National Program Action Plan for the Peanut Genome Initiative

Research Priorities and Funding Requests for 2007

Version 1.3 July 2007

Introduction

The peanut industry has attempted to use all available methods to reduce cost of production and improve quality of US peanuts. For example, funding has been provided for bioengineered peanuts since 1989. Some of these bioengineered products show promise for improving the productivity, protection and quality of U.S. peanuts, but have not advance to commercial production. However, the U.S. peanut industry recently has conducted a re-evaluation of this technology. This assessment was presented in the "*Biotech Peanut White Paper: Benefits and Issues*" at the winter meeting of the American Peanut Council, December 8, 2006 in Atlanta GA. During the discussion, it was recognized that international competitors in the peanut market (i.e., China, India) are developing bioengineered peanuts and will be releasing them in the next 3-5 years. In addition, this technology could lead to significant improvements in the cost of U.S. peanut production, nutrition and overall product quality. This technology also could provide the industry with greater flexibility to bring new traits to the market more rapidly, and to resolve seemingly intractable problems such as peanut allergy. Considering these and other factors, the industry reached consensus that with proper funding U.S. bioengineered peanuts could be commercially available in 5-7 years. However, gaps in U.S. peanut research facilities, research equipment needs and the availability of a trained workforce could impede progress toward that expectation. The following genomic technologies were identified as essential components of a successful effort to deliver agronomic bioengineered peanut cultivars in the expected time frame.

- ***Improved utility of genetic tools for peanut genomics research by developing useful gene markers.*** A goal of developing 150,000 expressed sequence tags was set for completion by 2008. Upon completion of the databases, marker assisted programs will be required for each trait of interest. Included in these projects will be population development so testing materials will be available to answer appropriate questions.

- ***Improved technologies for gene manipulation in genomes by developing useful transformation methods for functional genomic research in peanut.*** Although plant transformation techniques have been developed for peanut, procedures are cumbersome and time-consuming. Developing improved technologies will require additional funds to produce useful contributions for plant improvement projects.

- ***Frameworks for assembling the peanut genome by identifying and integrating the positions of expressed genes on genetic, transcript and physical maps.*** After expressed sequence tag (EST) libraries and markers are developed, genetic maps can be developed. Because the peanut has two genomes, maps will likely be required for the cultivated species and at least one of the diploid progenitors. Research to develop physical maps using BAC libraries will require a significant amount of DNA sequencing and labor.

- ***Improved knowledge of gene identification and regulation by providing baseline data and tools that facilitate the association of DNA-sequences in gene-rich regions of the peanut genome with a biological function.*** A detailed analysis of genomic regions containing genes and associated regulatory elements (promoters and enhancers) is essential to understanding the full genetic potential of peanut as a healthful and profitable food crop. Polyploidy, resulting from the joining of two genomes, increases the probability of two or more copies of many genes typically are present. Thus, specialized approaches must direct the accurate alignment and assignment of the sequence for each gene isoform to the proper chromosome in assembly of the peanut genome from a random assortment of DNA sequence fragments.

- ***Application of 'reverse genetics' to explore gene function will require developing specific populations*** **(**TILLING populations). Gene discovery from TILLING, for allergen identification or regulation of seed composition and quality, will facilitate determination of candidate gene function, hence the identity of the gene product or trait of interest.
- ***Understanding and overcoming immunological problems associated with peanut proteins.*** Making progress in this area will require good genomic databases to be assembled from which proteomic research can be conducted. Construction of proteomic maps and determining the effect of gene knockouts and substitutions for allergen genes on peanut allergy and seed composition will significantly enhance strategies for solving peanut allergen problems.
- ***Bioinformatic management of peanut biological information resources by establishing an interactive system for public distribution of data and information.*** USDA-ARS is funding a Legume database at the National Center for Genomic Research, Santa Fe, NM. To utilize this database on a continuing basis will require development of an *Arachis* database that in compatible with the Legume Information Network.

The *Strategic Plan for the Peanut Genomics 2004-2008* provides the peanut research community with a foundation for a comprehensive and integrated research approach to address these issues. Relevant performance measures or research objectives also are documented in the *National Program Action Plan for the Peanut Genome Initiative*. High priority performance measures for the initial necessary steps that will enable the timely development and delivery of bioengineered peanuts to the U.S. industry are presented in this document, *Research Priorities & Funding Request for 2007*. These statements define the actions that will be taken to solve a given problem, describe what is promised or will be produced, assign accountability for the work to be accomplished, indicate the anticipated benefits of the proposed work, and provide an estimate of year-one funding needs.

Research Need Summary:

Performance Measures for Genetic Tools & Breeding Methods

1.1 Develop a molecular map of the peanut genome
Year One Funding Need: Funded by NRI (Steven Knapp, coordinator)

1.2. Develop a comprehensive set of genetic markers for peanut
(SNP marker development for 3-5 key traits)
Year One Funding Need: $500,000

1.5. Develop genomic comparisons among legume species
Year One Funding Need: Funded by NSF (Doug Cook, coordinator)

Performance Measures for Plant Transformation Technology

2.1 Improve transformation efficiency and utility for peanut
(Expedited transformation system)
Year One Funding Need: $200,000

2.2 Improve transformation technology for functional genomics
(GM trait development)
Year One Funding Need: $150,000

2.3 Develop transgenic screens to understand gene function
(Gene insertion)
Year One Funding Need: $150,000

2.4 Develop biotech risk assessment and mitigation strategies
(Licensing and intellectual property rights)
Year One Funding Need: $75,000

Performance Measures for Genome Sequencing & Gene Discovery

3.1: Advance discovery and knowledge of abundantly-expressed genes in peanut
(Develop microarrays from markers)
Year One Funding Need: $133,000

Performance Measures for Functional Genomics & Proteomics

4.1 Proteomics research to characterize seed protein composition and function
(Develop proteomic maps, gene knock outs & substitutions to improve protein for edible and Biofuel)
Year One Funding Need: $225,000

4.2 Application of 'reverse genetics' to explore gene function
(Develop TILLING populations & collection distribution systems)
Year One Funding Need: $300,000

Performance Measures for Bioinformatics

5.3 Assure Peanut Genomic Resources are included in the Legume Information System
(Peanut genome database with links)
Year One Funding Need: $100,000

Total Year-One Funding Request: $1,833,000

Peanut Genome Initiative Research Priorities for 2007

Genetic Tools & Breeding Methods

Arachis hypogaea is a domesticated crop species that contains two different sets of chromosomes. These two genomes can be differentiated with DNA markers. Although a large amount of genetic variation is evident among accessions of cultivated peanuts, few molecular genomic tools are available to mine useful genes or improve the efficiency of genetic enhancement of peanut through plant breeding. These molecular markers are associated with resistance to only a few peanut diseases and insects. There is critical need to expand the scope of DNA marker coverage of entire peanut genomes, and to use DNA marker systems either from peanut and other legume crops to facilitate peanut crop improvement.

Goal: Improve the utility of genetic tools for peanut genomics research by developing useful gene markers.

Performance Measures:

1.1 Develop a molecular map of the peanut genome. Developing molecular maps for peanut is critical for identifying linkage relationships, comparative mapping among legumes, and better utilizing the genetic resources within the cultivated and related species. However, mapping the tetraploid peanut is extremely difficult.

Research Need: An alternative genetic map of a closely related diploid species to facilitate construction of a genomic map of *A. hypogaea*. New populations will be developed to adequately map the peanut genomes. To avoid complications associated with high levels of sterility, mapping within the A- and within the B-genomes at the diploid level will allow easier interpretation of data. Progenitor species of *A. hypogaea* will also allow diploid maps to more easily be transferred to the tetraploid species.

Anticipated Products:

- Useful populations for creating separate diploid maps with A- and B-genome species; and populations for inbred line selection with different A. hypogaea varieties

1.2. Develop a comprehensive set of genetic markers for peanut. Simple-sequence repeat (SSR) markers and other types of molecular marker systems such as single nucleotide polymorphisms (SNP) are valuable genetic tools for the identification of useful polymorphisms (mutations) in genes in *A. hypogaea* and wild species of the genus *Arachis*. Markers have utility in the characterization of candidate genes for specific traits from raw DNA sequence data, mapping the organization of the peanut genomes, anchoring physical maps of the genomes to genetic maps, and in improving the efficiency and effectiveness of peanut breeding.

Research Need: DNA-markers to enable development of genetic maps and comparative analyses of genomic regions that may harbor genes associated with peanut allergy & agronomic traits. DNA markers for specific phenotypic traits will be discovered in a database of expressed sequence tags (EST), the expressed portion of the peanut genome. Selected EST will be associated with bacterial artificial chromosomes (BAC), created with segments of peanut DNA sequence, that contain specific allergen-related genes. ‘Data-Mining’ these EST and comparative genomic analyses with relevant DNA-probes will locate all members of allergen gene families in peanut, and genes for agronomic traits as well.

Anticipated Products:

- Useful sets of SSR and SNP markers for the identification of candidate genes for important genetic traits from genome sequences, genome mapping and translational genomic approaches (marker–assisted selection) in peanut breeding.

STP 1.5 Develop genomic comparisons among legume species
EST resources provide a means to conduct comparative functional genomic analyses across genera and plant families such as *Glycine, Medicago* and *Phaseolus.* In lieu of advanced genomic data, this resource will improve analysis of the peanut genome.
Research Need: Studies are needed to investigate features that distinguish the major legume lineages from one another. Targeted sequences will be tagged to investigate traits unique to peanut, ones similar to other genera of legumes to compare species for gene discovery, and to investigate the origin of polyploidy and gene expression.
Anticipated Products:

- Test recognition of sequence-tag-sites in peanut with markers from other legume species. Develop reconstituted maps of the peanut genome with bioinformatics resources in LIS

Potential Benefits: Adequate supply of useful DNA-markers and genetic maps will accelerate discovery of gene rich QTL in the peanut genome. Gene markers and genetic maps will enable the design of more effective breeding strategies. Gains in breeding efficiency will enable the simultaneous 'stacking' or 'pyramiding' of multiple genes governing many desired traits that influence peanut allergy, quality, and productivity in elite peanut germplasm.

Plant Transformation Technology

Successful protocols have been developed for peanut transformation using both microprojectile bombardment and *Agrobacterium*-mediated gene transfer. However, it is unlikely that all U.S. laboratories capable of peanut transformation, working in concert, could analyze more than about 10 genes per year. There is critical need to expand the capacity of transformation systems to provide genome information that is essential to the timely achievement of technical advances in peanut enhancement, and the control and manipulation of genes that encode peanut allergens..

Goal: Improve the efficacy of technology for gene manipulation in genomes by developing useful transformation methods for functional genomic research in peanut.

Performance Measures

2.1 Improve transformation efficiency and utility for peanut. *Agrobacterium*-based transformation has certain advantages compared with "naked DNA" transfer procedures, such as microprojectile bombardment. Current protocols for inserting or deleting genes in peanut are limited by low transformation efficiency, and increased time in tissue culture. There is no adequate test system to identify the best recipient genotypes for specific agronomic goals.
Research Need: Protocols for higher efficiency in the recovery of transformed tissues. Peanut transformation has shown significant improvement and enabled production of cultivars with transgenic traits. However, the greatest need for research to enhance transformation efficiency is for applications in functional genomics. Transformation technology is essential for the discovery of gene function via changes in phenotype attributed to the deletion or insertion of specific genes. To meet pending demand for this technology, improvements will be made in areas that help ensure greater efficiency and effectiveness of peanut transformation.
Anticipated Products

- Greater peanut transformation capacity to accelerate the characterization of candidate sequences (genes and regulatory elements) for potential allergens that are generated by a large-scale peanut genomics program.

2.2 Improve transformation technology for functional genomics. Short-term increases in transformation capacity will be accomplished most effectively by developing two or more high-

throughput peanut transformation facilities capable of carrying out numerous simultaneous transformations, employing the most efficient protocols available.
Research Need: Organized collaboration among laboratories to eliminate duplication of effort, and to insure that modifications in transformation protocols are available.
Production of transgenic peanuts carrying genomic elements of interest will be made available as a service to the community of peanut genomic scientists. Regulatory approvals and appropriate agreements among collaborators will be needed to ensure efficient transfer of putative transgenic materials among collaborating institutions.
Anticipated Products

- Ability to generate and evaluate transgenics expressing multiple constructs
- Expanded capacity to test new gene promoters, selectable markers and terminators in peanut genomics programs.

STP 2.3 Develop transgenic screens to understand gene function.
General plant transformation protocols allow the introduction of DNA sequences of more than a few kilobase pairs into cells. The ability to introduce far larger DNA molecules (>100,000 bp) such as bacterial artificial chromosomes (BACs) with stacked genes for desirable traits would be an important tool for germplasm enhancement.
Research Need: Development of a transformation system capable of transferring these large sequences into peanut.
Advances in this technology will utilize BAC libraries of peanut genomic sequences with genes that mediate major constraints in peanut productivity, protection or product quality. This ability will generate transgenic plants that express "knock-out or silenced" genes and transformation events using both DNA array (transcription) and proteomic (translation) approaches.
Anticipated Products

- Germplasm with transgenic enhancement of product quality and safety
- Germplasm with transgenic resistance to diseases and pests

2.4 Develop biotech risk assessment and mitigation strategies. Many of the gene sequences and tools required for producing transgenic plants are subject to patents owned by industry. These tools may require licensing fees for use.
Research Need: Organized negotiation of license agreements with appropriate entities to ensure 'freedom to operate' with proprietary biological and transformation technology.
Transgenic peanuts will be evaluated and approved by governmental agencies charged with oversight of the safety of agricultural products for agriculture, humans, and the environment. Licensing agreements will be obtained for the traits and processes that are protected by patents and necessary for the development of these improved crops.
Anticipated Products

- Freedom to use the gene gun and *Agrobacterium tumefaciens* to transform plants with the desired foreign DNA sequences
- Freedom to use selectable marker genes such as hygromycin phosphotransferase and neomycin phosphotransferase to distinguish tissues containing transgenes
- Freedom to use gene promoters and terminators to turn introduced genes on and off
- Freedom to use genes conferring the traits of interest (e.g. pest resistance, oil quality)

Potential Benefits: Evaluation of existing transformation approaches will determine the optimal protocol that delivers greater efficiency and capacity. Advances in transformation capacity that are gained from the development of new methods will enable transformation of larger DNA sequences (e.g., from BAC libraries), tagging strategies that facilitate the cloning of important

genes by interrupting normal function and by marking genes with identifiable molecular sequences.

Genome Sequencing & Gene Discovery

The nuclear genome of the cultivated peanut is approximately 3 x 10^9 bp, similar to the size of the human genome. The peanut genome contains about 50,000 genes that collectively determine growth and development, response to biotic and abiotic factors in the growth environment, crop productivity and product quality. A detailed analysis of genomic regions containing genes and associated regulatory elements (promoters and enhancers) is essential to understanding the full genetic potential of peanut as a healthful and profitable food crop. Polyploidy, resulting from the joining of two genomes from ancestors with divergent evolutionary history, is a consideration in crafting research strategies for cultivated peanut. Because two or more copies of many genes typically are present, specialized approaches must direct the accurate alignment and assignment of the sequence for each gene isoform to the proper chromosome in assembly of the peanut genome from a random assortment of DNA sequence fragments.

Goal: Build a framework for assembling the peanut genetic genome by identifying and integrating the positions of expressed genes on genetic, transcript and physical maps.

Performance Measures

3.1: Advance discovery and knowledge of abundantly-expressed genes in peanut.
Sequencing of clones from cDNA of an expressed sequence tag (EST) database is the most efficient approach for gaining genomic information in peanut. ESTs can be used for gene discovery, genome annotation, comparative genomic analyses, and markers for genes that will lead to new cultivars with improved traits such as hypo-allergenicity and disease resistance.
Research Need: Adequate EST libraries to generate useful DNA-markers that are associated with peanut allergenicity & agronomic traits. A database containing about 200,000 ESTs (single-pass 5' sequence) will be constructed from flowering and peg/pod/seed tissues, from tissues responding to pathogens causing diseases, specifically the tomato spotted wilt (TSWV), leaf spots, Sclerotinia blight, and *Aspergillus*-aflatoxin, and to drought/heat stress. These will then be used to generate a set of 20,000 unigene clones that will be subjected to 3' sequencing. Selected ESTs will be associated with BACs for gene discovery, and the development of molecular markers and genetic maps of the peanut genomes.
Anticipated Products

- Useful EST libraries from specific peanut organs exposed to various environmental/ experimental conditions during various stages of plant development.
- Identification and characterization of candidate genes for allergens and agronomic traits from peanut EST sequences.
- Bacterial Artificial Chromosome (BAC) libraries with DNA sequences associated with allergen genes.

Potential Benefits: A database of expressed sequence tags (EST), the expressed portion of the peanut genome, will facilitate gene discovery and development of DNA markers. Use of these markers, on microarray chips to select progeny with desired gene combinations from segregating populations, will enhance the efficiency of peanut variety development

Functional Genomics & Proteomics

Proteins are the structural and enzymatic products encoded by genes. Examination of gene expression on a genomic scale at the mRNA level using micro-arrays does not provide information on post-translational modifications that affect the abundance and activity of gene products (proteins). High-resolution protein separation on two-dimensional gels followed by mass spectrometry of excised protein spots allows the rapid profiling of gene expression and protein metabolism. The use of probes created from known gene sequences or deduced from protein structure ('reverse' genetics) is an effective approach to discover and determine the function of the corresponding unknown gene sequence in peanut.

Goal: Improve knowledge of gene identification and regulation by providing baseline data and tools that facilitate the association of DNA-sequences in gene-rich regions of the peanut genome with a biological function.

Performance Measures

4.1 Proteomics research to characterize seed protein composition and function. Proteomics is the extensive characterization of proteins in biological organisms. The extent of genetic variability among peanut seed protein profiles has not been explored with high-resolution methods. Few attempts have been made to characterize changes in protein composition and allergenicity in the seed of peanuts with genetically inserted or deleted genes that mediate expression of allergenic proteins and other agronomic traits.

Research Need: Genetic regulation of allergen synthesis, structure, and composition. The diversity of allergen genes will be determined from ESTs representing all major peanut seed proteins. With clinically relevant serum, IgE-binding proteins will be identified for individual members of gene-families and correlated by sequence with the corresponding cDNAs. Homogenous purification of potential allergens will help characterize and distinguish proteins according to anaphylactic risk. Genes for specific high-risk allergen proteins will be eliminated by 'RNAi' or similarly effective technology. Annotated maps of the proteome will be developed to interpret immunological and metabolic changes in proteins that accumulate in compensation for 'deleted' allergens.

Anticipated Products

- Annotated high-density proteomic maps of developing & mature peanut seed of cultivars and germplasm exposed to various biotic and abiotic stresses.
- Annotated fingerprint patterns for allergenic proteins among experimental germplasm accessions of the USDA Peanut Germplasm Collection.
- A reference proteomic map from peanut leaf tissue.
- Micro-arrays with relevant DNA markers to expedite breeding progress in the development of elite peanut varieties.
- Knowledge of the genetic and metabolic regulation of peanut seed protein profiles.
- Genetic resources exhibiting unique gene insertions or deletions that influence the levels of potential allergenic proteins in peanut.
- An index of the allergenic potential of specific peanut proteins.

4.2 Application of 'reverse genetics' to explore gene function. Since cultivated peanut is a tetraploid, gene function may not simply be additive with respect to the combination of two putative progenitors of *A. hypogaea* into one species**.** Therefore, it would be most informative to create mutants in cultivated peanut. A genome-based approach to exploring the function of genes is called reverse-genetics or TILLING (Targeting Induced Local Lesions in Genomes).

Research Need: Peanut TILLING populations exhibiting useful mutations in allergen and other agronomic genes. Assigning gene function to DNA-sequences is hindered by a lack of

polymorphism (spontaneous mutations) within the peanut genome. Natural mutations in genes may be induced throughout the genome via chemical agents or high-energy radiation. TILLING will accelerate recovery of genes that may abate peanut allergy, and the development and use of precise DNA-markers in 'marker-assisted' breeding programs to select genotypes with specific mutations in members of allergen gene-families.

Anticipated Products

- Unique genetic resources for peanut that include DNA libraries of induced mutations in genes for allergenic proteins and agronomic traits.
- Germplasm and breeding lines with beneficial mutations in genes that govern the expression of allergens and other agronomic traits.

Potential Benefit: Advances in functional genomics and proteomics will provide assurance that targeted allergenic potential is arrested and will facilitate the assessment of possible collateral alteration of other potential allergens. These resources will enable strategies for reducing the allergenic potential of peanut, and to provide novel germplasm for biological characterization of metabolic and regulatory mechanisms governing protein synthesis in peanut seed. The use of gene markers on micro-array chips, to select progeny with desired gene combinations from segregating populations, will enhance the efficiency of peanut variety development.

Bioinformatics

Bioinformatics consists of data management (acquisition, storage, integration and dissemination) and data interpretation (data analysis, visualization and biological modeling). The scope of bioinformation includes DNA sequences, RNA expression levels, protein interactions, map positions and QTL. However, until robust databases are developed for peanut, bioinformation for peanut must be leveraged from other plant and legume species. Ability to find gene products that can be associated with biochemical pathways or networks, enables the discovery of the molecular basis for phenotypic traits from related species. The peanut bioinformatic database should be a component of the Legume Information System (LIS). LIS, a Congressionally mandated joint collaboration between USDA-ARS and the National Center for Genomic Research (NCGR), was conceived to be a comparative legume resource, and a first step towards leveraging bioinformation from model plants to gain insights into crop species. LIS is a data management system that includes an annotation pipeline of data analysis tools and legume (soybean, peanut, alfalfa, pea, dry beans) data bases developed by multiple groups.

Goal: Provide bioinformatic management of peanut biological information resources by establishing a state-of-art interactive system for public distribution of data and information

Performance Measures

5.3 Assure Peanut Genomic Resources are included in the Legume Information System (LIS). A distinct peanut information resource would facilitate the storage and use of bioinformation for peanut. There is great potential for high quality peanut information and enhanced comparative genomics approaches are based on an information resource that is distinct to the genus *Arachis.* Subsequent or concomitant incorporation of peanut data into LIS greatly enhances the utility of the peanut information resource.

Research Need: Development of ArachisDB. Bioinformatics involves management and interpretation of data from DNA sequences, forms of gene expression, protein interactions and the relationships of these data with traits such as allergens. With advances in Peanut Genomics a distinct information resource will facilitate the storage and use of bioinformation for peanut. Incorporation of peanut information into LIS will not only ensures that peanut research is

enhanced with related information from other legumes, it will also ensure that the diversity of allergen genes and potential allergens are identified among all food legumes.

Anticipated Products

- A state-of-art interactive bioinformatics resource for peanut, and other legumes.
- Advanced methods for comparative genomic analyses.

Potential Benefit: since inception, the legume research communities have guided the development of LIS. Future advances also require collaborative relations with other plant bioinformation groups to ensure representation in plant community defined ontologies and controlled vocabularies. Recognizing the power of comparative genomics among species to identify candidate genes, unique genes, and evolutionary relationships among genes for crop improvement, a Peanut Information Resource (ArachisDB) should be developed and should become a component of LIS