska
Department of Agronomy
322 Keim Hall
Lincoln, NE 68583-0915
Telephone: (402) 472-1536
Fax: (402) 472-7904
E-mail: jspecht1@unl.edu

Workshop Organizers

Kent Van Amburg
Smith, Bucklin & Associates, L.L.C.
540 Maryville Centre Drive
Suite LL5
St. Louis, MO 63141
Telephone: (314) 579-1588
Fax: (314) 579-1599
E-mail: kent_van_amburg@sba.com

H. Roger Boerma
University of Georgia
Dept. of Crop & Soil Sciences
Athens, GA 30602-7272
Telephone: (706) 542-0927
Fax: (706) 542-0560
E-mail: rboerma@arches.uga.edu

Maureen Kelly
AgSource, Inc.
Subcontractor, United Soybean
Board
600 Pennsylvania Avenue SE
Suite 320
Washington, DC 20003
Telephone: (202) 969-8902
Fax: (202) 969-7036
E-mail: mkelly@gordley.com

Dwayne Buxton
USDA – Agricultural Research
Service
5601 Sunnyside Avenue
Beltsville, MD 20705-5139
Telephone: (301) 504-4670
Fax: 301-504-5987
E-mail: drb@ars.usda.gov