SCNProDB: A database for the identification of soybean cyst nematode proteins

Savithiry Natarajan[1]*, Mona Tavakolan[2], Nadim W Alkharouf[2] & Benjamin F Matthews[1]

[1]USDA-ARS, Soybean Genomics and Improvement Laboratory, Beltsville, MD 20705, USA; [2]Department of Computer and Information Sciences, Towson University, Towson, MD 21252, USA; Savithiry Natarajan – Email: savi.natarajan@ars.usda.gov; Phone: 301-504-5258; Fax: 301-504-5728; *Corresponding author

Abstract:
Soybean cyst nematode (*Heterodera glycines*, SCN) is the most destructive pathogen of soybean around the world. Crop rotation and resistant cultivars are used to mitigate the damage of SCN, but these approaches are not completely successful because of the varied SCN populations. Thus, the limitations of these practices with soybean dictate investigation of other avenues of protection of soybean against SCN, perhaps through genetically engineering of broad resistance to SCN. For better understanding of the consequences of genetic manipulation, elucidation of SCN protein composition at the subunit level is necessary. We have conducted studies to determine the composition of SCN proteins using a proteomics approach in our laboratory using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) to separate SCN proteins and to characterize the proteins further using mass spectrometry. Our analysis resulted in the identification of several hundred proteins. In this investigation, we developed a web based database (SCNProDB) containing protein information obtained from our previous published studies. This database will be useful to scientists who wish to develop SCN resistant soybean varieties through genetic manipulation and breeding efforts. The database is freely accessible from: http://bioinformatics.towson.edu/Soybean_SCN_proteins_2D_Gel_DB/Gel1.aspx

Key words: Soybean, SCN, nematode, 2D-PAGE, MALDI-TOF-MS, LC-MS/MS, proteins

Background:
The soybean cyst nematode (SCN) is the major pest of soybean in the U.S. and caused ~457 to 819 million dollars in yield losses annually between 2003 and 2005 **[1]**. Some soybean cultivars are resistant to one race or isolate of SCN yet susceptible to another. The life cycle for SCN has different stages including egg, four juvenile (from J1 to J4), and adult **[2]**. The second-stage juvenile (J2) is important, because it is the only stage in which SCN is motile. After hatching from the egg, the J2 moves through the soil and infects by penetrating a host plant root, migrating toward the vascular tissue, and selecting a cell for feeding. The life cycle and pathogenic characteristics of SCN make it an excellent model for studying parasitism and virulence in plant-parasitic nematodes. Literature about gene expression during nematode infection in both susceptible and resistant soybean genotypes using microarrays are available **[3, 4, 5, 6]**. However,

not all expressed mRNAs are immediately translated into proteins. Furthermore, gene expression studies do not reveal protein turnover patterns and protein modifications that may be involved in signaling and communication, protein transport and targeting and other important phenomenon. Therefore, we need protein information which is important for developing SCN resistant varieties. Recent advancements in protein separation methods have led to greater use of proteomics to explore and understand mechanisms of resistance and susceptibility of plants to pathogens. We have published the characterization of SCN proteins previously **[7]**. Since limited information on SCN proteins is available, in this investigation, we developed a publicly available database for SCN proteins for use by the scientific community to identify genes encoding the proteins for resistance and so the genes can be cloned and transformed into susceptible soybean to develop SCN resistant varieties.
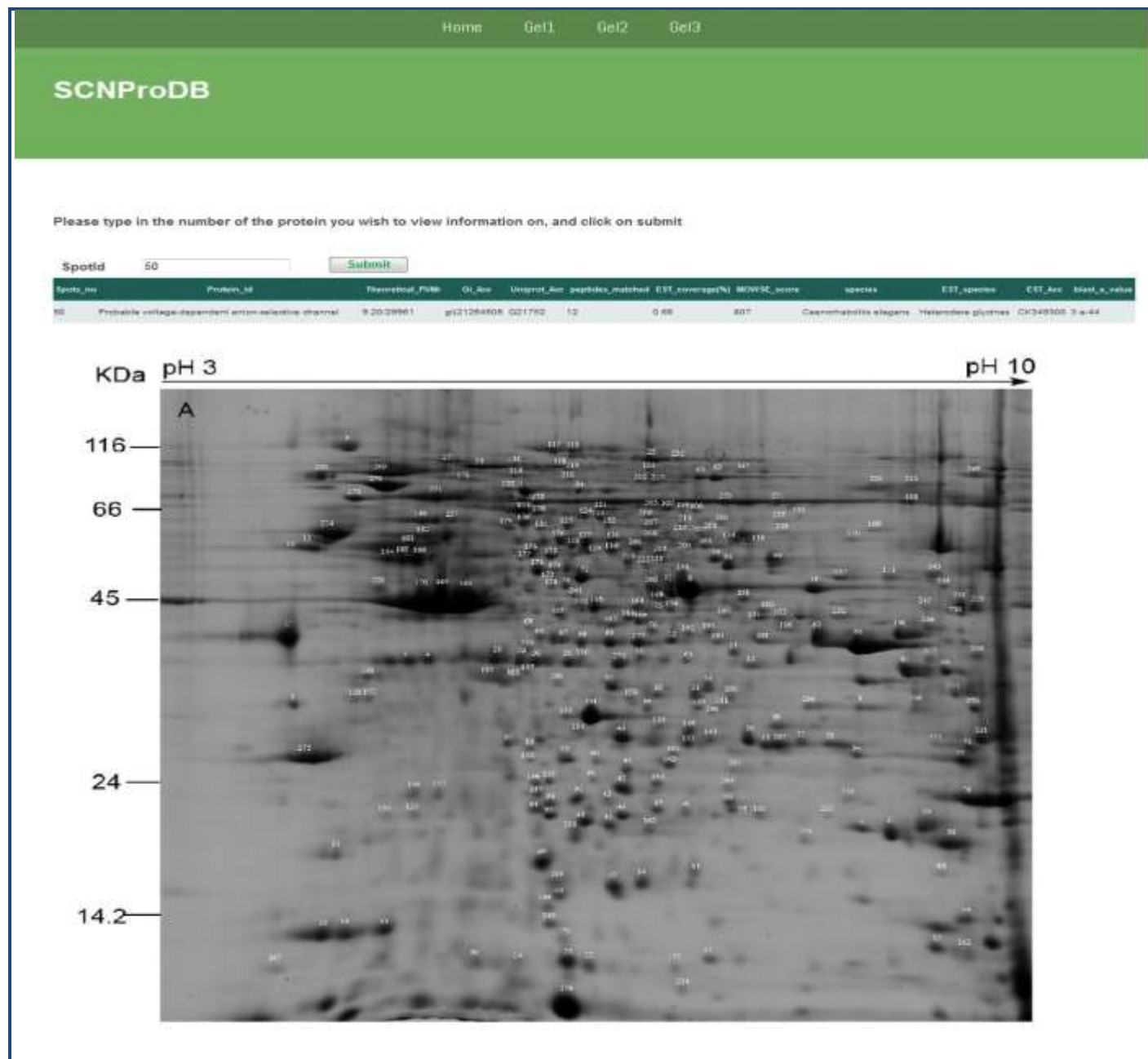
Figure 1: a snapshot of SCNProDB

Methodology:

The nematodes were grown at the United States Department of Agriculture, Beltsville MD, USA according to Klink *et al*. [4]. Proteins were extracted from J2 nematodes using a modified phenol extraction procedure [8]. The nematode protein samples (400µg) were separated in the first dimension using 13cm (Immobiline IPG Dry strips (GE Healthcare). The first dimensional separation was achieved through isoelectric focusing (IEF) using commercially available strips (pH 3-10, 4-7 and 6-11). The strips were loaded onto 12.5% SDS-PAGE gels and subjected to electrophoresis using the Hoefer SE 600 Ruby unit for the second dimensional protein separation. The 2D-PAGE gels were stained with Colloidal Coomassie Blue G-250 and scanned using Image Scanner TM II (GE Healthcare). From the 2D-PAGE, 803 distinct spots were manually excised from

gels and digested with trypsin. A LTQ Orbitrap XL hybrid linear ion trap, Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) was used for protein analysis. The resulting peptides were separated by reverse phase chromatography and searched against NCBI non-redundant databases and the invertebrates EST database using MASCOT search engine (http://www.matrixscience.com), which uses a probability-based scoring system [7].

Construction of database:

The web based database is composed of two main parts. The first one is a relational database built on Access 2007. The database contains several tables that are joined to each other based on a single primary key. This primary key is a unique number for each protein referred to as "SpotID". The SpotID

matches the number written next to the spots in the images. The second part is the web interface. Web pages were created by Active Server Pages (ASP.Net) using C# programming language. Both the database and the interface are housed on the Bioinformatics server at Towson University, MD, USA. The website consists of several sections each of which contain information about a 2D Gel image that represents isolated SCN proteins. Each page displays the 2D image and allows the user to obtain information about the proteins shown in the image. By entering the Spot ID of a protein, its corresponding information from the SCNProDB is retrieved. The website validates the user input to ensure that the entered number exists in the database. Valid inputs can only be integer numbers from 1 to 754. If the user inputs a character or an invalid number, an error message is displayed. For a valid number, a query is executed and returns the following protein information: assigned protein spot number, protein description, theoretical isoelectric point (pI), molecular weight (M_r), NCBI accession number (gi), Uniprot_accession number, peptides_matched, EST_coverage, molecular weight search (MOWSE) score, species, EST_species, EST_Acc, and blast_e_value.

The design is efficient and meets the needs of the biologists using the database. Figure 1 is a snapshot of 2D gel's main page. The database is accessible from: http://bioinformatics.towson.edu/Soybean_SCN_proteins_2D_Gel_DB/Gel1.aspx

Utility to 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.

References:
[1] Wrather JA & Koenning SR, *J Nematol.* 2006 **38**: 173 [PMID: 19259444]
[2] Niblack TL *et al. Annu Rev Phytopathol.* 2006 **44**: 283 [PMID: 16704359]
[3] Alkharouf NW *et al. Planta.* 2006 **224**: 838 [PMID: 16575592]
[4] Klink VP *et al Planta.* 2007 **226**: 1423 [PMID: 17653570]
[5] Klink VP *et al. Planta.* 2007 **226**: 1389 [PMID: 17668236]
[6] Klink VP *et al. Plant Molecular Biology*. 2011 **75**: 141 [PMID: 21153862]
[7] Chen X *et al. Proteomics.* 2011 **11**: 4742 [PMID: 22038885]
[8] Hurkman WJ & *Tanaka CK, Plant Physiol.* 1986 **81**: 802 [PMID: 16664906]