Transgenic Crops and Soybean Protein Analysis

Savithiry S. Natarajan, Ph.D.

United States Department of Agriculture
Agricultural Research Service

Soybean Genomics & Improvement Laboratory

Savi.natarajan@ars.usda.gov

Agricultural Research Service (ARS) MISSION

ARS conducts research to develop and transfer solutions to agricultural problems of high national priority and provide information access and dissemination in order to:

- Ensure high-quality safe food and other agricultural products
- Assess the nutritional needs of Americans
- Sustain a competitive agricultural economy
- Enhance the natural resource base and the environment
- Provide economic opportunities for rural citizens, communities, and society as a whole

Source: USDA

Beltsville Agricultural Research Center (BARC)

Source: USDA

Soybean Genomics and Improvement Laboratory (SG&IL)

Mission: To improve the quality of Soybean and to reduce the costs of production

Research Goals:

1) Define the Function and Interaction of Genes and Metabolites Controlling Soybean Resistance to Disease

2) Develop Genetic Markers and Genome Maps to Expedite the Deployment of Genes

3) Determine and define any unintended effects that may be associated with transgenic soybean

Soybeans

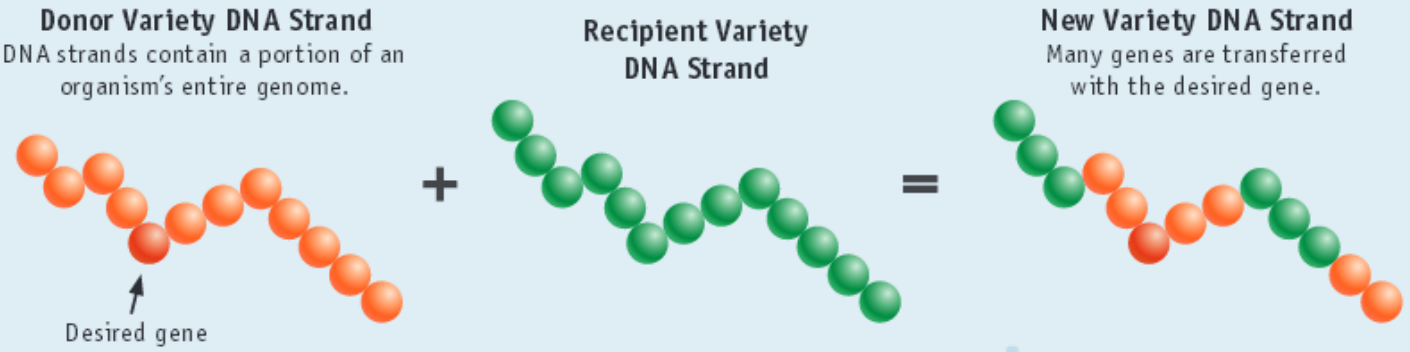
- Second most valuable crop in the US (Annual value ~$41 Billion)
- Inexpensive source of proteins
- More than 90 % of soybean planted in the US is GMO
- Health issues-presence of allergens and anti-nutrients

Source: Council for Biotech. Inf.

Traditional Plant Breeding Vs. Genetic Engineering

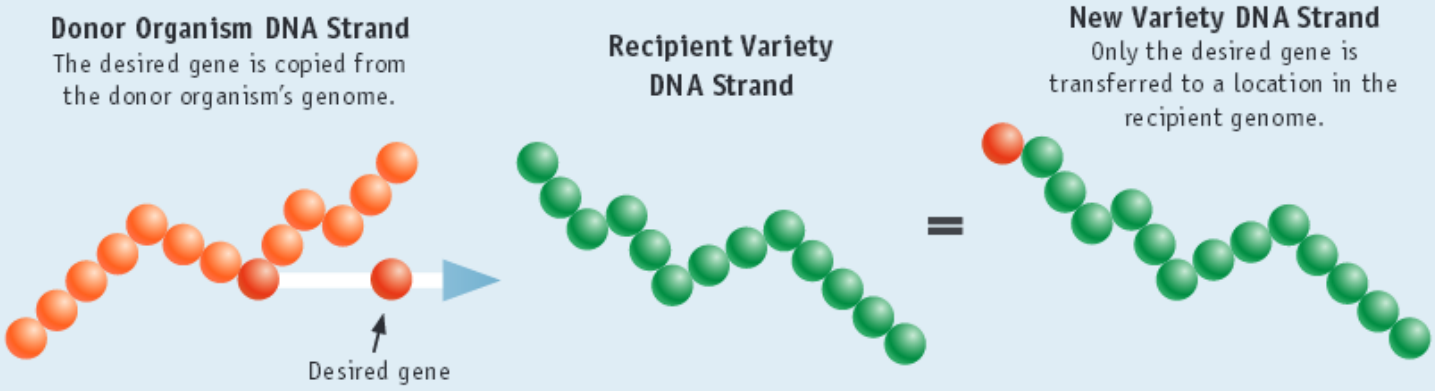
Traditional

The traditional plant breeding process introduces a number of genes into the plant. These genes may include the gene responsible for the desired characteristic, as well as genes responsible for unwanted characteristics.



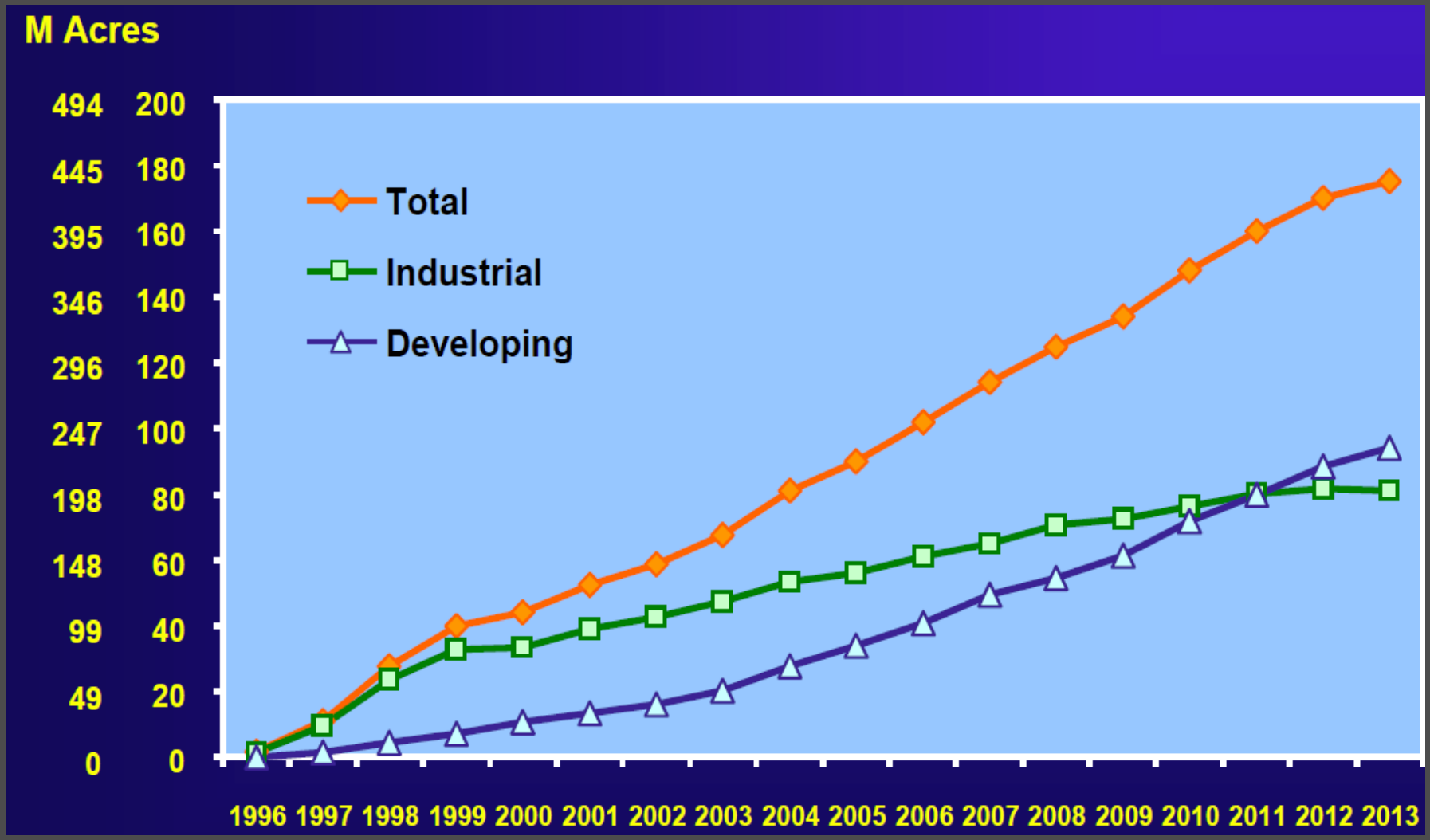
Genetic Engineering

Genetic engineering enables the introduction into the plant of the specific gene or genes responsible for the characteristic(s) of interest. By narrowing the introduction to one or a few identified genes, scientists can introduce the desired characteristic without also introducing genes responsible for unwanted characteristics.



Source: Consumer Information, FDA, 2013, www.fda.gov/consumer

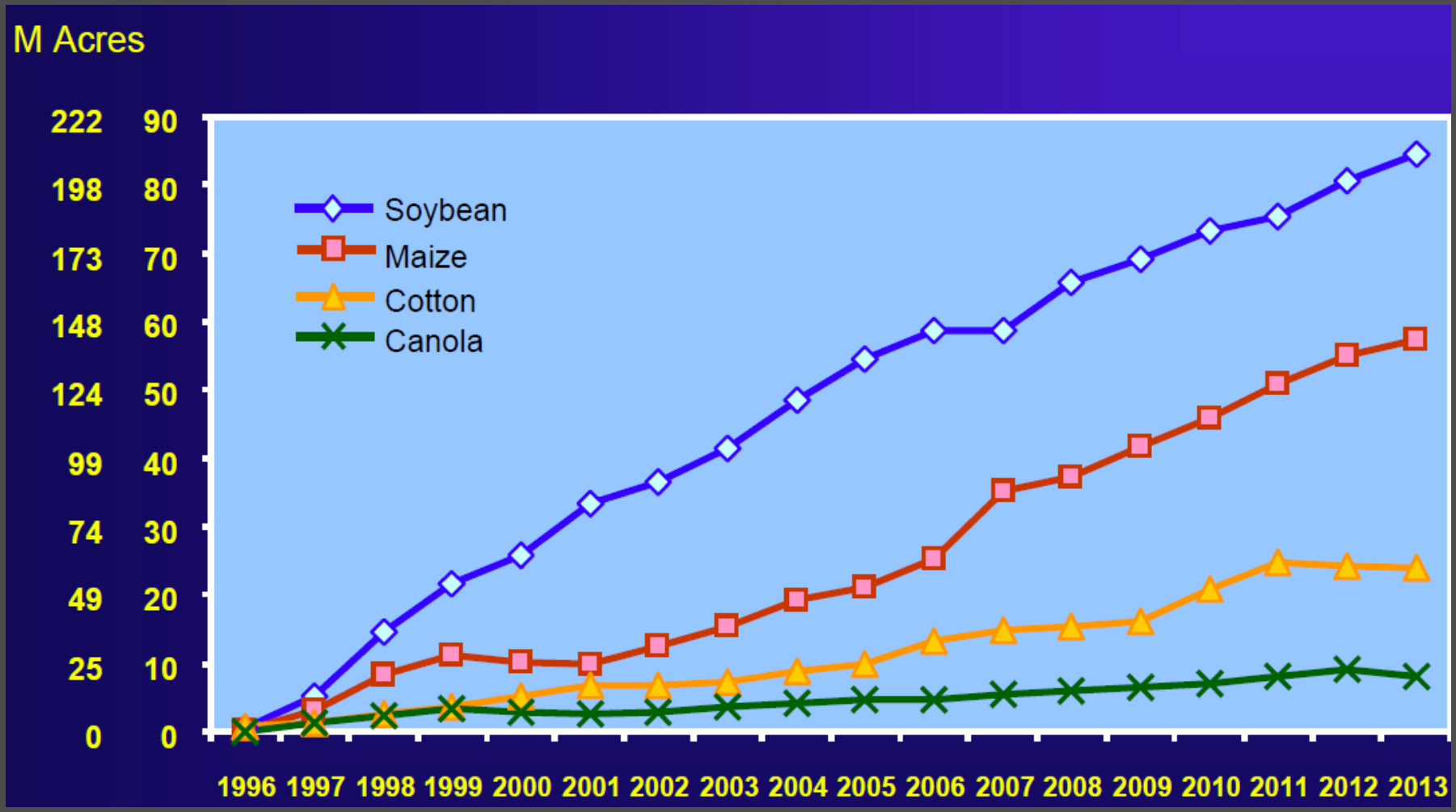
Global Cultivation of GMO Crops



Million hectares, Million acres

Source: ISAAA, Clive James, 2013

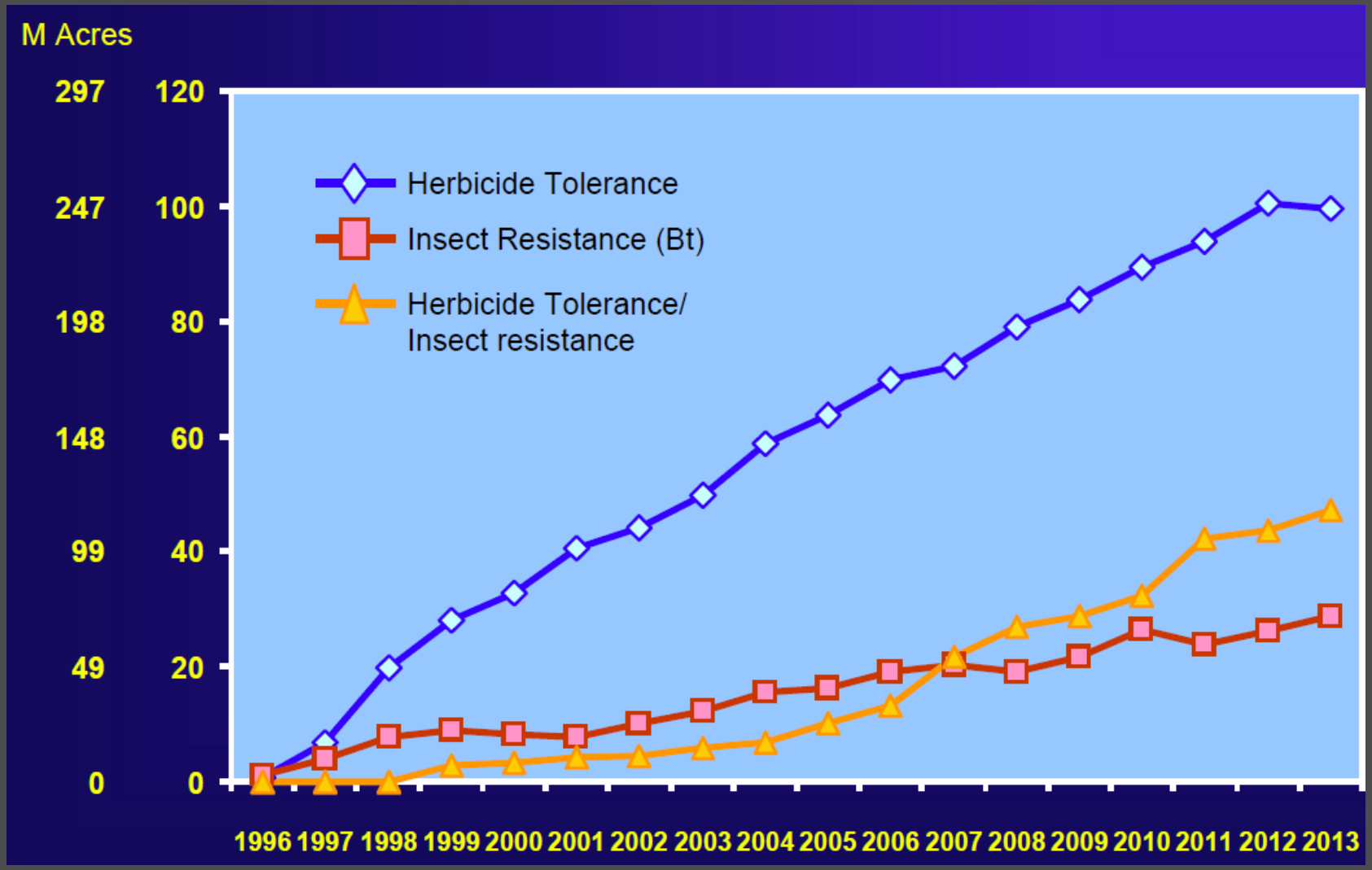
Global Area of GMO Crops (By Crop)



Million hectares, Million acres

Source: ISAAA, Clive James, 2013

Global Area of GMO Crops (By Trait)



Million hectares, Million acres

Source: ISAAA, Clive James, 2013

Proteomics

- To improve the probability of detecting any changes in protein profiles, profiling techniques such as proteomics are being evaluated
- The evaluation of substantial equivalence requires the availability of database based on natural variation of proteins
- Such a database is important for determining if the new biotech product falls within or outside the range of natural variation
- Proteomics allows us to build such a database for protein variation
- Proteomics can be a useful tool for the identification of precursors, post-translational modifications and also degradation products of proteins

3 Kinds of Proteomics

- Structural Proteomics
  - High throughput X-ray Crystallography/Modelling
  - High throughput NMR Spectroscopy/Modelling
- Expressional or Analytical Proteomics
  - Electrophoresis, MudPIT, Protein Chips, DNA Chips, 2D-HPLC
  - Mass Spectrometry, Microsequencing
- Functional or Interaction Proteomics
  - HT Functional Assays, Ligand Chips
  - Yeast 2-hybrid, Deletion Analysis, Motif Analysis

Source: Dr. David Wishart, University of Alberta

Research Challenges

- Analytical challenges for better separation and accurate identification of proteins
- Limitations in detecting very large/small, acidic/ basic proteins
- The sensitivity of current staining procedures sets limits on the amount of protein required for loading, identification and quantification
- The comparisons of datasets between different laboratories requires a standardized method to be developed for sample isolation and electrophoresis
- More background data on natural variation are needed, because the environment greatly influences protein abundance

Cellini et al. 2004. Food & Chem. Toxicol.42: 1089-1125

Research Objectives

- Develop and standardize effective proteomic methodologies and establish a database of soybean seed proteins
- Determine the natural variation of seed proteins of a wide range of soybeans to establish a baseline
- Determine expression of new protein/changes in proteins/quantity of the protein in transgenic soybean by comparing with the natural variation of seed proteins

Proteomic Tools

Protein Analysis - Expression and Identification

Sample Preparation

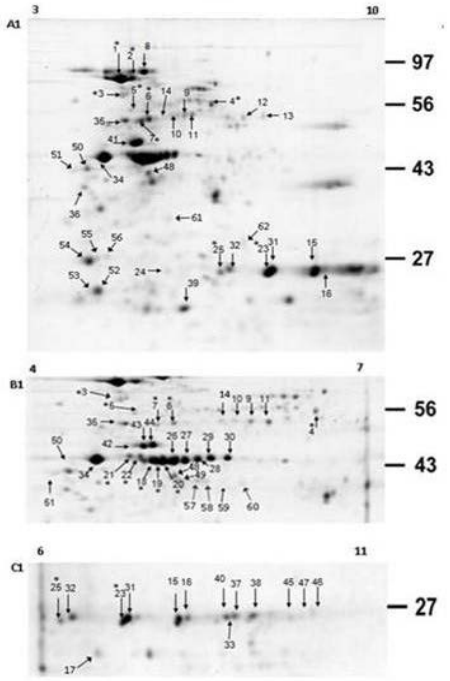
- Optimization and standardization of growth conditions
- Cell fractioning (supernatant, intracellular proteins, and membrane proteins)
- Purification steps (Phenol extraction, TCA/Acetone precipitation)

Isoelectric Focusing (1st Dimension)

- Separates proteins according to their isoelectric points
- Etta IPGphor: use of precast IPG strips with immobilized pH gradients (pH 3-10; pH 4-7; pH 6-11)

SDS-PAGE (2nd Dimension)

- Separates proteins according to their molecular weight (13 x 13 cm gels)
- Staining by Coomassie or Silver stain
- Each spot in the resulting 2nd dimensional array corresponds to a single / multiple protein species in the sample



2D-PAGE of Soybean Protein

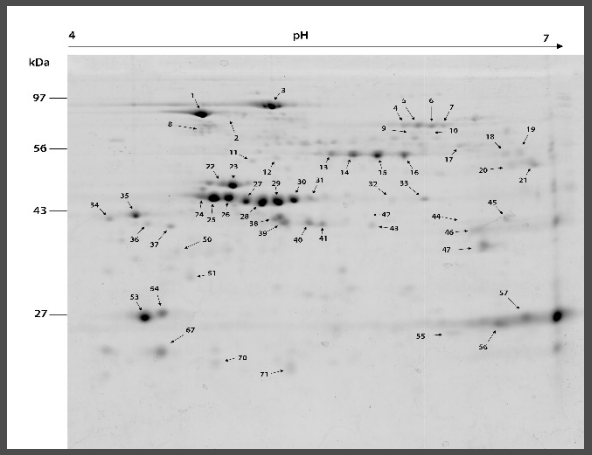
- Software analysis: spot detection, background correction, quantification of spot densities, inter-sample spot alignment
- Identification of protein spots by mass spectrometry (MALDI-TOF-TOF-MS/MS or LC-MS/MS)

MALDI-TOF-MS

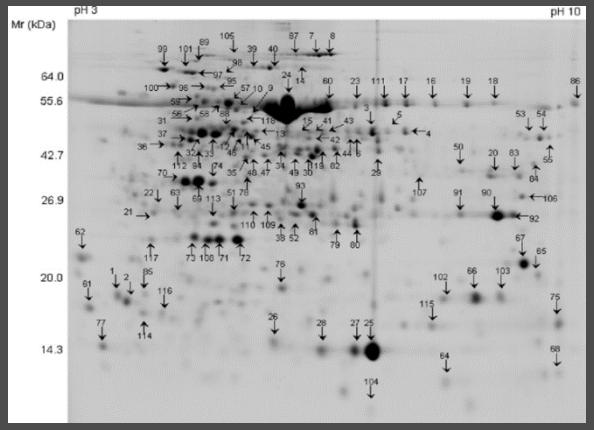
Common Methods of Mass Spectrometry

LC-MS/MS

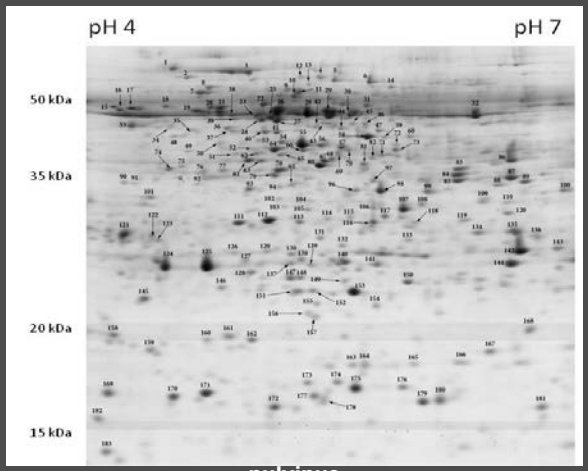
Protein Extraction



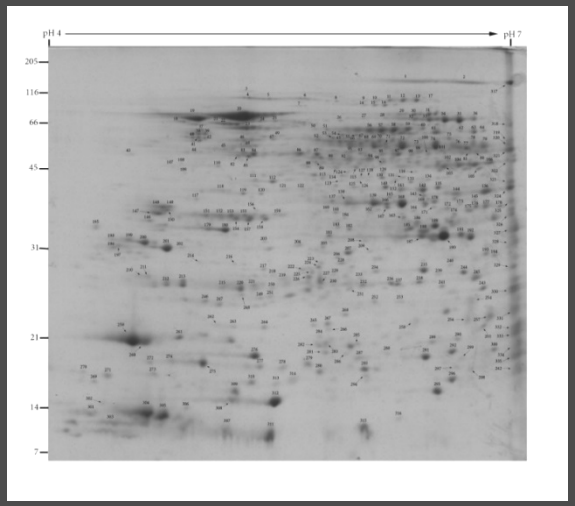
seed



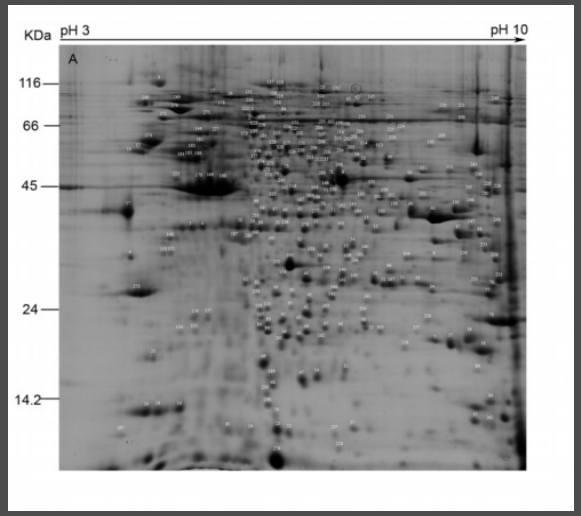
leaf



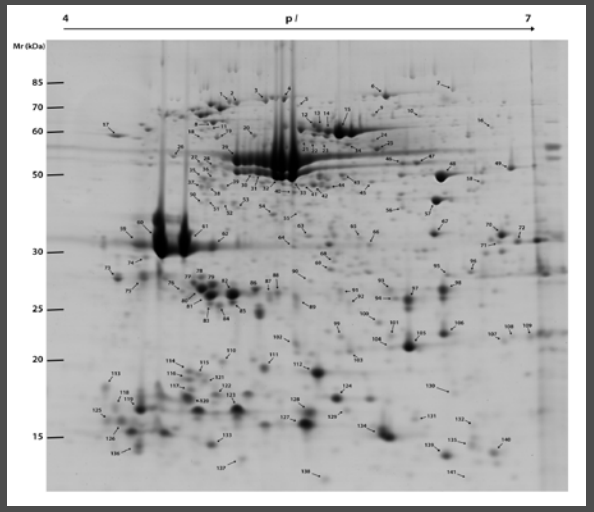
pulvinus



embryonic axis



Soybean nematode



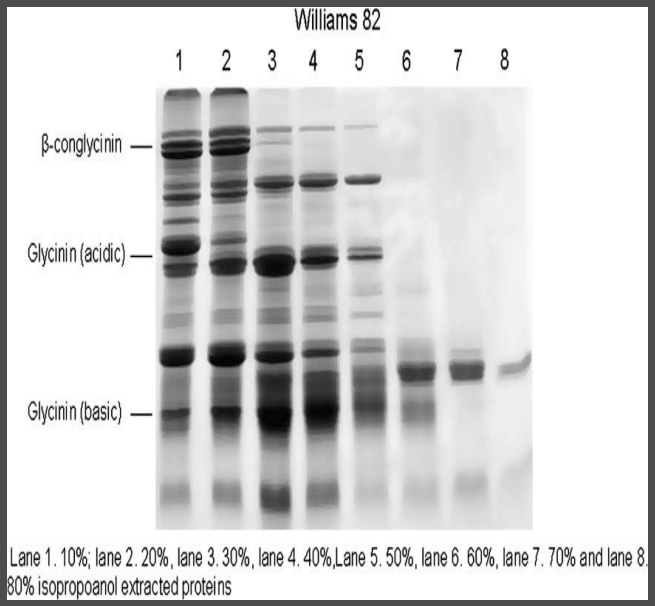
Common bean

- Developed an efficient protein extraction method for soybean proteins
- The modified TCA/acetone extraction method appears to be more efficient and more reliable, as compared to the thiourea, urea and phenol extraction methods, in solubilizing abundant and low abundant soybean seed proteins
- Developed several efficient protein extraction methodologies specific to additional agricultural proteomic investigations including common bean and soybean nematode

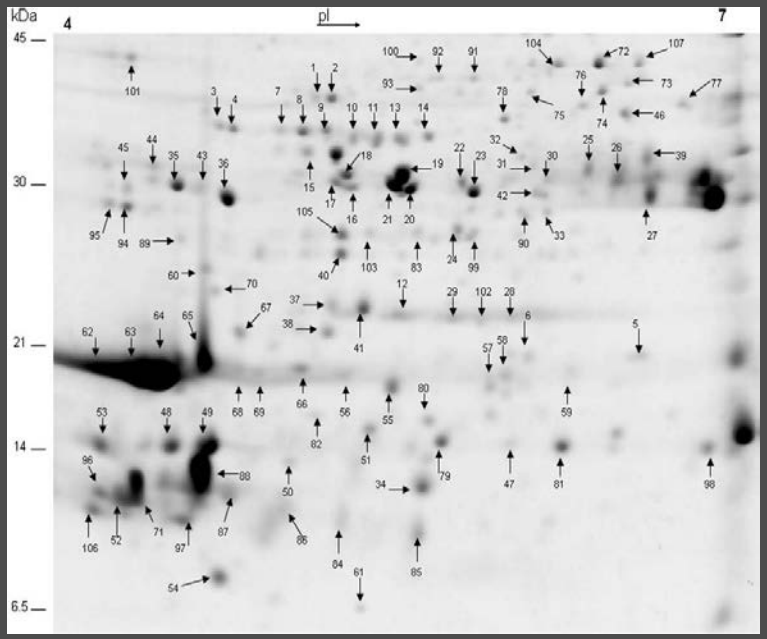
Natarajan et al. , 2013 JBAS 9: 424-437; Chen, MacDonald, Khan, Garrett, Matthews, Natarajan, 2011. Proteomics,11:4742-46; Natarajan et al. 2005. Anal. Biochem. 342: 214-220; Natarajan et al. 2013, JBAS 9: 309-332; Xu et al. 2006. Phytochemistry 67: 2431-2440

Extraction of Low Abundant Proteins-Isopropanol

- Large amounts of the abundant storage proteins, β-conglycinin and glycinin, in soybean *s*eeds hinder the isolation and characterization of low abundant seed proteins
- We investigated whether isopropanol extraction could facilitate resolution of the low abundant proteins
- Analyzed proteins by different concentrations (10, 20, 30, 40, 50, 60, 70 and 80%) of isopropanol by 1D-PAGE
- 40% isopropanal was selected for 2D-PAGE for preferential enrichment of low abundant proteins
- **Analysis of 2D-PAGE showed 107 proteins which were low abundant or absent by the conventional extraction procedure were clearly seen in the 40 % isopropanol extracts**



Lane 1. 10%; lane 2. 20%, lane 3. 30%, lane 4. 40%,Lane 5. 50%, lane 6. 60%, lane 7. 70% and lane 8. 80% isopropoanol extracted proteins



Conclusion:

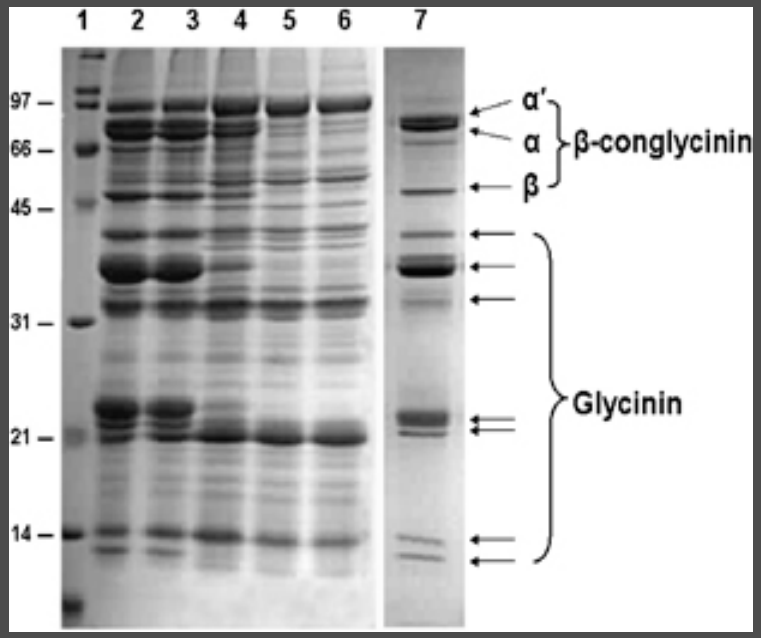
- Our results suggest that extraction of soybean seed powder with 40% isopropanol enriches lower abundance proteins and is a suitable method for 2D-PAGE separation and identification
- This methodology could potentially allow the extraction and characterization of low abundant proteins of other legume seeds containing highly abundant storage proteins

Natarajan et al. 2009. Analytical Biochem. 394: 259-268

Extraction of Low Abundant Proteins- CaCl2

We have developed a fast and simple fractionation technique using 10mM $CaCl_2$ to precipitate soybean seed storage proteins, glycinin and β-conglycinin

This method removes 87% of the highly abundant seed proteins from the extract



Lane 1, protein molecular weight markers in kDa

Lane 2, no calcium added

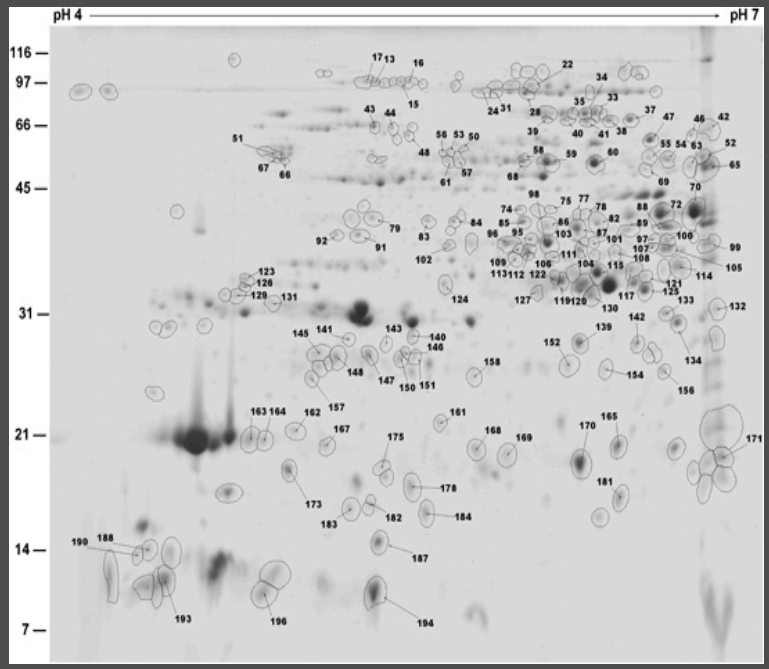
Lane 3, 1mM$CaCl_2$

Lane 4, 2mM $CaCl_2$

Lane 5, 5mM $CaCl_2$

Lane 6, 10mM $CaCl_2$

Lane 7, resulting precipitant from calcium chloride fractionation, demonstrating the removal of almost exclusively b-conglycinin and glycinin



Optimum concentration 10mM $Cacl_2$; Detected 541 protein spots; Identified 197 proteins

Fractionation also provided detection of 63 new phosphorylated protein spots and enhanced the visibility of 15 phosphorylated protein spots, using 2-D electrophoretic separation

Krishnan, Oehrle and Natarajan. 2009. Proteomics. 9: 3174-3188

Study the Natural Variation of a Wide Range of Soybeans Using Proteomics

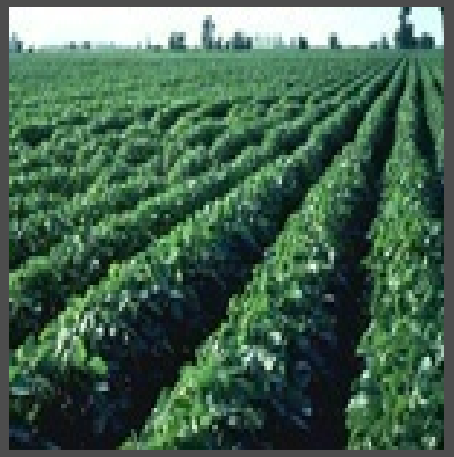
- Man made gene transfer (GMO) has been a major issue in crop production
- Gene transfer by conventional methods is acceptable
- Our initial research focus is to:

1. Examine the range of variability of proteins produced in a wide range of soybean genotypes that could readily be transferred using conventional breeding
2. Compare the range of protein variation in transgenic versus normal soybean using proteomics

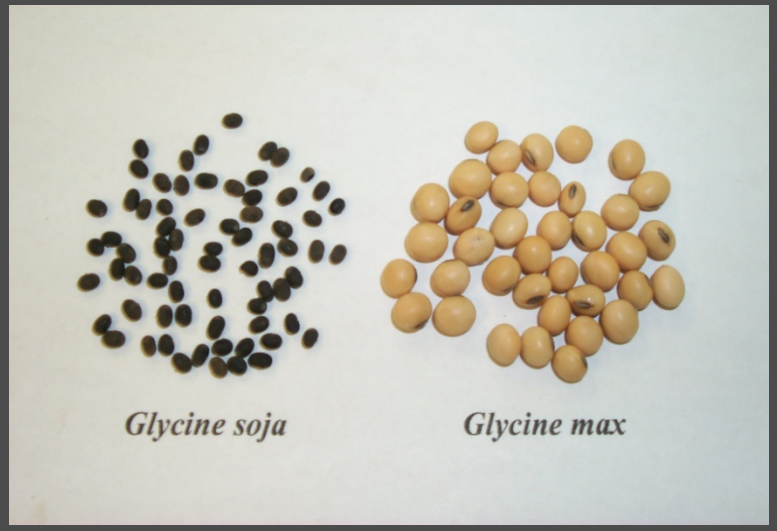
Plant Materials to Study Naturally Occurring Variation in Seed Proteins

- "Ancestral cultivars" (Old or first progeny) of N. Am. cultivated soybean estimated to have contributed approx. 81.5 % of the genes in current cultivars.
- Asian *G. max* (Landrace and Elite) accessions collected from Chinese Provinces, Korean provinces, Japanese prefectures. Collected from 22-50 degrees N, 104-140 degrees E.
- *G. soja* (Wild) Plant Introductions from China, Korea, and Japan. Collected from 23-50.2 degrees N, 106-140 degrees E.
- Selected 16 soybean genotypes from 4 groups-
  wild, Asian landrace, Ancestors and elite

Wild –*G. soja*



Cultivated –*G. max*



Natural Variation of Soybean Seed Proteins

- Storage proteins
- Allergen proteins
- Anti-nutritional proteins

Storage Proteins

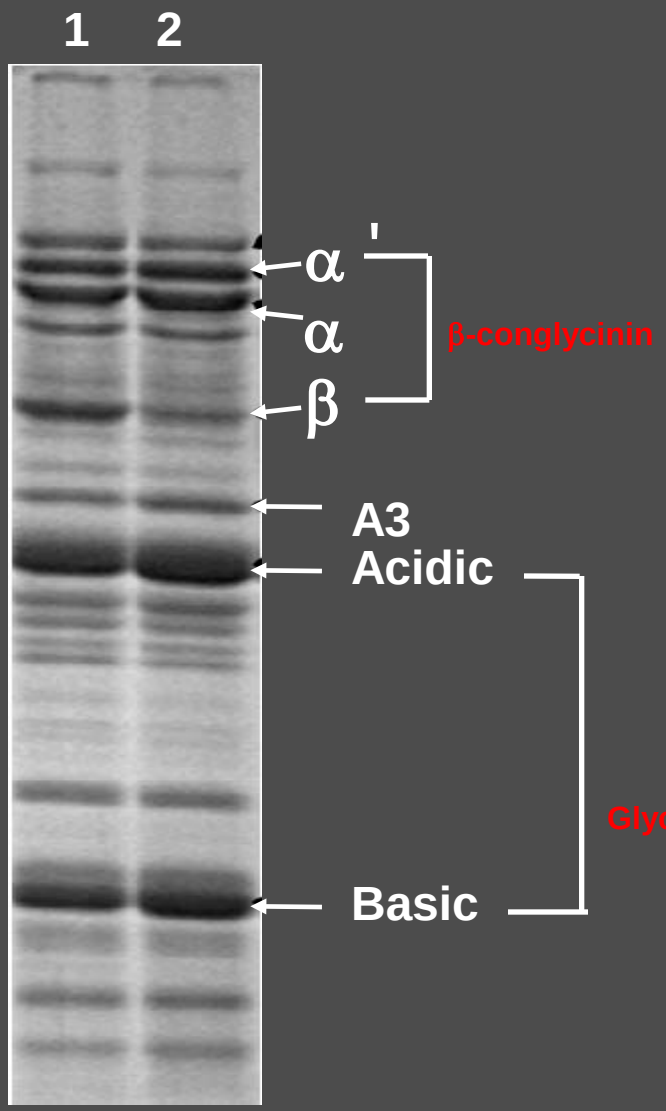
- **β-conglycinin**
  - α subunit of β-conglycinin
  - α′ subunit of β-conglycinin
  - β subunit of β-conglycinin

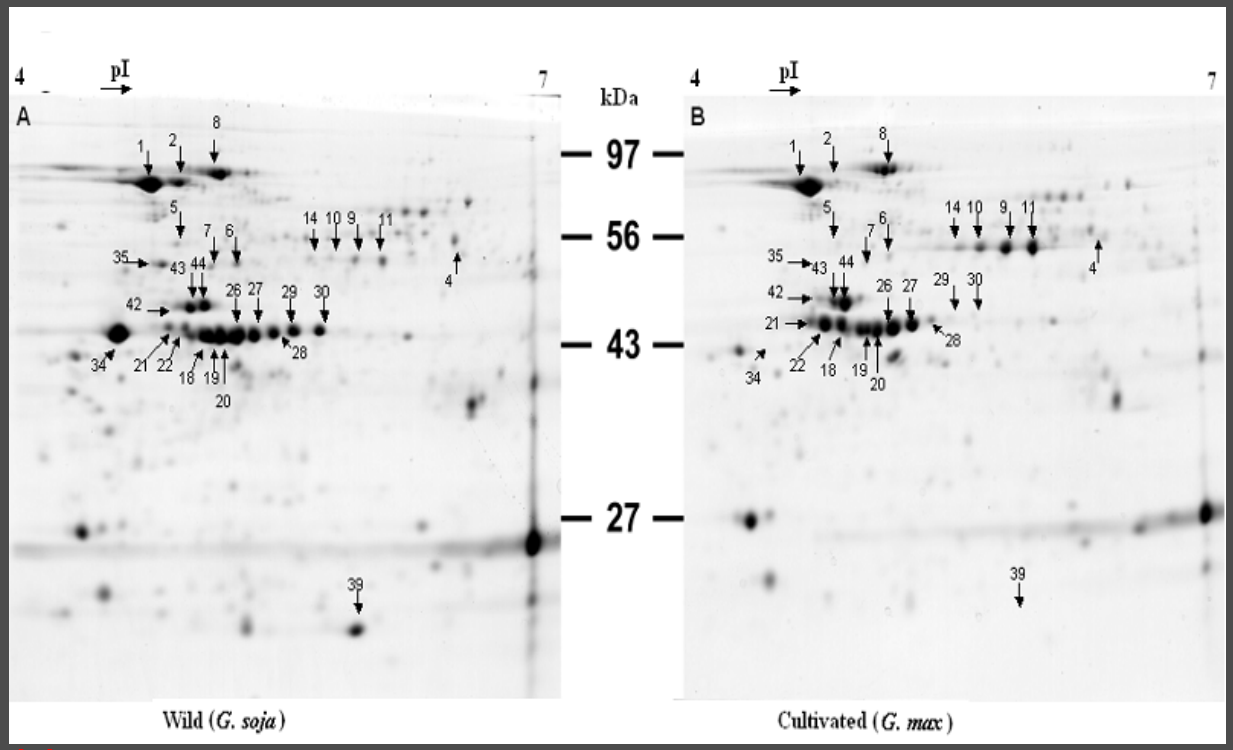
- **Glycinin**
  - Glycinin G1/A1aBx subunit
  - Glycinin G2/A2B1 subunit
  - Glycinin G3/A1ab1B subunit
  - Glycinin G4/A5A4B3 subunit
  - Glycinin G5/A3B4 subunit

- Identified and mapped an additional two genes in soybean variety Resnik, gy6 & Gy7

(*Beillinson et al. 2002. Physiol. Plant. 115: 585-597)*

Variation of Storage Proteins in Wild and Cultivated Soybeans





Natarajan et al. 2006. J. Agric & Food Chem, 54: 3114-3120
Krishnan, Natarajan, Mohmoud & Nelson. 2007. J. Agric. & Food Chem, 55(5):1839-1845
Mohmoud, Natarajan, Bennett, Mawhinney, Wiebold & Krishnan. 2006. J. Agric. Food Chem. 54: 3916-3922

Glycinin

- Important storage protein that lowers cholesterol levels in human serum
- Involved with structural properties of soybean products & Responsible for gel matrix structure
  - Hardness of tofu
  - Unfracturability of tofu
- Gel formed by glycinin is a turbid gel in contrast to transparent gel formed by β-conglycinin

Glycinin Proteins

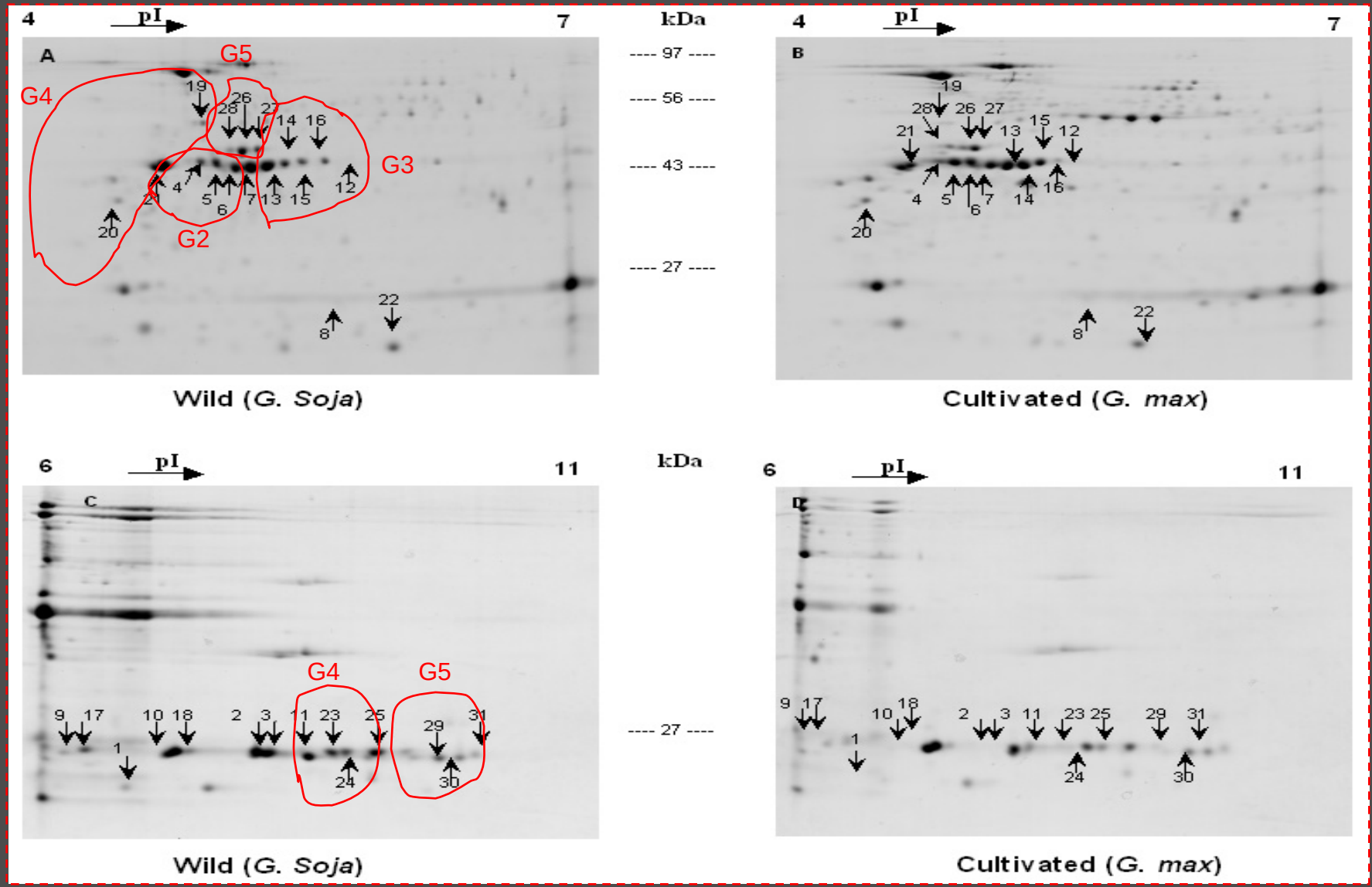
Two major groups

Group 1- Glycinin G1 / A1aBx subunit
Glycinin G2 / A2B1 subunit
Glycinin G3 / A1ab1B subunit

Group 2- Glycinin G4 / A5A4B3 subunit
Glycinin G5 / A3B4 subunit

- 84% homology within the group
- 45-49% homology between the group
- Various subunits play an important role in tofu gel formation

2D-PAGE of Glycinin Proteins



Natarajan et al., 2007. Pl. Physiol. & Biochem. 45:436-444

Allergens & Anti-Nutrients of Soybeans

Allergens

- Gly m Bd 60K (α, α′ and β subunit of β-conglycinin, acidic polypeptides of G1 and glycinin G2)
- Gly m Bd 30K (Soybean vacuolar protein)
- Gly m Bd 28K (closely related to peanut Ara h 1)
- Soybean hydrophobic proteins (Gly m 1.0101, Gly m 1.0102)
- Soybean hull protein (Gly m 2)
- Soybean profilin (Gly m 3, homologous to Bet v 2, a birch pollen allergen with sequence identity of 73%)

Anti-nutrients

- Lectins
- Saponins
- Protease inhibitors
- Phytic Acid
- Lipoxygenase

Natarajan et al. 2006. J. pl. Biochem & Biotech. 15: 103-108
Cordle 2004. J. Nutr. 134: 1213S-1219S
Beardslee et al. 2000. Int. Arch. Allergy. Immunol. 123: 299
Krishnan et al. 2009, 57: 938-943

Variation of Allergen Proteins in Sixteen Soybean Genotypes

Soybean allergen composition in sixteen genotypes

Soybean genotypes	Accession	Protein Spot Number of Allergenic Proteins			
		Gly m Bd 60 K		Gly m Bd 30 K	Gly m Bd 28 K
		α subunit of β conglycinin	Glycinin G2		
Wild (*G. soja*)	PI366120	1, 3, 4, 5, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16	18, 19
	PI393551	1, 2, 3, 4, 5, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16	18, 19
	PI407027	1, 2, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI407282	1, 3, 4, 5, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
Landrace (*G. max*)	PI423954	1, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI89138	1, 2, 3, 4, 5, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI594777	1, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI59845	1, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
Ancestor (old)	PI548445	1, 3, 4	8, 9, 10, 11, 12, 13, 14	16, 17	18, 19
	PI548298	1, 3, 4	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI548318	1, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI548362	1, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
Modern (elite)	PI536635	1, 3, 4	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI533655	1, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI525453	1, 3, 4, 6, 7	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19
	PI513382	1, 3, 4	8, 9, 10, 11, 12,13, 14, 15	16, 17	18, 19

Characterized allergen proteins in sixteen soybean genotypes from four groups

Major proteomic variation was observed between wild and cultivated soybean genotypes rather than among genotypes in the same group

Xu, Caperna, Garrett, Cregan, Bae, Luthria & Natarajan. 2007. J. of the Sci. of Food and Agric. 87: 2511-2518

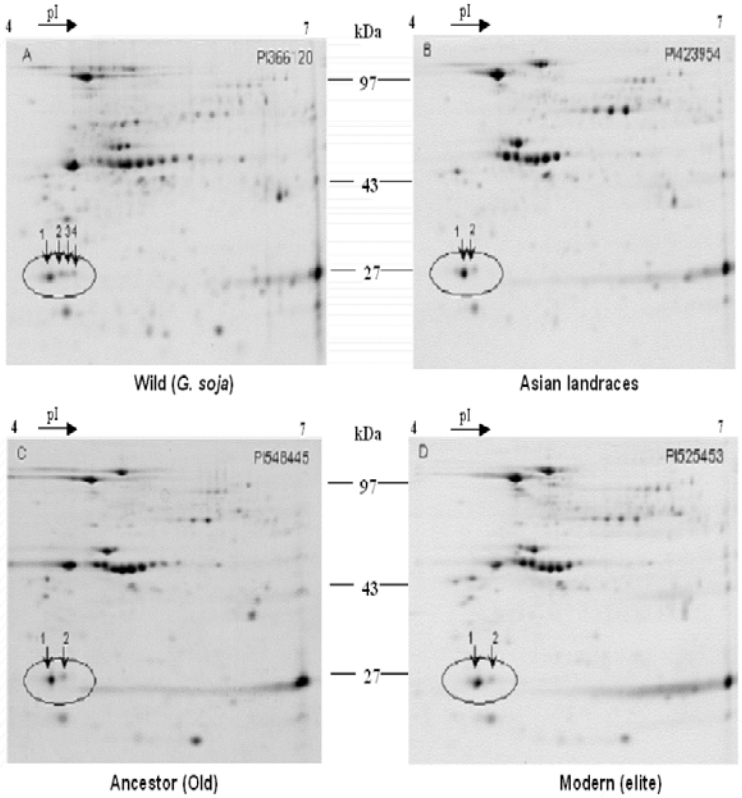
Natarajan et al. 2006. J. Pl. Biochem & Biotech. 15:103-108

Soybean Anti-nutritional Proteins

- Kunitz Trypsin Inhibitor (KTI) is an anti-nutritional protein found in soybean that acts as a proteinase inhibitor
- Involved in respiratory hypersensitivity reactions
- The total seed protein of soybean contains about 6 % proteinase inhibitor
- 32% DNA sequence homology with a rye grass pollen allergen

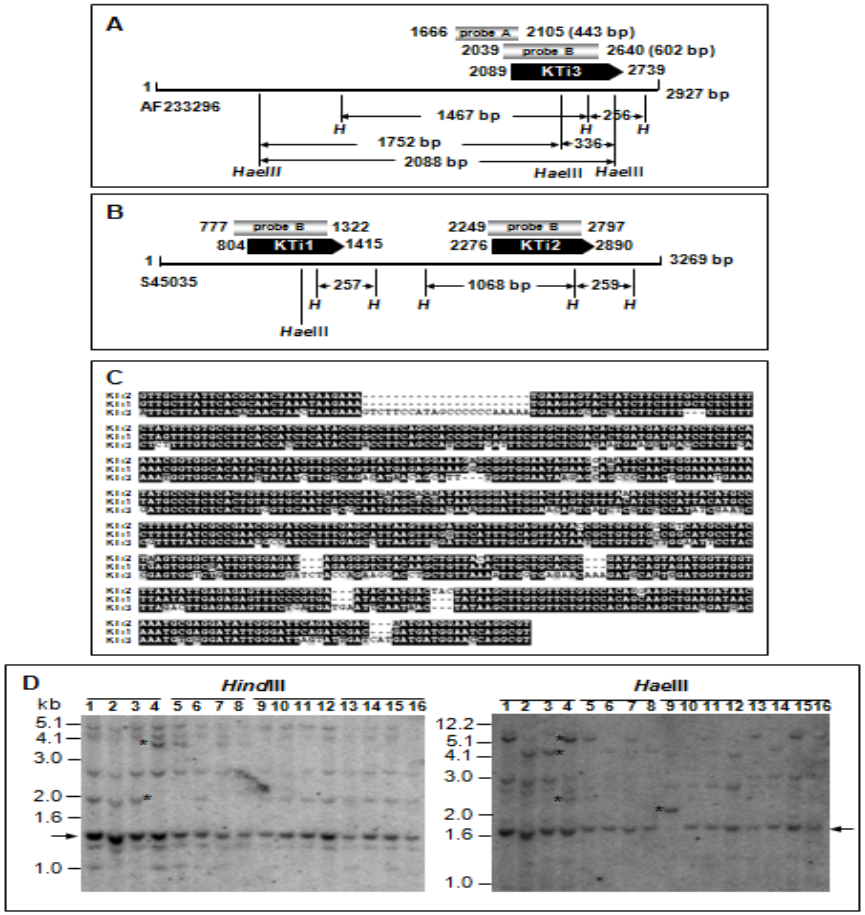
Variation of KTI in Sixteen Soybean Genotypes

Figure1



2D-PAGE

Figure 2



DNA blot analyses of soybean KTI genes

Natarajan et al., 2007. J. Pl. Physiol. 164:756-763

Protein Variation in Soybean Genotypes

- Generation of baseline data
  - Natural protein variation exists in different cultivars of soybeans and maturity groups
  - SoyProDB was established containing comprehensive details of soybean proteins (storage, allergen, and anti-nutritional) http://bioinformatics.towson.edu/Soybean_Seed_Proteins_2D_Gel_DB/Home.aspx
  - Examined the variability of proteins produced in a wide range of soybean genotypes including wild and cultivated soybeans
  - These data will be used to determine if the protein variability in soybeans resulting from transgenic soybean falls outside the range occurring in wild soybean and soybean landraces

Tavakolan, Alkharouf, Khan, and Natarajan 2013, Bioinformation.9(3): 165–167

Summary

- Allergens and anti-nutritional proteins are major concerns
- Advanced technologies are useful for identifying unintended protein expression in new crops
- Different methods have been developed in our lab to determine soybean abundant and low abundant protein profiles
- Multiple isoforms of β-conglycinin and glycinin were identified
- Soy allergen & anti-nutritional (KTI) proteins were characterized

Summary (cont.)

- Natural variation was observed in different classes of soybean seed proteins
- Evaluation of substantial equivalence requires the availability of a database of natural variation of proteins
- A database, SoyProDB, was developed containing comprehensive details of soybean proteins
- Natural protein variation data will be used to determine if the protein abundance in transgenic soybean are within the limits of natural variation