

## SoyProDB: A database for the identification of soybean seed proteins

Mona Tavakolan<sup>1</sup>, Nadim W Alkharouf<sup>1</sup>, Farooq H Khan<sup>2</sup> & Savithiry Natarajan<sup>2\*</sup>

<sup>1</sup>Department of computer and information sciences, Towson University, Towson, MD 21252, USA; <sup>2</sup>USDA-ARS, Soybean Genomics and Improvement Laboratory, Beltsville, MD 20705, USA; Savithiry Natarajan - Email: savi.natarajan@ars.usda.gov; Phone: 301-504-5258; Fax: 301-504-5728; \*Corresponding author

Received January 11, 2013; Accepted January 13, 2013; Published February 06, 2013

### Abstract:

Soybean continues to serve as a rich and inexpensive source of protein for humans and animals. A substantial amount of information has been reported on the genotypic variation and beneficial genetic manipulation of soybeans. For better understanding of the consequences of genetic manipulation, elucidation of soybean protein composition is necessary, because of its direct relationship to phenotype. We have conducted studies to determine the composition of storage, allergen and anti-nutritional proteins in cultivated soybean using a combined proteomics approach. Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) was implemented for the separation of proteins along with matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) and liquid chromatography mass spectrometry (LC-MS/MS) for the identification of proteins. Our analysis resulted in the identification of several proteins, and a web based database named soybean protein database (SoyProDB) was subsequently built to house and allow scientists to search the data. This database will be useful to scientists who wish to genetically alter soybean with higher quality storage proteins, and also helpful for consumers to get a greater understanding about proteins that compose soy products available in the market. The database is freely accessible.

**Availability:** [http://bioinformatics.towson.edu/Soybean\\_Seed\\_Proteins\\_2D\\_Gel\\_DB/Home.aspx](http://bioinformatics.towson.edu/Soybean_Seed_Proteins_2D_Gel_DB/Home.aspx)

**Key words:** Soybean, 2D-PAGE, MALDI-TOF-MS, LC-MS/MS, Conglycinin, Glycinin, Allergen proteins.

### Background:

Soybean is the second most valuable agricultural commodity and an inexpensive source of proteins for humans and animals in the United States. Soybean seed proteins possess unique physiochemical properties which are suitable for various human and animal food uses. Soybean seed proteins are used in baby formula, flour, protein supplements, concentrates, and textured fibers. Soybean seeds contain 40-50% protein on a dry matter basis and consist predominantly of globulin type of proteins. Soybean storage proteins account for ~70-80% of total seed proteins and are deposited in protein bodies, which are specialized membrane-bound organelles. These storage proteins are largely responsible for the nutritional and physicochemical properties of soybeans. Soybeans are also sources of several

secondary metabolites including isoflavones, saponins, phytic acid, flatus-producing oligosaccharides, and goitrogens.

In recent years, the application of proteomic tools such as two-dimensional polyacrylamide gel electrophoresis, matrix-assisted laser desorption/ionization time of flight, and liquid chromatography mass spectrometry, has become popular as a powerful methodology for accurately detecting and examining changes in protein composition. These tools have been extensively used to examine the composition of both natural and transgenic soybean storage protein profiles and determine seed qualities of soybeans. Information about soybean seed proteins will be useful for scientists to understand their functional characteristics for subsequent modification of the

protein through genetic alteration in order to obtain a valuable trait. However limited sources are available to retrieve soybean protein information from publicly available databases [1, 2]. Therefore, we developed a database for soybean seed proteins which are easily accessible to the scientific community as well as the general public.

### Identified Soybean Proteins

The majority of the proteins described in our database are in the form of storage, allergen, and anti-nutritional.  $\beta$ -conglycinin, one of the two major storage proteins, is a trimeric 7S globulin glycoprotein consisting of three types of subunits:  $\alpha$ ,  $\alpha'$  and  $\beta$ , in seven different combinations with the molecular weight of 180 kDa. The second storage protein, glycinin, is a hexameric 11S globulin (360 kDa), consists of acidic (A) and basic (B) polypeptides, and is encoded by five subunits. Based on physical properties, these five subunits are classified into two distinct major groups; group I consisting of G1, G2, and G3 proteins, and group II consisting G4 and G5 subunits. The group I subunits contain more methionine residues than group II. This is an important feature for plant breeders desiring to increase the methionine content in soybean seeds to improve their nutritional quality. Beillinson *et al.* [3] identified and mapped two additional genes in soybean variety Resnik. The soybean allergen proteins exist as three major types. The first major allergen, soybean Gly m Bd 60K, is described as seed storage protein. Based on the many reports, some subunits of  $\beta$ -conglycinin and glycinin are considered as Gly m Bd 60k allergen protein members. Krishnan *et al.* [4] reported that all subunits of  $\beta$ -conglycinin are possible allergens. Gly m Bd 30K is the second major allergen protein of soybean.

This protein was previously described in soybean seed as P34, a 34-kD vacuolar protein [5]. Elimination of P34 from soybean seeds may enhance food safety and make the use of soybean products available to sensitive individuals. Recently, Herman *et al.* [6] developed a transgenic soybean lacking p34 allergen. The third allergen, Gly m Bd 28K is a less abundant protein of soybean which was originally isolated from soybean meal as a 28 kDa glycosylated protein [7]. Like allergens, anti-nutritional proteins in soybean act as limiters for soybean applications as a food or feed. Kunitz trypsin inhibitor (KTI) is one of the abundant anti-nutritional proteins, which can inhibit trypsin, an important animal digestive enzyme. In addition, KTIs have been characterized as food allergens in humans and have 32% sequence homology with a rye grass pollen allergen [8]. Seed lectins are also anti-nutritional proteins present in soybeans and account for about 10% of total protein in some legumes.

### Methodology:

Protein was extracted from cultivated soybean seed, *G. max* PI 423954. Extraction procedures used were a modified TCA/acetone method and isopropanol method [9, 10]. 2D-PAGE separation was performed using IPG strips with pH 3.0-10.0, 4.0-7.0 and 6.0-11 ranges. Protein spots were initially analyzed by MALDI-TOF-MS, and those not positively identified were subjected to LC-MS/MS. Protein identification was performed by searching against National Center for Biotechnology Information (NCBI) non-redundant database using the Mascot search engine (<http://www.matrixscience.com>). The identified proteins were classified in to 3 major protein groups: storage, allergen, and anti-nutritional.

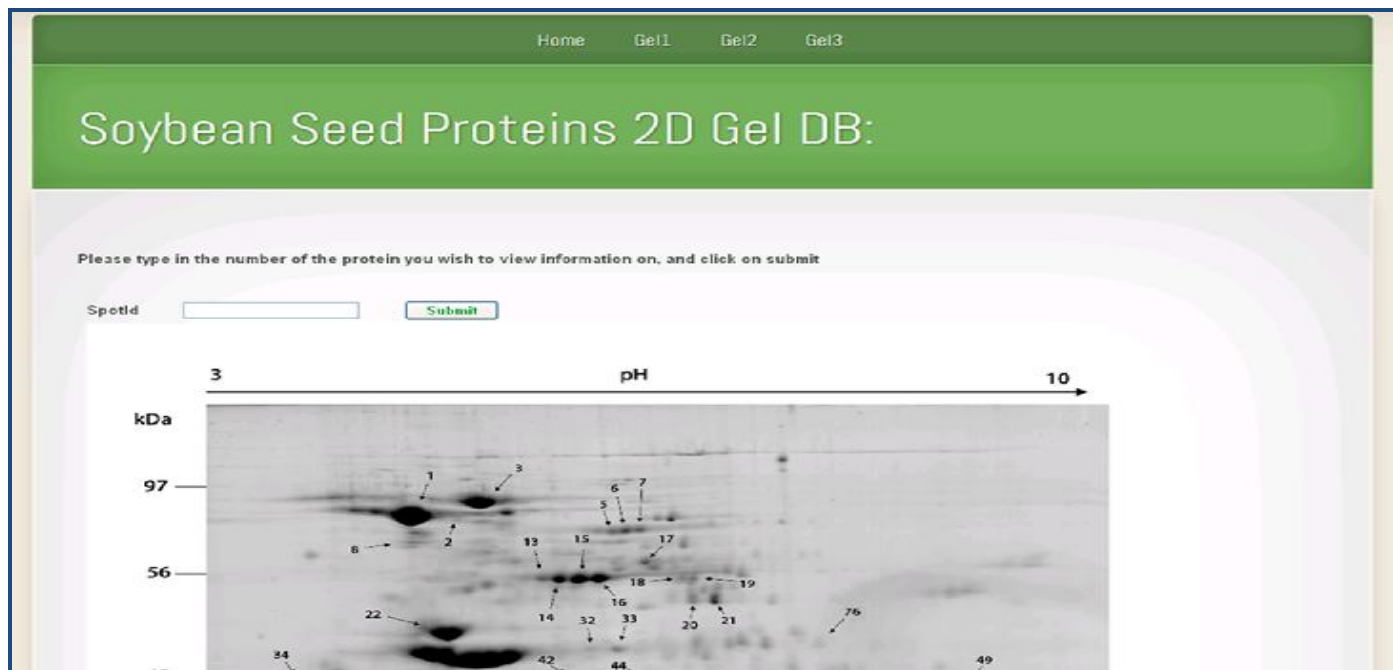


Figure 1: A snapshot of SoyProDB interface.

### Database methodology

The web based database is composed of two main parts. The first one is a relational database built on Access 2007. The second part is the web interface. Web pages were created using Active Server Pages (ASP.Net) using C# programming language. Both the database and the interface are housed on the ISSN 0973-2063 (online) 0973-8894 (print) Bioinformatics 9(3):165-167 (2013)

bioinformatics server at Towson University, MD, USA. The site houses several 2D Gel images showing isolated soybean seed proteins. Each protein is given a unique number (SpotID). Information on each protein on the gels is stored in the database. The website allows users to enter the SpotID for the protein of interest, and retrieves the corresponding information

on that protein from the SoyProDB. The design is simple, yet efficient, and meets the needs of the biologists using the database. **Figure 1** is a snapshot of the database's main page. The database is accessible from: [http://bioinformatics.towson.edu/Soybean\\_Seed\\_Proteins\\_2D\\_Gel\\_DB/Home.aspx](http://bioinformatics.towson.edu/Soybean_Seed_Proteins_2D_Gel_DB/Home.aspx).

#### Utility of the biological community:

The database is of interest to biologists working with soybeans and/or seed proteins. It provides an easy and visual means to identify key proteins in soybean seeds. The web interface allows scientists to access the data using any web browser.

#### Caveats:

More gels need to be added to this database as well as a more comprehensive biological interpretation of the data. Work on that is underway.

#### Acknowledgement:

The authors thank Dr. Benjamin F. Matthews for the critical review of this manuscript. Funding for this research was provided by ARS project 1275-21000-224-00D. Mention of trade name, proprietary product or vendor does not constitute a guarantee or warranty of the product by the U.S. Department of

Agriculture or imply its approval to the exclusion of other products or vendors that also may be suitable.

#### Reference:

- [1] Hajduch M *et al.* *Plant Physiol.* 2005 **137**: 1397 [PMID: 15824287]
- [2] Ohyanagi H *et al.* *Front Plant Sci.* 2012 **3**: 110 [PMID: 22661982]
- [3] Beilinson V *et al.* *Physiol Plant.* 2002 **115**: 585 [PMID: 12121465]
- [4] Krishnan HB *et al.* *J Agric Food Chem.* 2009 **57**: 938 [PMID: 19138084]
- [5] Ogawa T *et al.* *Biosci Biotechnol Biochem.* 1993 **57**: 1030 [PMID: 7763872]
- [6] Herman EM *et al.* *Plant Physiol.* 2003 **132**: 36 [PMID: 12746509]
- [7] Tsuji H *et al.* *Biosci Biotechnol Biochem.* 1997 **61**: 942 [PMID: 9214751]
- [8] van Ree R *et al.* *J Allergy Clin Immunol.* 1995 **95**: 970 [PMID: 7751518]
- [9] Natarajan SS *et al.* *Anal Biochem.* 2005 **342**: 214 [PMID: 15953580]
- [10] Natarajan SS *et al.* *Anal Biochem.* 2009 **394**: 259 [PMID: 19651100]

Edited by P Kanguane

Citation: Tavakolan *et al.* *Bioinformation* 9(3): 165-167 (2012)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited