TOWSON UNIVERSITY
OFFICE OF GRADUATE STUDIES

GENOMIC AND PROTEOMIC ANALYSIS IN SOYBEAN, AND BUILDING OF RELATED DATABASES

by

Mona Tavakolan

A Dissertation

Presented to the faculty of

Towson University

in partial fulfillment

of the requirements for the degree

Doctor of Science in Information Technology

Department of Computer and Information Science

Towson University

Towson, Maryland 21252

July 2014

TOWSON UNIVERSITY
OFFICE OF GRADUATE STUDIES

DISSERTATION APPROVAL PAGE

This is to certify that the dissertation prepared by MONA TAVAKOLAN, entitled GENOMIC AND PROTEOMIC ANALYSIS IN SOYBEAN, AND BUILDING OF RELATED DATABASES, has been approved by the thesis committee as satisfactorily completing the dissertation requirements for the degree of Doctor of Science in Information Technology, Department of Computer and Information Science.

Chairperson, Dissertation Committee	**Date** July 21, 2014
Committee Member	**Date** 7/21/2014
Committee Member	**Date** 7/21/2014
Committee Member	**Date** 7/21/2014
Dean, College of Graduate Studies and Research	**Date** July 25, 2014

Acknowledgment

I would like to express the deepest appreciation to my advisor, Dr. Nadim Alkharouf, for his valuable guidance, patient, and encouragement. I am extremely grateful for his assistance and suggestion throughout my project.

Special thanks to my thesis committee members Dr. Darush Davani, Dr. Sungchul Hong and Dr. Yuanqiong Wang for investing time to review my dissertation and giving precious advice. I would also like to express my gratitude to Dr. Ben Mathews for giving me the chance to be part of an excellent research team from which I gained valuable knowledge and experience.

On the personal side, I want to thank my parents, my sister, my brother and my husband for all their great support. I have no word to express my sincere appreciation and love for all they did to make this research possible. None of this would have been possible without their practical and emotional support.

Abstract

GENOMIC AND PROTEOMIC ANALYSIS IN SOYBEAN, AND BUILDING OF RELATED DATABASES

MONA TAVAKOLAN

Soybean (*Glycine max*) is the second most valuable and inexpensive source of proteins for humans and animals in the United States. In this dissertation we analyzed proteomic and genomic data related to soybean. On the proteomics side, we used 2d gel data to identify proteins present in soybean seeds, which is of importance since the seeds are used often in agriculature as feed for animals, or incorporated into food products for human consumption. We also used 2D gel data to identify soybean proteins in the soybean cyst nematode (SCN), which is a major pest of soybean in the United States and causes million dollars losses annually. For both these analysis, we developed publicly available databases that are accessible through the web for researchers to access the data freely.

From the genomics side, we used RNA-Seq data to analyze soybean treated with different plant hormones. The goal here was to see how genes in soybean react when treated with one of these hormones: Salicylic acid (SA), Jasmonic acid (JA), auxin (IAA), and ethylene (Eth). Many of these hormones are used in the agirculature industry, so knowing their effect at the molecular level is of importance to human health. We also used pathway visualization tools to map the gene expression measurements we got from the RNA-Seq experiemnts onto biological pathways, to better understand their effects.

Table of Contents

List of Figures

List of Tables

Chapter 1: Specific Aims and Objectives

Soybean (*Glycine max*) is the second most valuable agricultural commodity and inexpensive source of proteins for humans and animals in the United States. Soybean is a species of legume in the U.S. and their seeds have numerous uses. The two primary products are oil and meat. Fat-free soybean meal is a cheap source of protein for animal feeds; the oil is used in many industrial applications. Recently, scientists focus to identify different proteins in soybean.

The transcriptome is the complete set of transcripts (mRNA levels for each gene) in a cell, and their quantity, for a specific developmental stage or physiological condition. Understanding the transcriptome is essential for interpreting the functional elements of the genome and revealing the molecular constituents of cells and tissues, and also for understanding development and disease. Various technologies have been developed to deduce and quantify the transcriptome. Recently, the development of novel high-throughput DNA sequencing methods has provided a new method for both mapping and quantifying transcriptomes. This method, termed RNA-Seq (RNA sequencing), has clear advantages over existing approaches and is expected to revolutionize the manner in which transcriptomes are analyzed. RNA-Seq is a developed approach to transcriptome profiling that uses deep-sequencing technologies. RNA-Seq also provides more precise measurement of levels of transcripts and their isoforms than other methods.

The goal of this study is to use computational methods to identify and characterize genes of soybean using RNA-Seq method and proteins in soybean using 2D gel analysis. To that aim we had 4 specific objectives.

First objective is to use 2d gel analysis software: The application of proteomic tools such as two-dimensional polyacrylamide gel electrophoresis has become a powerful methodology to detect and examine changes in protein composition accurately. 2D gel is a powerful tool for identifying proteins that are differentially expressed across treatment conditions. Due to the number of proteins and the huge variation shown in gel images, the differential analysis of 2D gel images is challenging. In recent years, the application of proteomic tools such as two-dimensional polyacrylamide gel electrophoresis 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. The information about soybean seed proteins will be useful for scientist to understand its 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. Therefore, we developed a pipeline and database for soybean seed proteins which are easily accessible to the scientific community as well as the general public.

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 the soy products available in the market.

Second objective – RNA Seq analysis: here the objective is to calculate the amount of mRNA a gene is producing which can be measured by sequencing all the mRNA molecules in a sample and then aligning them to the genes in a given genome. The number of reads (sequences)

that align to specific genes can be used to measure differential expression of genes based on these counts, biologists can categorize the genes. We develop a relational database to store and manage the huge amount of data generated by these experiments. Then, we made the database publicly available via the World Wide Web, and developed web based a user interface to allow users to query the data. Our database is one of the first publicly available web based plant RNA-Seq databases.

Salicylic acid (SA), Jasmonic acid (JA), ethylene (Eth), and auxin (IAA) are amongst a collection of empirically studied plant hormones involved in the plant defense response. Otherwise known as phytohormones, these signaling molecules are regulated through numerous elegant and precise biochemical cascades in a tissue-specific manner. Upon pathogen attack, phytohormones drive synthesis of defense response elicitors in plants. Thus, quantifying host transcripts following phytohormone treatment and identifying transcription factor binding sites found in promoters of differentially expressed genes could reveal potentially novel insight into defense response signaling and co-regulatory interplay. In this study, we quantified transcripts in soybean roots following treatment with salicylic acid, jasmonic acid, ethylene and auxin. Statistical analysis of promoter sequences of differentially expressed transcripts revealed numerous over-represented transcription factor binding. Our results provide a biologically catalog of differential transcripts and corresponding binding sites over-represented following soybean root treatment with salicylic acid, jasmonic acid, ethylene and auxin.

Third objective is to compare the result of different analysis tools on soybean sequence data. Sequence analysis is a discipline that grew enormously in recent years. Analysis tools are combination of mathematics; statistics and computer science to analyze RNA-Seq. RNA-Seq analysis tools take the aligned reads from two or more conditions and reports genes and

transcripts that are differentially expressed. Each tool has its own protocol and different results. We collect the common result between some tools and visualize it to make it easy for biologists' scientist.

Our fourth and final objective is to visualize transcript expression onto biochemical pathways. Analysis results are always huge. Processing and comparing these huge data sets is difficult and time consuming, however visualization can make it easier. Visualization means to map quantified transcript expression onto biochemical pathways and makes these huge datasets traceable. We used Cytoscape software to map the results to pathways for time series and single time point transcriptomics analysis. It serves the result as visual pathways for transcriptomics knowledge discovery and functional annotation.

Chapter 2: A database for the identification of soybean proteins by 2D-Gel method

2.1 Introduction

Gel electrophoresis is a method of separating proteins based on their size and electric charge. In this method, an electric field is generated which applies force to the charged proteins causing them to move. The electric field is generated by cathode and anode probes place apart from each other. The proteins move in a Polyacrylamide gel medium. The motion of shorter molecules is faster in this medium whereas longer molecules experience more resistance along the way. This phenomenon causes the molecules with different lengths to separate, a process commonly called sieving. In addition to sieving the protein molecules, the gel medium maintains the proteins after sieving is complete. In the next steps, the proteins are stained and identified.

Two categories of gel may be used for this process. "Denaturing gel" unfolds the natural structure of molecules which turns them into a linear chain. As a result, the motion of the unfolded molecule primarily depends on its linear length and its mass-to-charge ratio. "Native gel" leaves the structure of the macromolecule unchanged. The physical size and structure of the protein also can affect the mobility of the molecule.

2D gels are widely used in molecular biology to measure whole-call protein expression. In a single gel, tens of thousands of proteins can be detected. The basics of it are very similar to 1D-PAGE except that two different characteristics of the protein are used to separate them in two physical dimensions. First, a method called Isoelectric Focusing (IEF) is utilized, in which the proteins move until they reach their isoelectric point. In IEF, the pH of the gel increases linearly from one end to the other. Immobilized pH gradient (IPG) strips could be employed to create the

pH gradient. In each pH level, the proteins are charged except for the isoelectric point where the charge will be zero. Therefore, applying the electric field shifts the proteins to their isoelectric point.

In the next step, the proteins are sieved in the perpendicular direction based on their molecular mass. The proteins are initially treated with sodium dodecyl sulfate (SDS) in a process which denatures them into straight molecules. Application of SDS negatively charges the proteins forcing them to have roughly the same mass-to-charge ratio. Similar to 1D-PAGE, an electric field is applied but in the perpendicular direction. As noted before, this will separate the molecules depending on their length.

Finally, the resulting gel is stained to reveal the spots where proteins are accumulated. The size, location, and density of these spots provide clues to identify the molecular weight and quantity of the proteins. The gels are commonly scanned to be accessible by others. These scans are augmented by reference maps created after identifying and labeling each protein.

Identification of soybean proteins is important since it is 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. 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 soybean proteins and a web based database named soybean protein database (SoyProDB) was subsequently built to allow scientists to search the data. This database is useful for both scientists and consumers. The database is freely accessible from:

http://bioinformatics.towson.edu/Soybean_Seed_Proteins_2D_Gel_DB/Home.aspx

2.2 Methodology

Protein was extracted from cultivated soybean seed, G. max PI 423954. Extraction procedures used were a modified TCA/acetone method and isopropanal 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.

2.3 Database methodology

The web based database is composed of two main parts. 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 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 2.1 is a snapshot of 2D gel's main page. Figure 2.2 is a snapshot of 2D gel's second page. Figure 2.3 is a snapshot of 2D gel's third page. The database is accessible from:

http://bioinformatics.towson.edu/Soybean_Seed_Proteins_2D_Gel_DB/Home.aspx

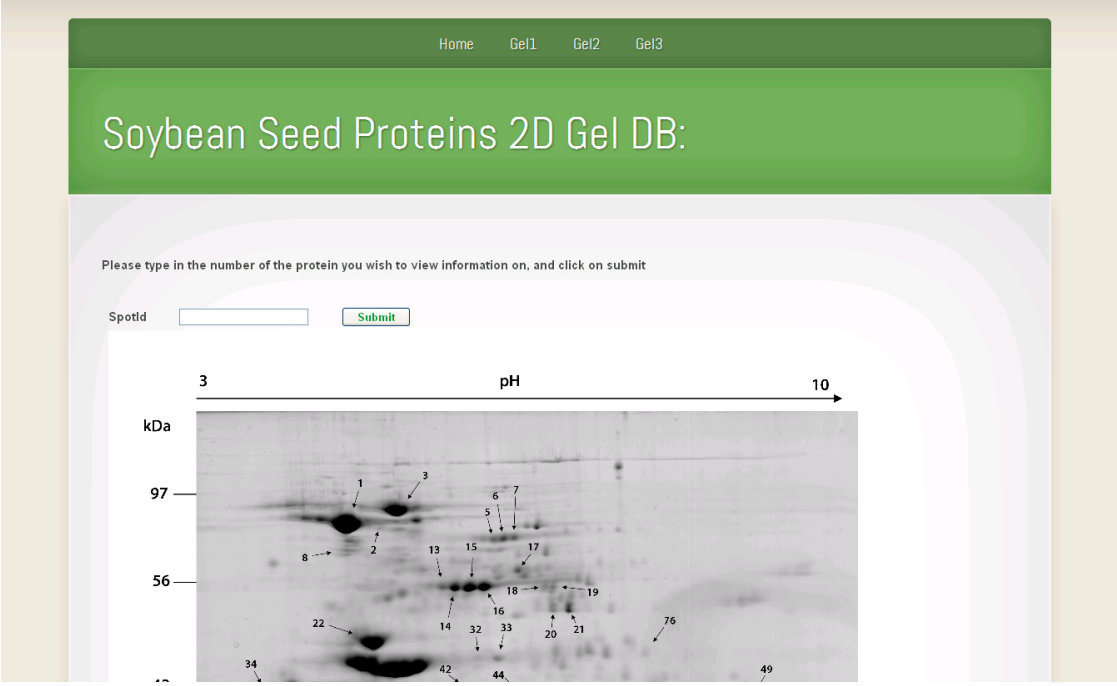


Figure 2.1 A snapshot of proteome database 2D gel interface

Figure 2.2 A snapshot of proteome database 2D gel interface

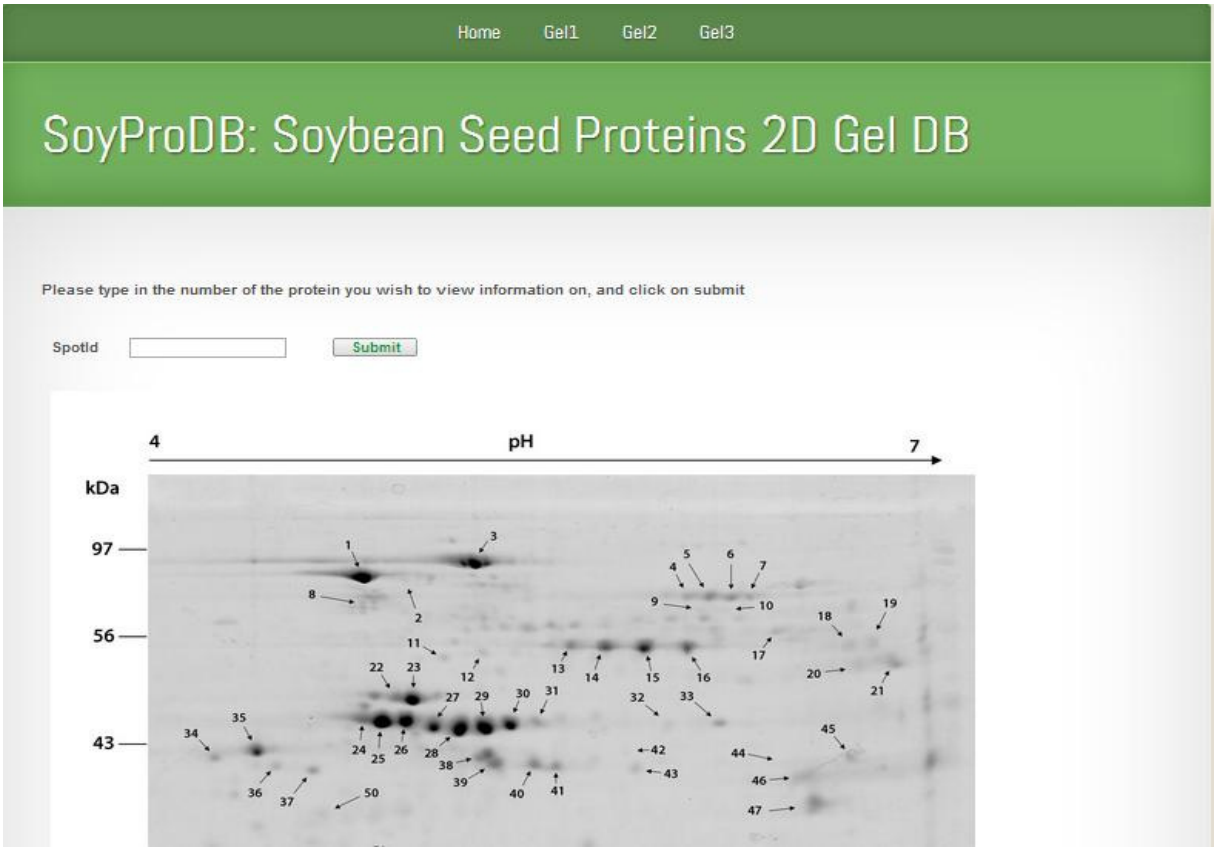


Figure 2.3 A snapshot of proteome database 2D gel interface

2.4 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.

2.5 Conclusion

Our analysis resulted in the identification of several soybean proteins by 2D-Gel method. 2D gels are used in molecular biology to measure whole-call protein expression. A web based database named soybean protein database (SoyProDB) was subsequently built to stores data. It allows scientists to search the data and retrieves the information of proteins. This database will

be useful to scientists who wish to genetically alter the soybean with higher quality storage proteins, and also helpful for consumers to get a greater understanding about proteins that compose the soy products available in the market.

Our collaboration with the USDA on this project resulted in peer reviewed publications:

- **Tavakolan M**, Alkharouf NW, Khan FH, Natarajan S."SoyProDB: A database for the identification of soybean seed proteins" Bioinformation 2013.
- **Tavakolana M**, Natarajanb S, Alkharouf NW and Mathews B. "LAProDB: A database for soybean low abundance proteins" In preparation for Bioinformation 2014.

Chapter 3: SCNProDB: A database for the identification of soybean cyst nematode proteins

3.1 Introduction

The soybean cyst nematode (SCN) is the major pest of soybean in the U.S. and causes ~457 to 819 million dollars in yield losses annually between 2003 and 2005[11]. 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 [12]. The second-stage juvenile (J2) is important, because it is the only motile stage. 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 is available [13, 14, 15, 16]. 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.

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. Our analysis resulted in the identification of several hundred proteins and a web based database (SCNProDB) containing protein information. The database is freely accessible from:

http://bioinformatics.towson.edu/Soybean_SCN_proteins_2D_Gel_DB/Gel1.aspx

3.2 Methodology

The nematodes were grown at the United States Department of Agriculture, Beltsville MD, USA. According to Klink et al [14], proteins were extracted from J2 nematodes using a modified phenol extraction procedure [18]. The nematode protein samples (400µg) were separated in the first dimension using 13cm (Immobiline IPG Drystrips (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 stained 2D-PAGE gels with Colloidal Coomassie Blue G-250 were scanned using ImageScannerTM 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 (ThermoFisher Scientific, San Jose, CA) was used for protein analysis. The resulting peptides were separated by reverse phase chromatography and searched against NCBInr databasesand the invertebrates EST database [17].

3.3 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 “SpotID” 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: protein description, Theoretical_PI/Mr, Uniprot_Acc, peptides_matched, EST_coverage, 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 3.1 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

3.4 Utility of the biological community

The database is of interest to biologists working with soybeans 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.

Home Gel1 Gel2 Gel3

SCNProDB

Please type in the number of the protein you wish to view information on, and click on submit

SpotId 50 Submit

Spots_no	Protein_id	Theoretical_PI/Mr	Gi_Acc	Uniprot_Acc	peptides_matched	EST_coverage(%)	MOWSE_score	species	EST_species	EST_Acc	blast_e_value
50	Probable voltage-dependent anion-selective channel	9.20/29961	gi|21264505	Q21752	12	0.68	807	Caenorhabditis elegans	Heterodera glycines	CK349300	3 e-44

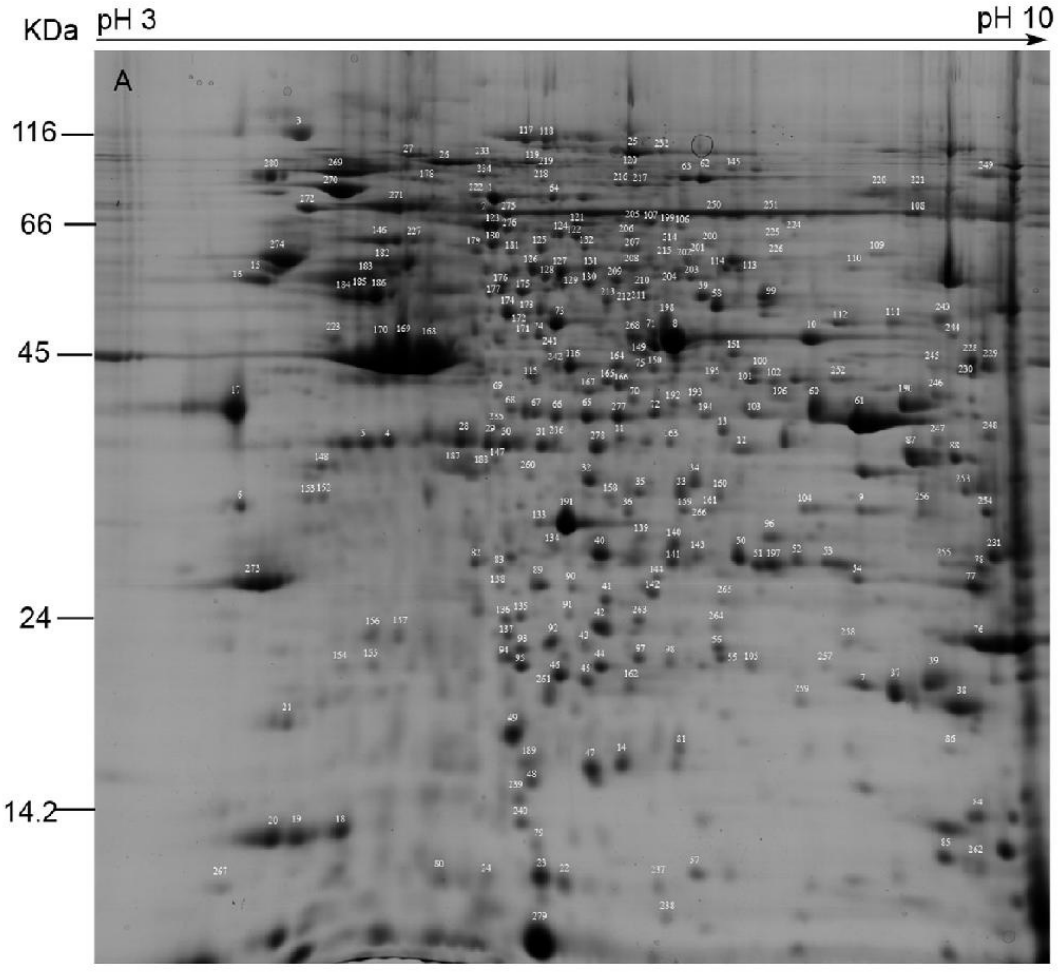


Figure 3.1 A snapshot of SCNProDB

3.5 Conclusion

We have conducted studies to determine the composition of SCN proteins using a proteomics approach using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) to separate SCN proteins and to characterize the proteins further. Our analysis resulted in the identification of several hundred proteins. In this investigation, we developed web based database (SCNProDB) containing protein information. This database will be useful to scientists who wish to develop SCN resistant soybean varieties through genetic manipulation and breeding efforts.

Our collaboration with the USDA on this project resulted in a peer reviewed publication: Natarajanb S, **Tavakolana M**, Alkharouf NW and Mathews B. "SCNProDB: A database for the identification of cyst nematode proteins using 2D Gels" Bioinformation 2014.

Chapter 4: Analysis of next generation sequence data related to soybean treated with different hormones

4.1 Introduction to plant hormones

Phytohormones are plant hormones which regulate virtually every aspect of plant development. Systematic interplay amongst such molecules, even at very low concentrations, drives biochemical signaling cascades which govern processes such as seed germination, leaf senescence, and flower differentiation. In the face of herbivory and pathogenesis, this tightly regulated interplay, synergistic or antagonistic, is necessary for orchestrated and perfectly timed signaling cross-talk. Throughout the past several decades, our understanding of plant hormones and their corresponding receptors has grown orders of magnitude, with a bulk of such advances discovered in the Arabidopsis thaliana model. Amongst the first empirically studied phytohormones were auxin (IAA), abscisic acid (ABA), gibberellin (GA), and ethylene (ETH). These phytohormones were termed "classical" phytohormones due to their early discovery and core ability to regulate plant organ differentiation [19, 20]. More recently discovered signaling molecules, such as jasmonic acid (JA), salicylic acid (SA), and brassinosteriods (BS), have also joined the phytohormone catalog. Upon perception of external stimuli, such as pests or pathogens, virtually all phytohormones are involved in systematic cross-talk [21, 22, 23, 24]. Thus, treating plant tissues with various phytohormones and quantifying transcript abundance in these samples could provide a transcriptomic into both systematic cross-talk and defense-response signaling.

In this study, we treated soybean roots with one of four phytohormones: SA, JA, IAA, and ETH. Therefore, we analyzed them to find out differentially expressed genes. Our results

provide heavily induced, moderately induced, weakly induced, weakly suppressed, moderately suppressed, and heavily suppressed across all treatments relevant to SA, JA, IAA and Eth regulation.

4.2 Background on RNA-Seq

The vast amount of data coming from the genome sequencing programs will be of limited use unless it can be linked to ways of measuring. Measuring the expression of selected genes needs comparison of gene expression patterns under different conditions, for example control and sample tissues. Gene expression begins when genes are transcribed into messengers RNAs, which are then translated to produce proteins. The transcriptome and proteome, unlike the genome, are changeable in response to conditions, and depend on the state of development, the environment, and the type of tissue. Most gene expression experiments are intended not simply to detect the genes being expressed at a given time, but to detect differences in gene expression under different conditions.

Devising ways of comparing sequences has never been straightforward, not just because of the vast amount of information. The difficulties arise because of the many ways DNA and proteins can change during evaluation. Mutation and selection over millions of years can result in considerable divergence between present day sequences derived from the same gene. Bases at originally corresponding positions, and the amino acids they encode, can change as a result of point mutation, and the sequence length can be quite different as a result of insertions and deletions. Gene duplicate are common as well, and in many cases mutation has disabled one copy of a gene so that is either no longer express or, if transcribe, does not produce a functional protein [21, 22].

On superficial inspection, such changes in gene sequence and length can effectively mask any underlying sequence similarity. To reveal it, the sequences have to be aligned with each other to maximize their similarities. Alignment methods are at the core of many of the software tools used to search the databases, such as Bowtie tool.

As the result of mutation, even the sequences of the same protein or gene from two closely related species are rarely identical. Ideally, what we want to achieve when comparing sequences is to line up them in such a way that, when they do drive from a common ancestor, bases or amino acids derived from the same ancestor base are aligned. Without information to the contrary, this is the best achieved by maximizing the similarity of aligned regions.

Sometimes because of the difference in length, it also creates false matches between non corresponding positions. To solve this problem, gaps are introduced into one or both the sequences so that maximum similarity is preserved. There is never just one possible alignment between any two sequences, and the best one is not always obvious, especially when the sequences are not very similar to each other. At the heart of sequence comparison and database-searching methods are algorithms for testing the fit of each alignment generated, giving it a quantitative score, and filtering out the unsatisfactory ones according to preset criteria [22, 23].

Alignment can reveal homology between sequences. Homology genes are the genes derived from the same ancestral gene. During their evolutionary history they will have diverged in sequence as result of accumulating different mutations.

Because homology implies a common ancestor, it can also imply a common function or structure for two homologous proteins, which can be a useful pointer to function if one of the proteins is known only from its sequence.

The operation of natural selection tends to result in the acceptance of mutations that preserved the folding and function of a protein, whereas those that destroy folding or function will be eliminated. However, similar or identical aligned residues may simply be due to relatively recent divergence of the two sequences, and so care must be taken not to overestimate their functional importance. Moreover, mutation and selection can generate proteins with new functions but relatively little change in sequence. Therefore, sequence similarity does not always imply a common function.

Conversely, there are proteins with very little sequence similarity to each other but in which a common protein fold and function are preserved. Consequently, low sequence similarity does not necessarily rule out common function or homology. Such cases require extra information, such as structural or biochemical knowledge to demonstrate their true relationship.

Sequences can also be significantly similar to each other, and yet not be evolutionarily homologous, as a result of convergent evolution for similar function. In this case, identical or very similar aligned residues can be argued to have an important functional role. Convergent evolution does not, however, usually produce highly similar sequences of any great length.

All these consideration have to be taken into account when analyzing the result of a database search. An alignment of two sequences is, in effect, a hypothesis about which pairs of residues have evolved from the same ancestral residue. But an alignment in itself does not imply an evolutionary order of events, so that the two alternatives of homology and convergent evolution cannot usually be distinguished without additional information [23, 24].

Sequence comparison methods have to take account of such factors as the type of mutation that occur over evolutionary time, differences in the physicochemical properties of

amino acids and their roles in determining proteins structure and function, and the selective pressures that result in some mutations being accepted and other being eliminated. One has to consider the evolutionary process that are responsible for sequence divergence and find a way to include the salient features in practicable schemes for testing the goodness of fit of the alignment. These must be quantitative and hence involve a score. Such scoring schemes can be incorporated in algorithms designed or generate the best possible alignments. Finally, ways must be found to discriminate between fortuitously good alignments and those due to a real evolutionary relationship.

4.3 Methods

Soybean plants were grown in a green-house under natural light at 25 C. Two weeks after sowing plants, roots from plants were treated with a hormone. Sets of four roots were treated with a single hormone, SA, JA, ETH and IAA, respectively. SA, JA and IAA were dissolved in 1 mL ethanol and dispersed in water to produce 5 mM and 0.5 mM concentrations solutions, respectively. Roots of each of four plants were submerged into 250 mL containing SA, JA, IAA or Eth for 8 hr in the dark. Plants treated with ETH were placed in an airtight cabinet in the dark. ETH was injected to achieve a final concentration of 1 l/L. To serve as a baseline, two control plants were also grown and maintained in the dark for 8 hr, but treated with none of the SA, JA, IAA and ETH.

Total RNA was extracted from roots of each plant using the RNeasy Mini Kit. Two RNA libraries were generated from untreated roots as well as roots treated with SA, JA and IAA. For roots treated with ETH, only one RNA library was generated. cDNA libraries were generated for all nine RNA libraries and sequenced using the Illumina paired-end sequencing recipe by expression analysis.

Paired-end FASTQ files from each of the sequenced lanes were mapped to the Phytozome Glycine max transcriptome [25] using the Bowtie short-read aligner [26]. Reads mapping to more than one transcript were assigned to the transcript with the highest alignment score.

Figure 4.1 shows the bioinformatics pipeline used for the RNA-Seq analysis. The pipeline can be divided into the following steps: 1) Gene models functional annotation, 2) RNA-Seq alignment to reference genome/gene predictions, 3) Gene expression estimation, 4) Identify differentially expressed genes.

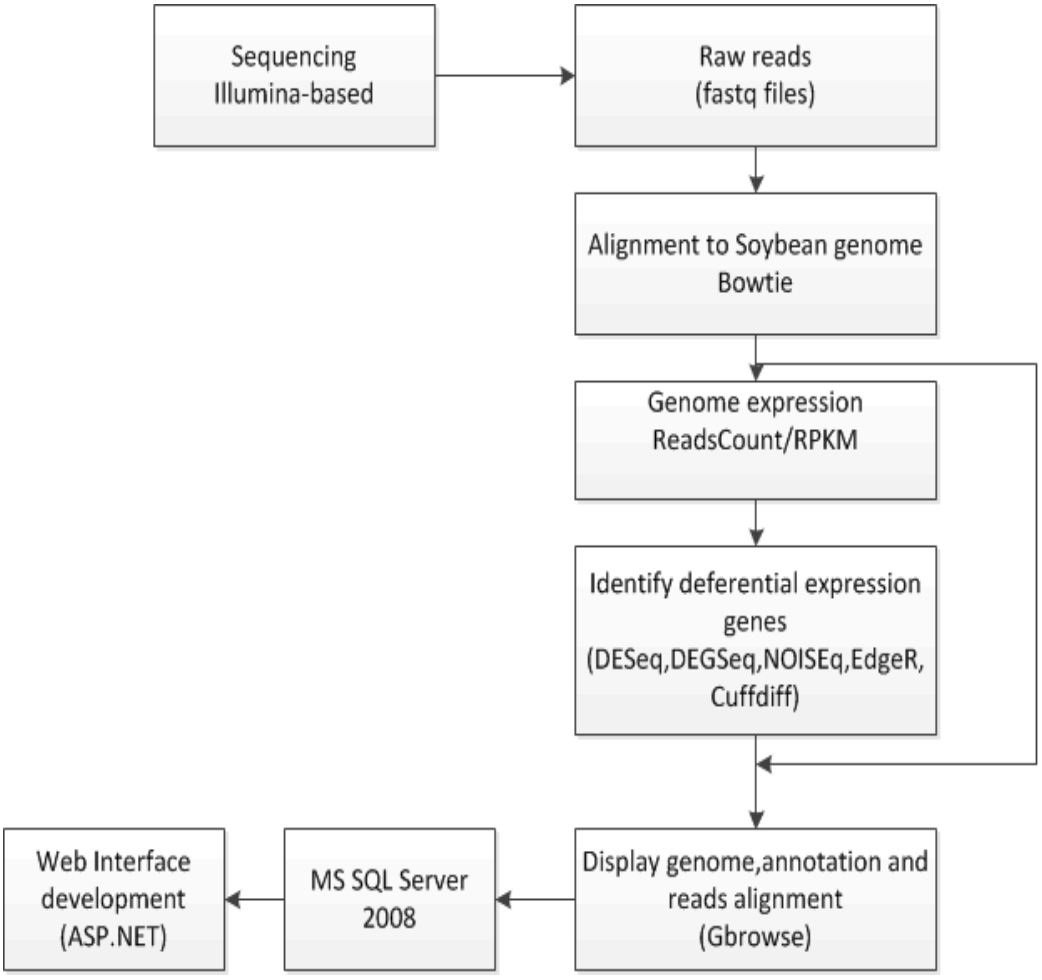


Figure 4.1 Shows the bioinformatics pipeline used for the RNA-Seq analysis

Short reads will be aligned to transcript. It includes several outputs such as number of reads, numbers of alignment, percentage of alignment and number of reads which fail to align. For alignment we use bowtie.2. Bowtie is an ultrafast, memory-efficient short read aligner geared toward quickly aligning large sets of short DNA sequences (reads) to large genomes. Bowtie can also output alignments in the standard SAM format, allowing Bowtie to interoperate with other tools supporting SAM, including the SAMtools. Bowtie runs on the command line under Windows, Mac OS X, and Linux. The first argument to bowtie is the index for the genome to be searched. The second argument is the name of a FASTQ file containing the reads. Bowtie print many lines of output. Each line is an alignment for a read. The name of the aligned read appears in the leftmost column. The result of bowtie is a SAM file which is a huge file. The UNIX shell command used to build the reference index:

Bowtie-build ref_seq.fa ref_seq_index

Where ref_seq.fa is the FASTA file of the reference, this file can be single/multi FASTA formatted. ref_seq_index is the name of the generated index. The UNIX shell command used to perform the alignment:

Bowtie -v -S ref_index input_file output_file

Where -v is the option to print a summary on the terminal after the alignment process has finished, -S is the option to write the resulted alignments in SAM format, ref_index reference index file name, input_file is the sequence reads file in FASTQ format and output_file is the output file name.

The perl scripts were written to extract the number of reads for each gene. Alignment output will be parsed and then used to calculate important variables such as (reads count per

gene, RPKM as normalize value). Sometime more reads mapped to a transcript if it is long so normalization is needed. By normalization total sequence from different conditions can be compared and feature of different length would be considered. Here we are using RPKM value as a normalize value. RPKM stands for Reads Per Kilo base per Million mapped reads. The formula to calculate it is as below:

RPKM: C/LN

Where C is number of map able reads on a feature (e.g. transcript, exon, etc.), L is length of feature (in kb) and N is total number of map able reads (in millions). Based on reads count per gene, differential expression calculated by DEGSeq software and used to categorize genes to suppressed, unchanged and induced groups. The statistical package R used for analysis.

DEGseq identify differentially expressed genes between two samples with technical replicates directly. This function is useable when users already have the gene expression values such as the number of reads mapped to each gene. DEGseq has different arguments. geneExpFile is a file containing gene expression values for replicates of sample and it is the input file for DEGseq. geneCol is a gene id column in geneExpFile. expCol is expression value columns in geneExpFile for replicates of sample. Here, there are 4 columns of expCol which replicate1 and replicate2 are two (groups of) technical replicates of a sample and other two are replicates of the controls. GroupLabel modify the label of group on the plots. The first GroupLable is sample and the next one is control otherwise the result of DEGseq program would be vice versa. Header is a logical value indicating whether geneExpFile contains the names of the variables as its first line. Header has to be TRUE if the file contains the names of the variables. Sep argument is the separator character of the input file. If sep = "" which is the default for read.table the separator is

white space, that is one or more spaces, tabs, newlines or carriage returns. Method modifies the method to identify differentially expressed genes. DEGseq will do normalization on its process.

After running DEGseq, it produces a file as an output. The output file has an each gene id with corresponding values. These values are such as pValue, zScore, qValue, foldChange. OutputDir shows the output directory.In the context of this study, outlier transcripts were defined with statistically significant log2 fold-change variances following two treatments.

Following each of the four phytohormone treatments, Phytozome accession identifiers for the top 400 induced soybean transcripts were retrieved. For all accession identifiers, its promoter sequence 2kb upstream from its transcription start site was subsequently identified. Thus, four sets of promoter sequences were produced, each termed a query set. To effectively identify transcription factor binding sites (TFBS) over-represented within each query set, a matching control set was generated from promoter sequences of non-differential transcripts following all treatments. Each control set was generated to match length distribution of its respective query set and differ in GC content by at-most 2%. To model TFBSs, 71 publicly available position weight matrices (PWMs) were retrieved from AthaMap [29]. The Marina software [30] was used to statistically contrast abundance of all 71 PWMs within each query and control set. Following abundance analyses, a PWM abundance matrix was generated by Marina to represent frequency of individual PWMs within all promoter sequences. To effectively quantify magnitude of TFBS classification to either the query or control set, a LASSO classifier was built for each matrix using the glmnet R package [31].

Each classifier was built using 10-fold cross-validation whereby the model was built using nine-tenths of the matrix and tested on the remaining one-tenth. Magnitude of TFBS

classification was modeled as real-number weights, with positive and negative weights indicative of classification towards the query or control set, respectively. All TFBSs with negative weights were assigned a weight of 0 to indicate no classification towards the query set. Approximately 290 million 50bp paired-end reads were generated upon sequencing all cDNA libraries. Phred quality (q) scores throughout all lanes were either 38 or 39, indicating base calling accuracy of approximately 99.99% (Table 4.1). Similarly, percentage of reads mapping to the soybean transcriptome consistently ranged between 89% to 93%, averaging a respectable alignment percentage of 90%. To determine linearity between each replicate, Pearson (p) correlation coefficients were derived given read counts, shown in Table 4.1. The high Pearson coefficients associated with these treatments (p>=0.98) thus indicates presence of robust, biologically-sound replicates.

Treatment	Replicate	# reads	Map (%)	SRA	q-score
Untreated	1	42,820,946	89	SRR976388	38
	2	40,705,166	89	SRR976389	38
JA	1	40,631,916	90	SRR976390	39
	2	43,081,146	89	SRR976392	38
IAA	1	40,906,124	91	SRR976393	38
	2	39,896,498	92	SRR976395	38
ETH	1	39,837,842	90	SRR976396	38
Total	-	287,879,638	mean=90	-	mean=38

Table 4.1 Sequencing the soybean root transcriptome upon treatment with four phytohormones. On average, approximately 90% of reads mapped to the soybean transcriptome

Of the 73,026 transcripts and 53,917 genes making up the Phytozome Glycine max 1.1 transcriptome build, 6,632 transcripts and 5,498 genes were rendered DE upon treatment with at least one phytohormone. DEG transcripts were subsequently stratified based on their log2 fold-change, f, and classified into one of six categories: heavily induced (f >= 5), moderately induced (2 <= f < 5), weakly induced (1.5 <= f < 2), weakly suppressed (-2 < f < -1.5), moderately

suppressed (5 < f <= -2), and heavily suppressed (f<= -5). Across all treatments, heavily induced and heavily suppressed transcripts were the least abundant transcript categories.

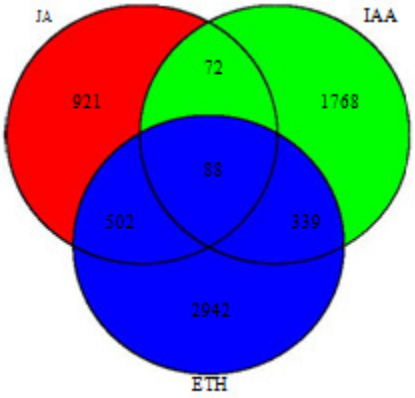


Figure 4.2 Transcript distributions post-treatment and replicate linearity

As shown in Figure 4.2, of the 6,632 DE transcripts, 88 transcripts were DEG following all three treatments. Almost one-third of these transcripts (n=34) were collectively induced. Of these transcripts, 15 transcripts were identified to functionally encode numerous enzymes central to metabolism and biosynthesis such as phosphoethanolamine (EC: 3.1.3.75), glucosucrase (EC: 3.2.1.26), glutathione S-transferase (EC: 2.5.1.18), and flavonol synthase (EC: 1.14.11.23).

Thus, there exists a core set of induced transcripts following SA, JA, IAA and Eth treatment which collectively contribute towards the regulation of critical basal operations. Following all treatments, the top 200 most induced transcripts DEG solely after at least one treatment were identified and termed a transcript set. For each transcript set, the top 20 enriched Gene Ontology (GO) Processes were identified. Such results could therefore provide a glimpse into systematic interplay orchestrated solely by each individual phytohormone. Of the 20 enriched GO Processes within induced transcripts, 10 were exclusively enriched following IAA treatment. Such GO enrichments were exclusively associated with classical IAA regulatory roles

such as root development and post-embryonic differentiation.

An additional 2 enrichments ("response to jasmonic acid stimulus" and "secondary metabolic process") were statistically significant following both JA and ETH treatments. This mutually-shared set of ontologies should come as no surprise considering the well studied co-regulation between JA and ETH. Two of the most enriched GO Processes within induced transcripts were "oxylipin metabolic process" and "oxylipin biosynthetic process", both exclusive to transcripts following JA treatment. This should come as no surprise since oxylipins represent a family of fatty acids such as JA, signaling molecules known for their ability to regulate stress response signaling [35, 36, 37].

Contrasting GO enrichments within induced transcripts to those from suppressed transcripts yields pronounced differences between the two enrichment sets. Such results shed light on systematic interplay by revealing expression profiles that could potentially capture individual hormone-driven signaling dynamics.

A total of 44 transcripts were rendered outliers following JA and ETH treatment. Of this set, 16 transcripts were induced following both treatments. An additional 11 transcripts were identified to be induced following JA treatment but suppressed following ETH treatment. One such loci, Glyma02g09130.2, encoded aAlliin lyase (EC: 4.4.1.4), an enzyme shown to play critical roles in herbivory stress response [38, 39]. An additional subset of 16 transcripts were induced following ETH treatment and suppressed following JA treatment. Such transcripts encoded numerous transcription factors (TF) critical to ethylene interplay. Four genomic loci from this set (Glyma13g18330.1, Glyma13g18370.2, Glyma19g34670.2, Glyma10g04170.2) encoded an AP2/ERF TF, a regulatory protein critical to defense and ethylene signaling [40, 41]. Two additional soybean loci (Glyma14g06710.1, Glyma01g32310.1) encoded shikimate O-

hydroxycinnamoyltransferase (EC: 2.3.1.133) and Peroxidase (EC: 1.11.1.7), respectively. Both enzymes are well-known for their roles in flavonoid and phenylpropanoid biosynthesis, two key defense signaling pathways [42].

A set of 36 transcripts were identified as outliers following both JA and IAA treatment. Interestingly, only 3 of these transcripts (Glyma13g06880.1, Glyma07g05740.1, Glyma03g27120.1) were induced after JA treatment but suppressed after IAA treatment. Of the remaining 33 transcripts, 20 were induced following IAA treatment but suppressed following JA treatment. Surprisingly, five of these 20 transcripts (Glyma14g06710.1, Glyma13g18330.1, Glyma13g18370.2, Glyma10g04170.2, Glyma19g34670.2) were also outlier transcripts following JA and ETH treatments. Of these 20 outliers, the most induced following only IAA treatment was Glyma16g03600.1, a vital gene encoding 1-aminocyclopropane-1-carboxylate synthase (ACC; EC: 4.4.1.14), a key regulator of ETH biosynthesis [43, 44]. Lastly, 13 of the 36 total outlier transcripts were induced following both IAA and JA treatment. Four of these transcripts (Glyma17g18040.1, Glyma05g21680.1, Glyma02g13910.1, Glyma17g18080.1) encoded GH3 proteins which have been shown in soybean plants to undergo rapid induction following IAA treatment [45].

A set of 72 outliers were identified following both ETH and IAA treatment. From this set, four transcripts (Glyma02g09130.2, Glyma18g06230.1, Glyma13g34221.1, Glyma06g14640.1) were induced solely after IAA treatment but suppressed after ETH treatment. A further four outlier transcripts (Glyma05g21680.1, Glyma17g18080.1, Glyma17g18040.1, Glyma10g04160.1) were induced following both IAA and ETH treatment.

The database is created to provide an integrated view of results. Based on the result, up-regulated and down-regulated genes have been categorized. Whenever, the signature is “TRUE”

it means gene is either up-regulated or down-regulated. Positive fold change determines up-regulated and negative fold change classifies down-regulated gene.

4.4 Conclusion

In this study, we treated soybean roots independently with one of four phytohormones: SA, JA, IAA, and ETH. Transcripts isolated from treated roots were sequenced using RNA-Sequencing (RNA-Seq), generating almost 300 million high-quality reads. Functional analysis of differential transcripts revealed numerous distinct transcriptomic profiles following either individual or pairs of treatments. In addition, statistical analysis of soybean transcription factor binding sites identified numerous proximal regulatory elements in promoters of differentially expressed genes. Our results provide a potentially novel catalog of biologically-sound and over-represented soybean transcription factor binding sites relevant to SA, JA, IAA and Eth regulation.

Chapter 5: A database for analysis of next generation sequence data related to soybean treated with different hormones

5.1 Introduction

Recently, scientists have focused on identifying different proteins in soybean. Plant hormones play a crucial role in growth and development. Little is understood about their mode of action though, therefore studying their effects at the molecular level is imperative. We have developed a relational database to store different soybean proteins treated with various phytohormones.

It is a web-based database which allows users to have an integrated view of results. Based on the result, up-regulated and down-regulated genes have been categorized. The database was developed using Microsoft SQL Server 2008 to house experiment data, annotations and gene expression profiles of Soybean genes. It can be accessed publicly from a web-based interface; this website provides search and browse functionalities to allow scientists to access and search the data. It can be accessed from:

http://bioinformatics.towson.edu/Soybean/

5.2 Background

Genes encode proteins and proteins modify cell function. Protein production starts at transcription DNA to messenger RNA (mRNA). The mRNA is then translated into protein. Thus, control of these processes plays a critical role in determining what proteins are present in a

cell. In addition, the way in which a cell processes its RNA transcripts and newly made proteins also greatly influences protein levels.

The amount of mRNA that hits a gene can be measured by sequencing all the mRNA molecules in a sample and then aligning them to the genes in a given genome. The number of reads (sequences) that align to specific genes can be used to measure differential expression of genes. Biologists can categorize the genes based on these counts [19, 20].

5.3 Description of the Entities and Relationships

The SDEG database was designed, implemented, and hosted using Microsoft SQL Server 2008 Enterprise Edition. The entity-relation diagram is shown in Figure 5.1. The diagram shows the tables where the RNA-Seq analysis is stored, the attributes of each of the tables, and the relationships between the tables. The SDEG database contains 10 tables that vary in the number of attributes and the size.

SDEG stores are relevant data relating to every RNA_Seq experiment, such as gene No, reads count, and fold-change, significant. SDEG stores RNA_Seq experiments, and each experiment is given a unique sample ID. Each experiment has been done in different conditions.

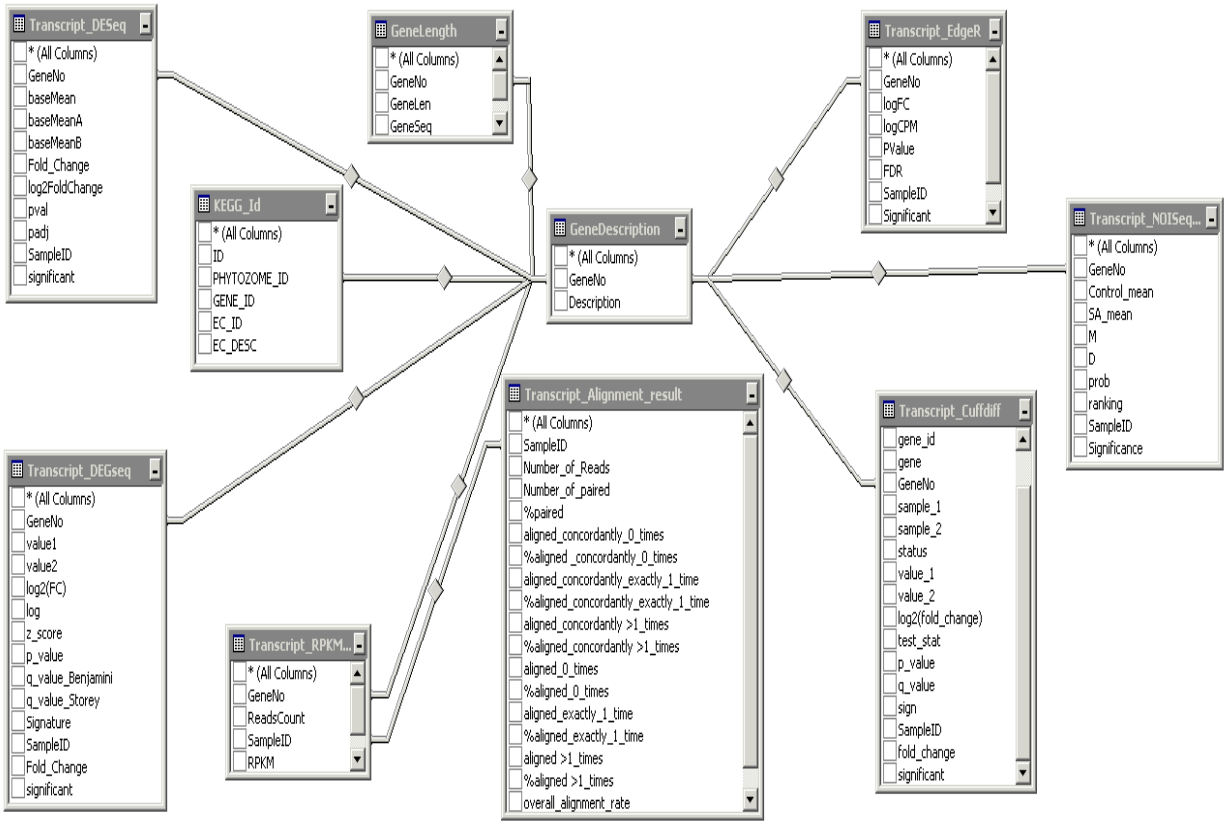


Figure 5.1 ER diagram

5.4 SDEG Web Interface

SDEG is a relation database built on SQLServer2008 and is housed at Towson University in MD. The database server is housed on a dual Intel Pentium server with processor speeds of 1 GHz each with 1 GB of memory. The web server is Microsoft's IIS (Internet Information Services) version 6.0 running on Windows 2003 Server operating system. Active Server Pages (ASP) and ASP.Net were used to build the user interfaces and query pages. SQL code is embedded within ASP to query the database.

SDEG's web site is a secure site requiring user authentications and is password protected. The web interface to SDEG is divided into five major categories, search gene by No, search gene by DES, Accessdata, compare sample_DEGSeq, transcript compare sample_DEGSeq.

Search gene by No provides users to search genes by their numbers. It shows the gene number, the description for the gene and the RPKM values in all the samples individually as it is shown in the Figure 5.2.

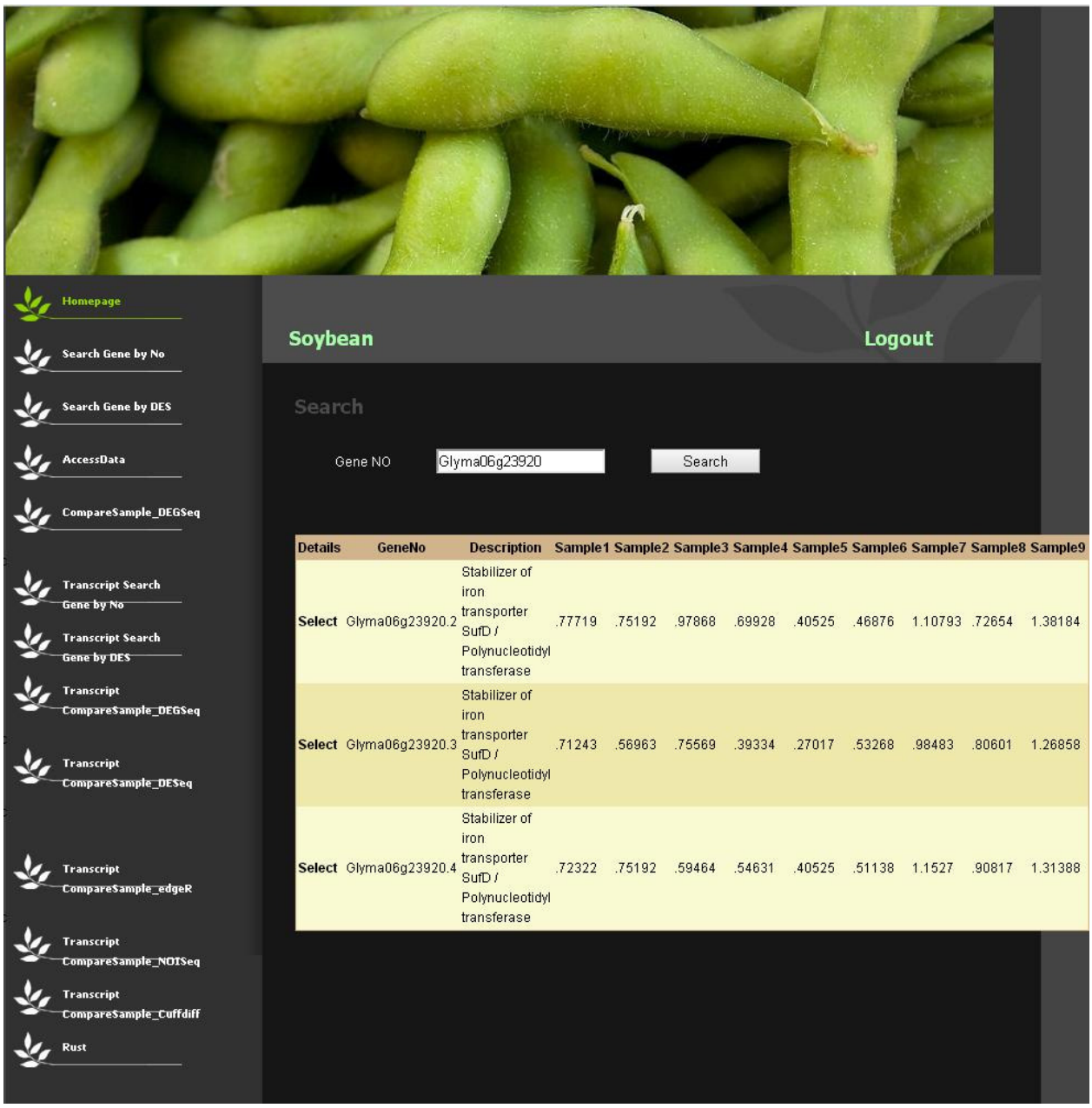


Figure 5.2 Search gene by No

In the case that user are more interested in the details, the select option next to gene number can be selected. As the result, gene No, gene length, description, sequence of gene will be shown. The Figure 5.3 shows the result.

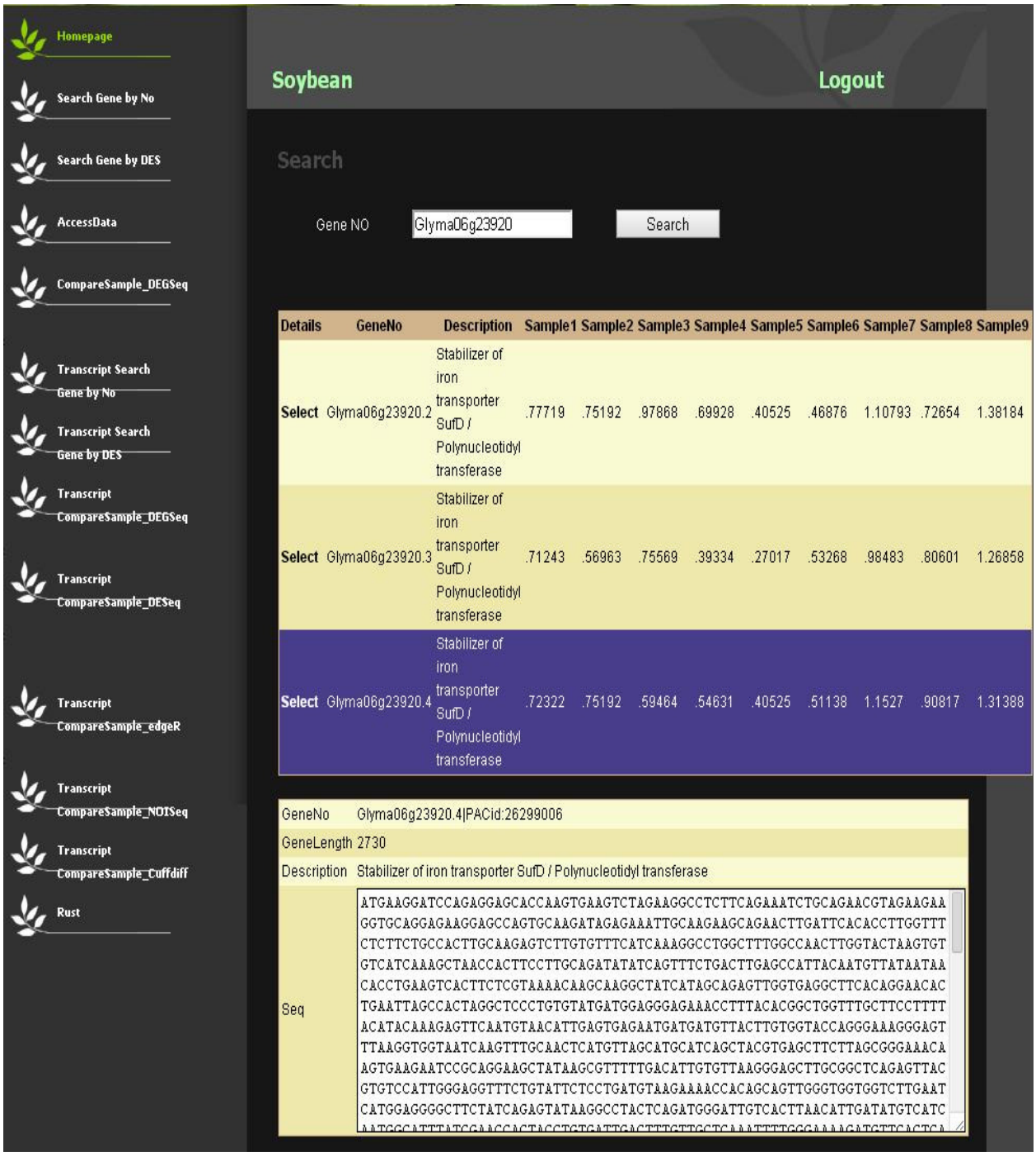


Figure 5.3 Search gene by No (2)

Search gene by DES provides users to search genes by their descriptions. It shows the gene description and the number of the gene and the RPKM values in all the samples individually as it is shown in the Figure 5.4.

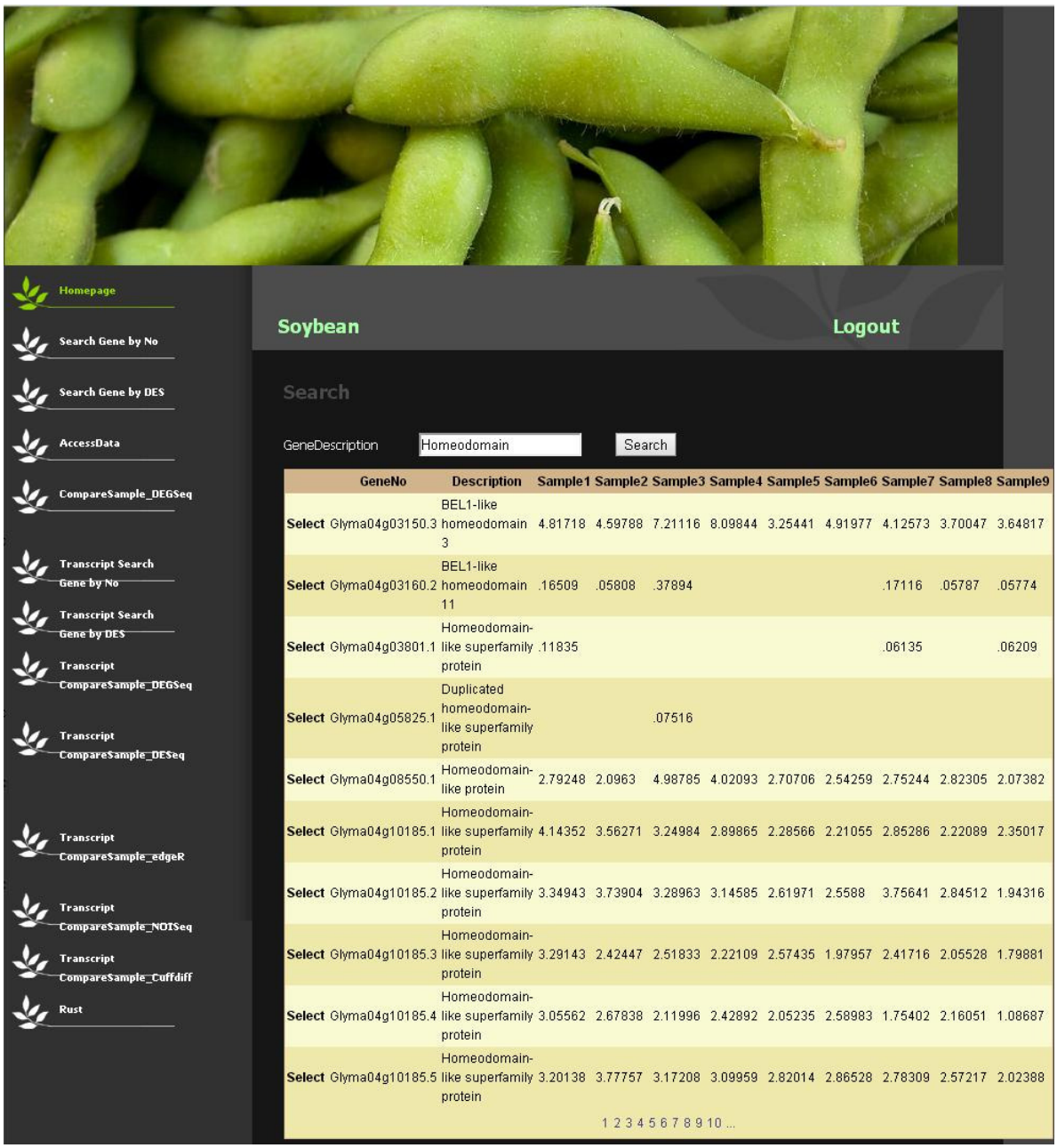


Figure 5.4 Search gene by DES

If user is more interested in the details, the select option next to gene number can be selected. As result, gene No, gene length, description, sequence of gene will be shown. The Figure 5.5 shows the result.

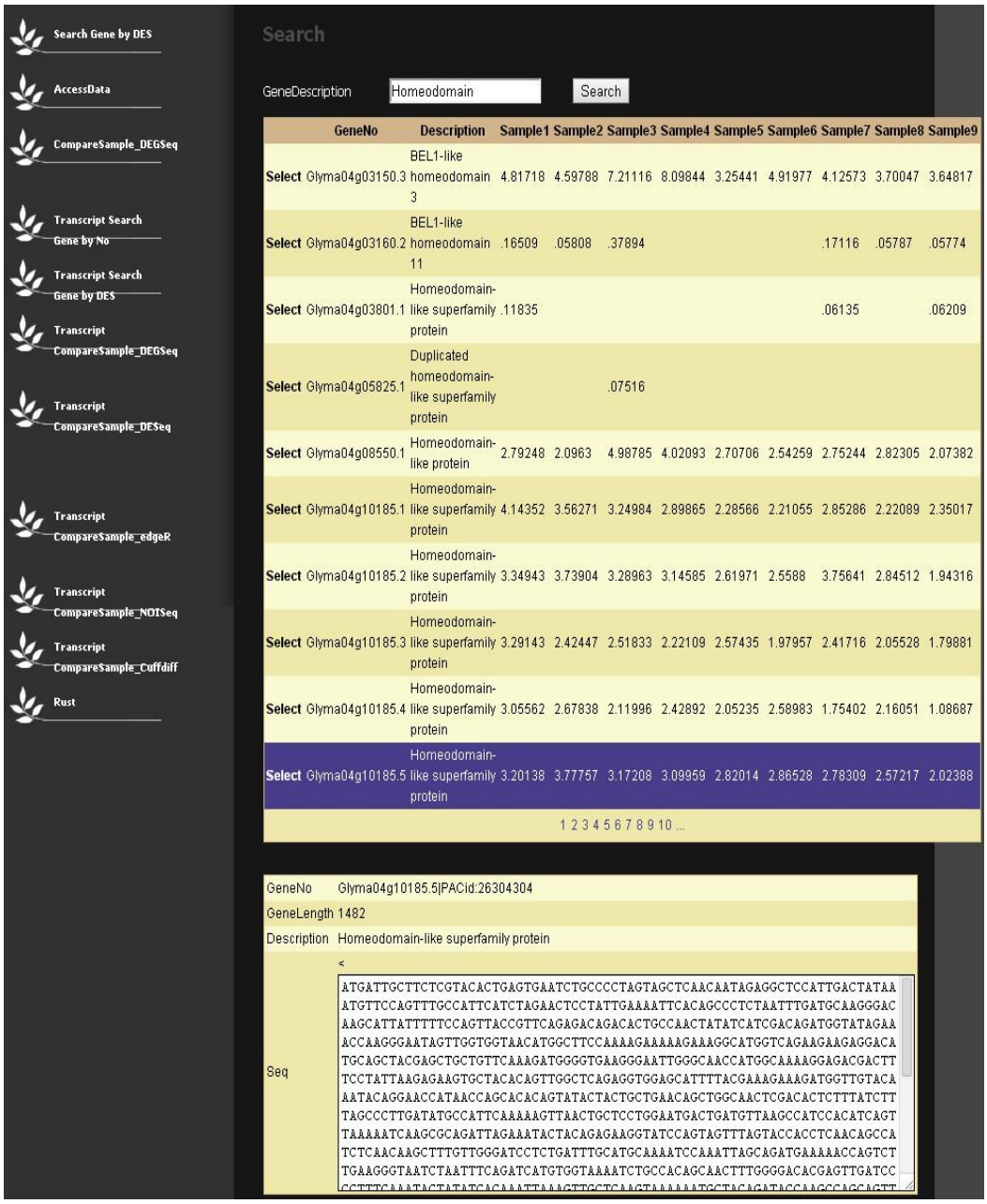


Figure 5.5 Search gene by DES (2)

Access data shows number of reads, number of reads which aligned zero time, number of reads which aligned exactly one time, number of reads which aligned more than one times for each experiment. These numbers are the result of bowtie alignment. Figure 5.6 shows the access data page for all the experiments.

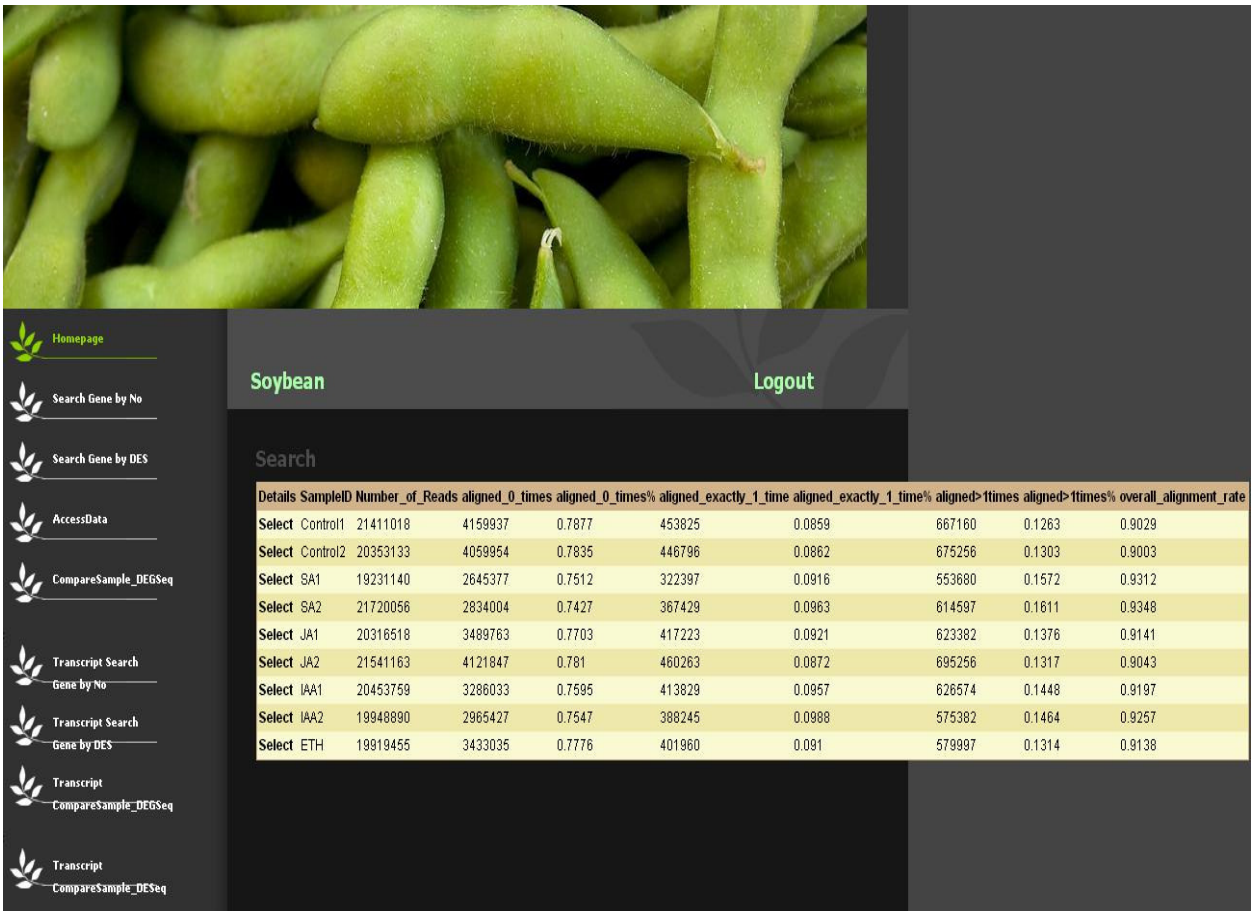


Figure 5.6 Access data

User can choose each of the experiment by clicking select label next to it and it shows gene No, gene description, reads count, RPKM of the genes on that experiment. Figure 5.7 shows the values for one of the experiment.

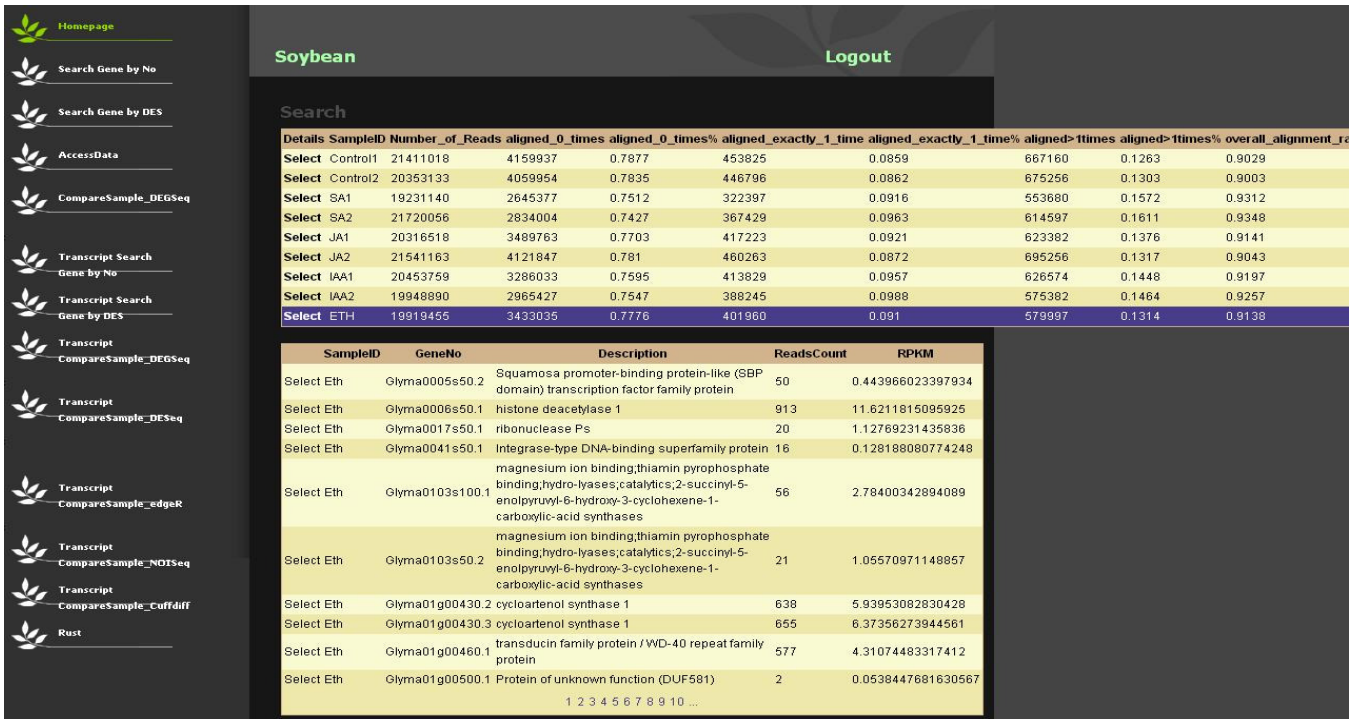


Figure 5.7 Access data (2)

If user is more interested in the details, the select option next to gene number can be selected. As the result, gene No, gene length, description, sequence of gene will be shown. The Figure 5.8 shows the result.

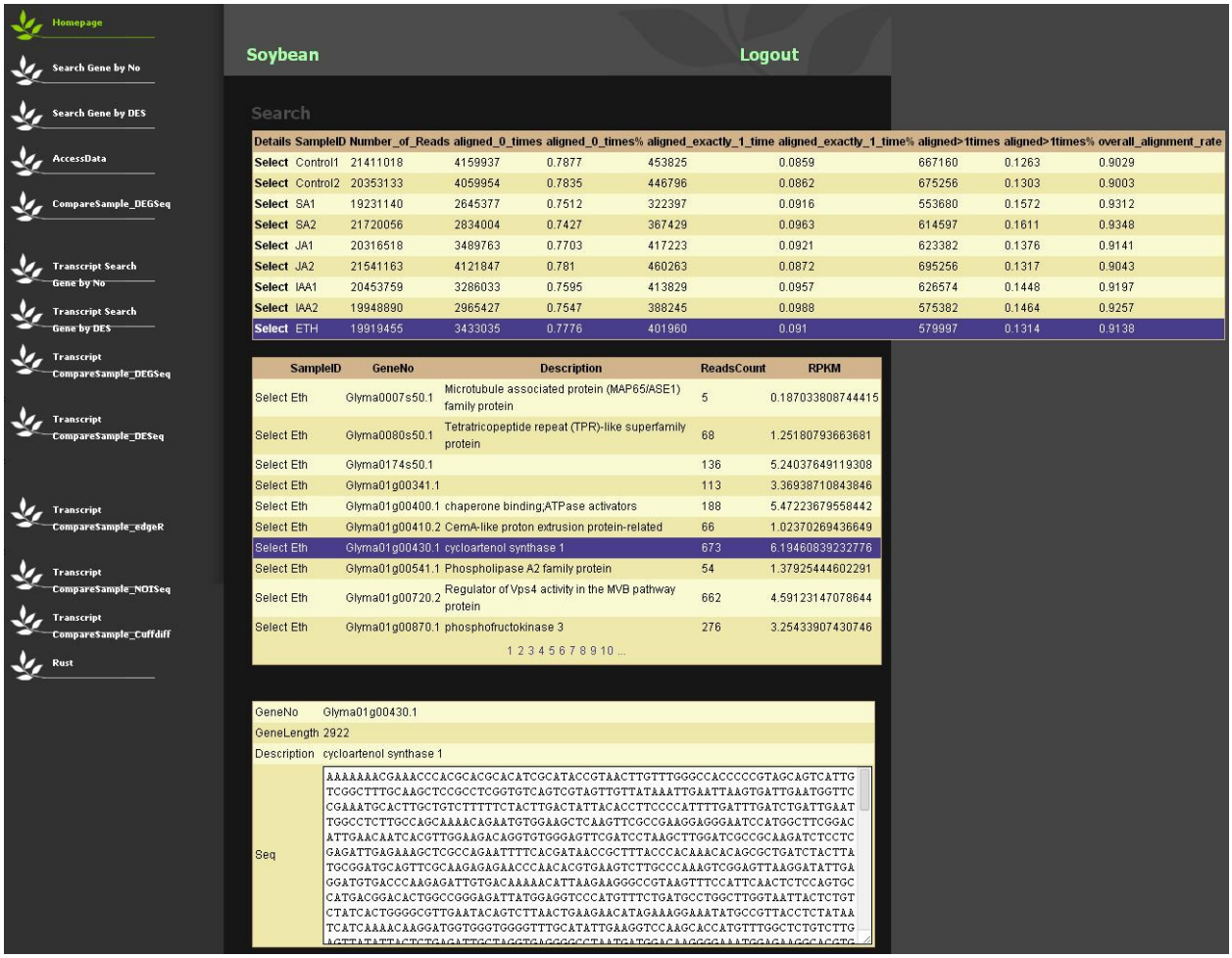


Figure 5.8 Access data (3)

Compared sample DEGSeq needs samples, significance and fold-change to be clarified as inputs. It shows gene No, gene description, number of reads for the experiment, number of reads for the control, RPKM values. Figure 5.9 shows the compared sample result by DEGSeq program.

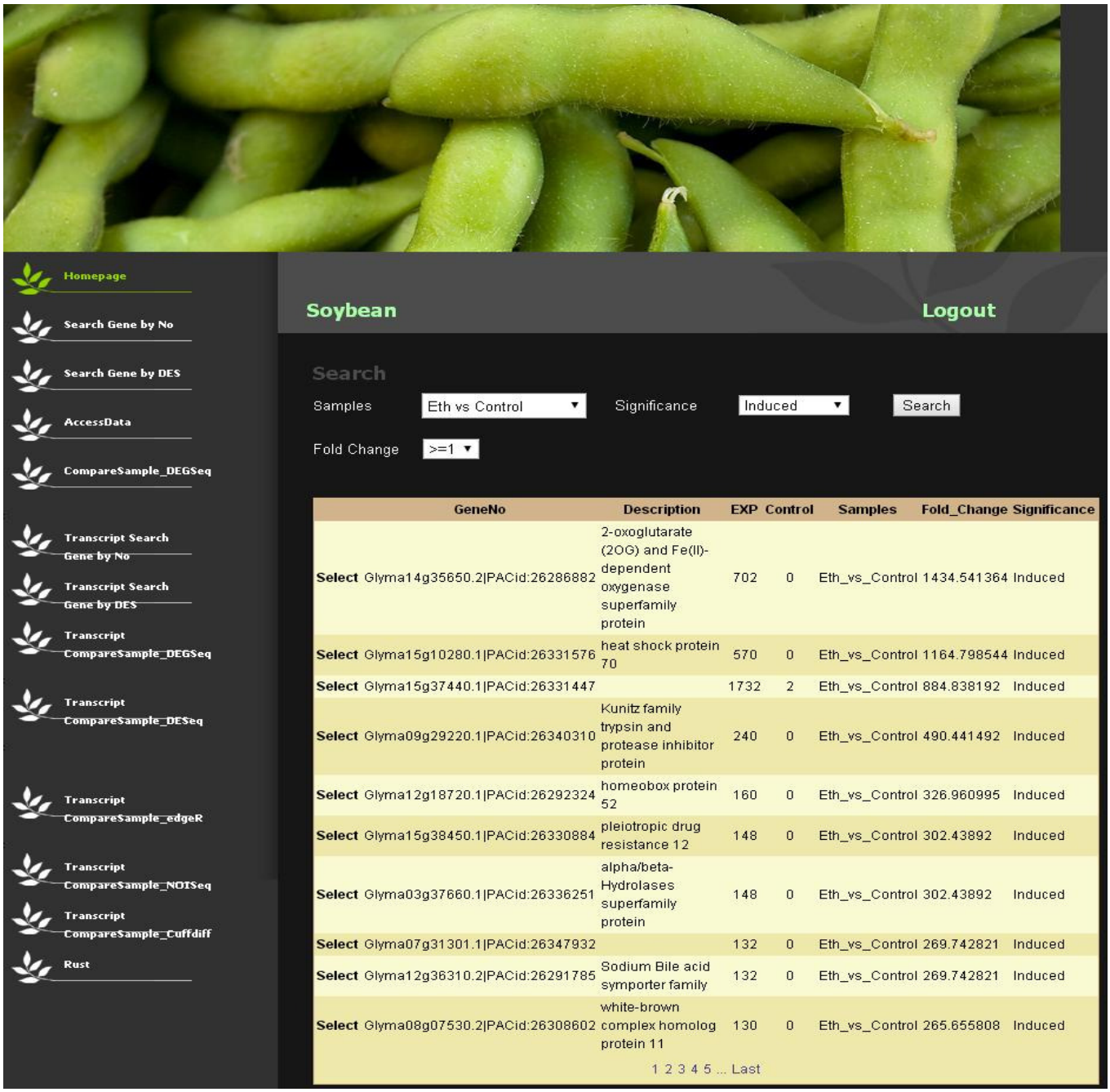


Figure 5.9 DEGSeq

If user is more interested in the details, the select option next to gene number can be selected. As the result, gene No, z-score, p-value, q-value, signature, fold-change, significance of gene will be shown. The Figure 5.10 shows the result.

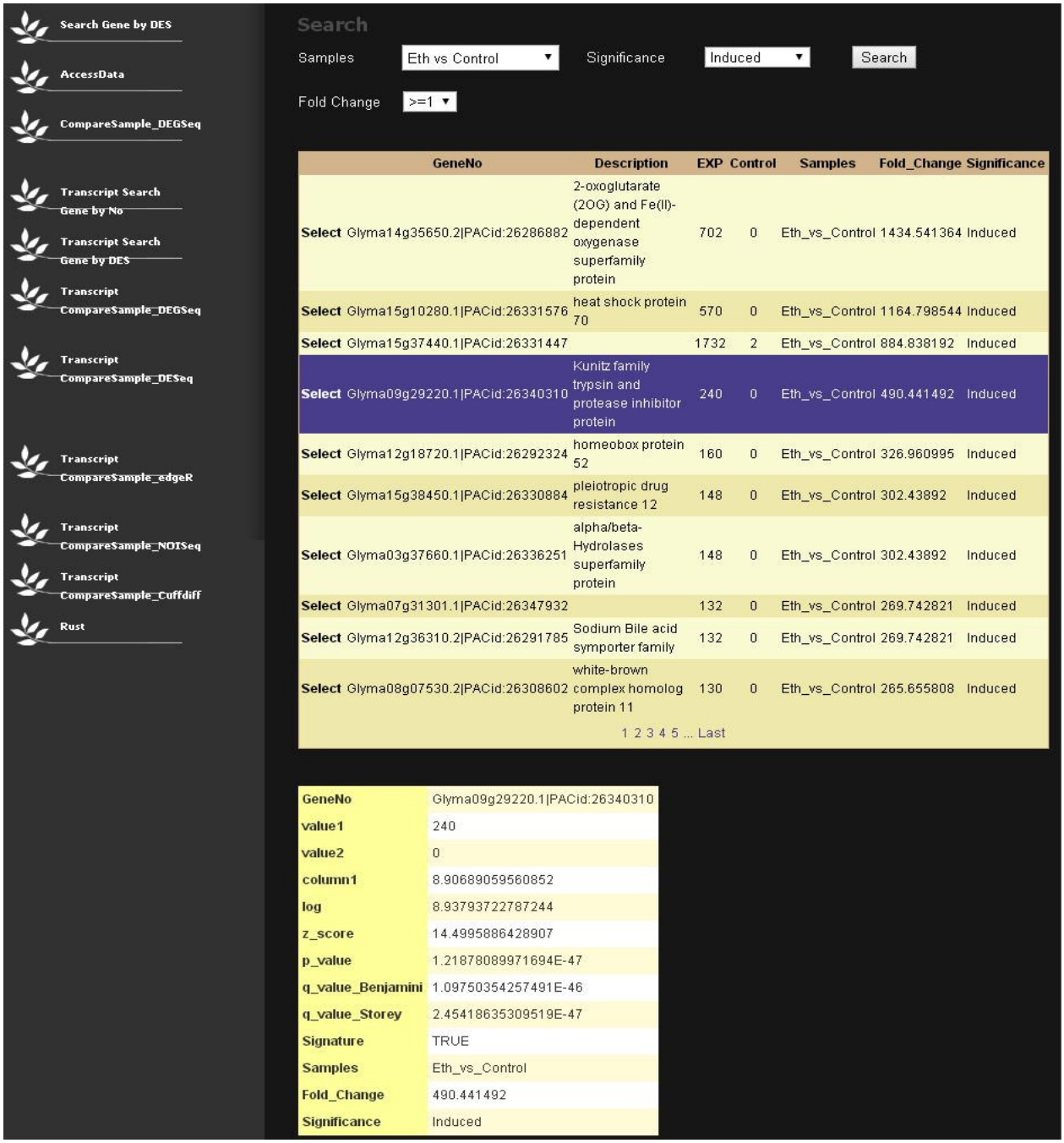


Figure 5.10 DEGSeq(2)

5.5 Conclusion

We analyzed next generation sequence data related to soybean treated with different hormones to elucidate the molecular and biochemical pathways that are altered in soybean when treated with these hormones. Therefore, we developed a relational database to store and manage the huge amount of data generated by these experiments. We also made the database publicly available via the World Wide Web, and created a web based user interface to allow users to query their data. Our data base will be one of the first publicly available web based plant RNA-Seq databases.

Our collaboration with the USDA on this project resulting in a publication:

Tavakolan, M., Alkharouf, NW. and Matthews, B. "SoyProDB: A database for analysis of next generation sequence data related to soybean treated with different hormones". In preparation for BMC plant biology.

Chapter 6: Compare the result of different analysis tools on soybean sequence data

6.1 Introduction

We treated soybean roots independently with one of four phytohormones: Salicylic acid (SA), Jasmonic acid (JA), auxin (IAA), and ethylene (Eth). Transcripts isolated from treated roots were sequenced using RNA-Sequencing. RNA-Seq analysis tools take the aligned reads from two or more conditions and reports genes and transcripts that are differentially expressed. DEGSeq, DESeq, NOISeq, edgeR, and cuffdiff are tools which used to compute differential expression values. Each tool has its own algorithms, inputs and outputs. Therefore, common dataset among all tools are more accurate.

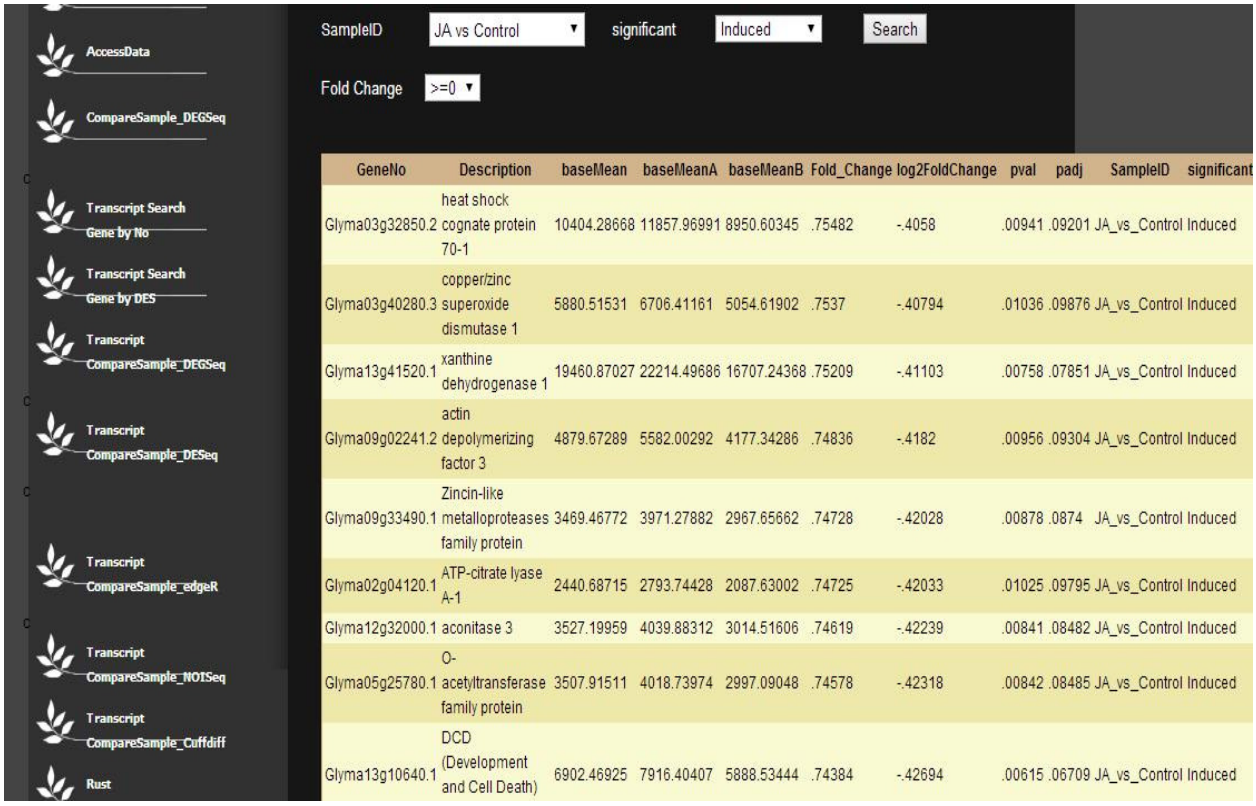
6.2 The result of DESeq analysis tool on soybean sequence data

As input, the DESeq package expects count data. The count data must be raw counts of sequencing reads. Only the actual counts allow assessing the measurement accurately. DESeq will provide nonsensical results if it has any inputs rather than raw counts. Therefore, normalized counts will give useless results. DESeq program must have a replicates raw data.

The input file has five columns. The first one is gene id and the other four columns are raw counts which first two columns are two groups of technical replicates of a sample and other two are replicates of the controls. The best data are useless without metadata in DESeq package. DESeq uses familiar idioms in Bioconductor to manage the metadata that go with the count table. The metadata can be divided into three groups: information about the samples (table columns), about the features (table rows), and about the overall experiment. The metadata modifies whether the samples are treated or untreated and if they are single-end or pair-end.

Since, DESeq package accepts count data as input. It needs normalization process to produce meaningful results. The normalization process is the first step of processing in DESeq. Each column of the count table divides by the length of the gene as a normalized value. The count values are brought to a common scale and comparable by normalization. After normalization, it is straight-forward to look for differentially expressed genes. To compare two conditions, to see if there is differential expression between conditions "nbinomTest" function need to be called. It returns the results with useful information such as baseMean, pvalue. baseMean modifies normalized counts, It is average over all samples from both conditions (treated and untreated conditions). baseMeanA shows normalized counts from first condition which is treated samples. baseMeanB shows normalized counts from second condition which is untreated condition. FoldChange means fold change from both conditions. pval is for the statistical significance of the change. According to fold-change values and baseMean values the genes are categorized to up-regulated or down-regulated groups.

Compared sample DESeq shows GeneNo, gene description, baseMean A, baseMean B, fold-change, pval, sampleID and significant. Figure 6.1 shows the compared sample result by DESeq program.



GeneNo	Description	baseMean	baseMeanA	baseMeanB	Fold_Change	log2FoldChange	pval	padj	SampleID	significant
Glyma03g32850.2	heat shock cognate protein 70-1	10404.28668	11857.96991	8950.60345	.75482	-.4058	.00941	.09201	JA_vs_Control	Induced
Glyma03g40280.3	copper/zinc superoxide dismutase 1	5880.51531	6706.41161	5054.61902	.7537	-.40794	.01036	.09876	JA_vs_Control	Induced
Glyma13g41520.1	xanthine dehydrogenase 1	19460.87027	22214.49686	16707.24368	.75209	-.41103	.00758	.07851	JA_vs_Control	Induced
Glyma09g02241.2	actin depolymerizing factor 3	4879.67289	5582.00292	4177.34286	.74836	-.4182	.00956	.09304	JA_vs_Control	Induced
Glyma09g33490.1	Zincin-like metalloproteases family protein	3469.46772	3971.27882	2967.65662	.74728	-.42028	.00878	.0874	JA_vs_Control	Induced
Glyma02g04120.1	ATP-citrate lyase A-1	2440.68715	2793.74428	2087.63002	.74725	-.42033	.01025	.09795	JA_vs_Control	Induced
Glyma12g32000.1	aconitase 3	3527.19959	4039.88312	3014.51606	.74619	-.42239	.00841	.08482	JA_vs_Control	Induced
Glyma05g25780.1	O-acetyltransferase family protein	3507.91511	4018.73974	2997.09048	.74578	-.42318	.00842	.08485	JA_vs_Control	Induced
Glyma13g10640.1	DCD (Development and Cell Death)	6902.46925	7916.40407	5888.53444	.74384	-.42694	.00615	.06709	JA_vs_Control	Induced

Figure 6.1 The result of DESeq analysis tool on soybean sequence data

6.3 The result of NOISeq analysis tool on soybean sequence data

The other tool which we used in this study is NOISeq. Differential expression analysis maybe carried on both read counts or normalized expression data such as RPKM values, and also any other normalized expression values. However, in our study row counts have been used as inputs to be able to compare the results with other tools. Two pieces of information must be provided to the readData function of NOISeq. The expression data and the factors defining the experimental groups to compared. However, some additional annotations such as gene length need to be provided in order to normalize the count data. The expression data must be provided in a matrix or a data.frame object. Factors are given in a data.frame. The levels of the factor are

“Sample” and “Control”. The order of the samples (columns) in the expression data has to be the same as factors.

The feature length is needed to perform an exploratory analysis before the differential expression computation. Since the length is a matrix, the feature names or IDs must be in the row names of the object. For reading these data and to start working with all the functionalities from NOISeq package, the function readData has to be used.

The NOISeq package computes differential expression between two experimental conditions given the expression level of the considered features. The package takes read counts from RNA-seq as the expression values, and previously normalized read counts. Since, there are some noises in the analysis the normalization step is very important in order to make the samples comparable and to remove possible biases in the data. It is useful to filter out low expression data prior to differential expression analysis. The normalization method which has been used is RPKM method. The expression values are divided by length for RPKM method.

NOISeq calculates the differential expression statistics for different parameters such as M which is the log2-ratio of the two conditions and D which is the value of the difference between conditions. If M and D values are higher than in noise, the gene will be considered as differentially expressed. Noise distribution is comparing all pairs of replicates within the same condition.

Thus, by comparing the M and D values against the noise distribution, NOISeq obtains the probability of differential expression for this feature. Two replicates in one of the experimental conditions are enough to run the algorithm. NOISeq shows GeneNo, gene description, control_mean, sample_mean, M, D, prob and ranking. Figure 6.2 shows the result by NOISeq program.

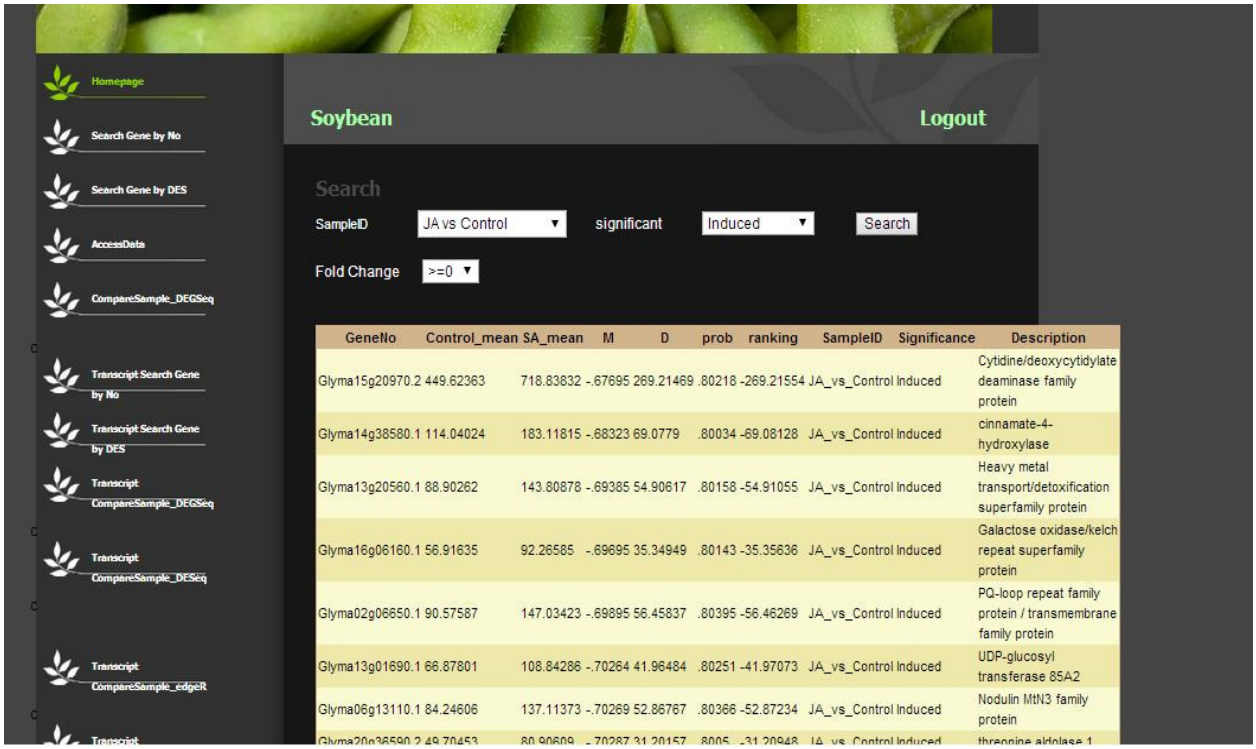


Figure 6.2 Shows the result by NOISeq program

6.4 The result of edgeR analysis tool on soybean sequence data

edgeR works on a table of integer read counts which shows the total number of reads aligning to each gene. Each file is contained gene expression data for a sample. Files at least should have two columns, one for the counts and one for a gene identifier. First column is transcript identifier and the second column is counts. Transcript identifier is unique. The files are assumed to be tab-delimited and to contain column headings. First, the counts have been read into an R session. Because the count data is contained in a single tab-delimited text file with multiple columns, one for each sample, then the simplest method is using read.delim.

edgeR stores data in a simple list-based data object called a DGEList. The function readDGE makes a DGEList object directly. The data.frame samples contain a column group,

identifying the group membership of each sample. edgeR program filters out tags with very low counts. Since there are two replicate samples in each group, it keeps tags that are expressed at a reasonable level in at least two samples. edgeR shows GeneNo, gene description, logFC, false discovery rate (FDR), sampleID, and significant. Figure 6.3 shows the result by edgeR program.

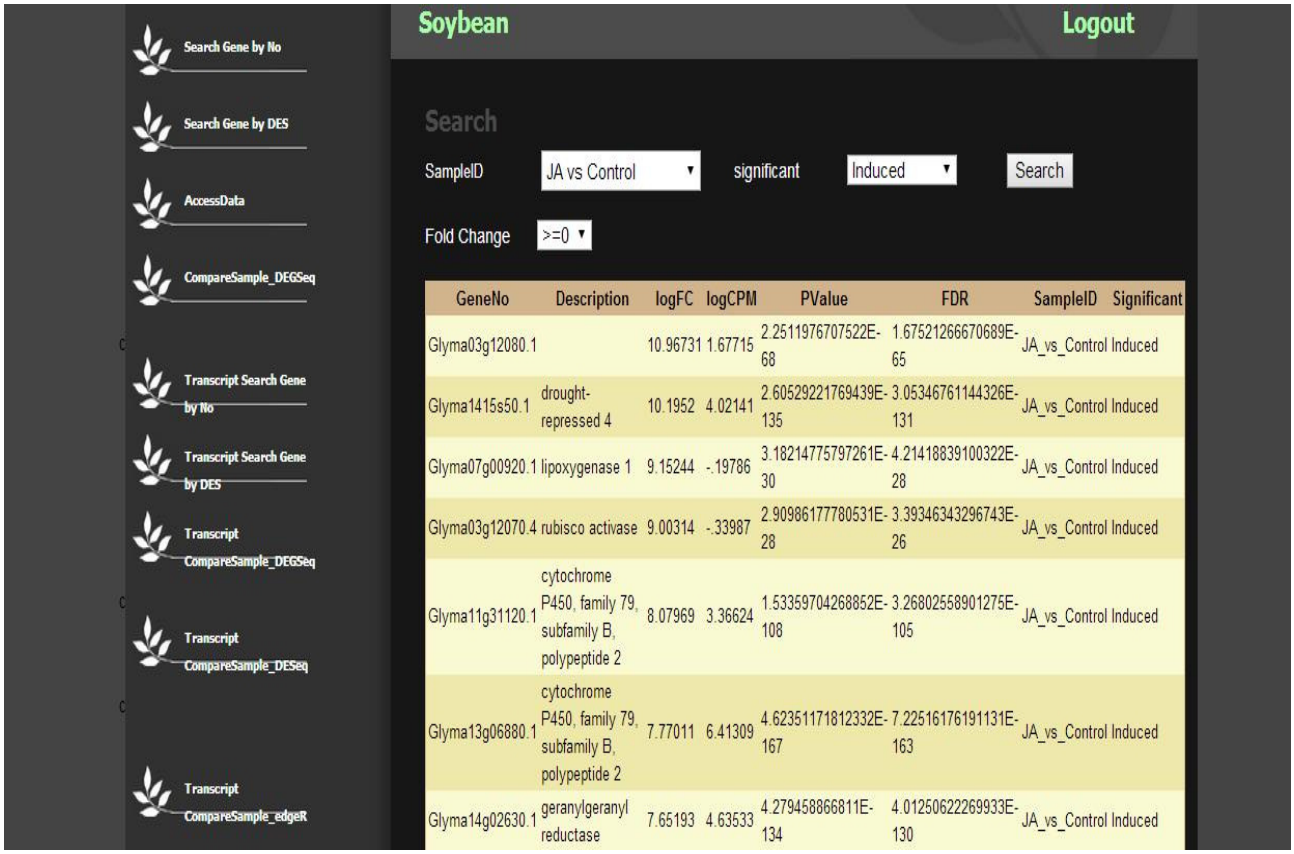


Figure 6.3 Shows the result by edgeR program

6.5 The result of cuffdiff analysis tool on soybean sequence data

Cufflinks runs on Linux or Mac OS. Cufflinks takes a text file of SAM alignments, or a binary SAM (BAM) file as input. SAM is a standard short read alignment which allows aligners to attach custom tags to individual alignments. Cufflinks requires that the alignments you supply

have some of these tags. SAM stands for Sequence Alignment/Map format. It is a TAB-delimited text format consisting of a header section, which is optional, and an alignment section. For faster processing of SAM files, we then converted all SAM formatted files into BAM formatted files. BAM is the binary version of a SAM file. BAM files are generated by compressing SAM file using the BGZF compression algorithm. The UNIX shell commands used to convert the SAM file to BAM is:

samtools faidx index.fasta

samtools import ref_index.fa.fai output_file.sam output_file.bam

samtools sort output_file.bam output_file.sort

samtools index output.sort.bam

Where ref_index.fa.fai is the index file of the reference genome, output_file.sam is the SAM file to be converted and output_file.bam is the converted file name. Cufflinks produces transcript.gtf output file. This transcript.gtf file contains Cufflinks' assembled isoforms. The program cuffcompare compares assembled transcripts to a reference annotation. Cuffcompare takes Cufflinks' GTF output as input and produces an index file. Cuffcompare reports a GTF file containing the "union" of all transfrags in each sample. If a transfrag is present in both samples, it is thus reported once in the combined gtf.Cuffcompare produces the following output files:

- <outprefix>.tracking: This file matches transcripts up between samples. Each row contains a transcript structure that is present in one or more input GTF files. Because the transcripts will generally have different IDs cuffcompare examines the structure of each the transcripts, matching transcripts that agree on the coordinates and order of all of their introns, as well as strand would be saved in this file. Matching transcripts are allowed to

differ on the length of the first and last exons, since these lengths will naturally vary from sample to sample due to the random nature of sequencing.

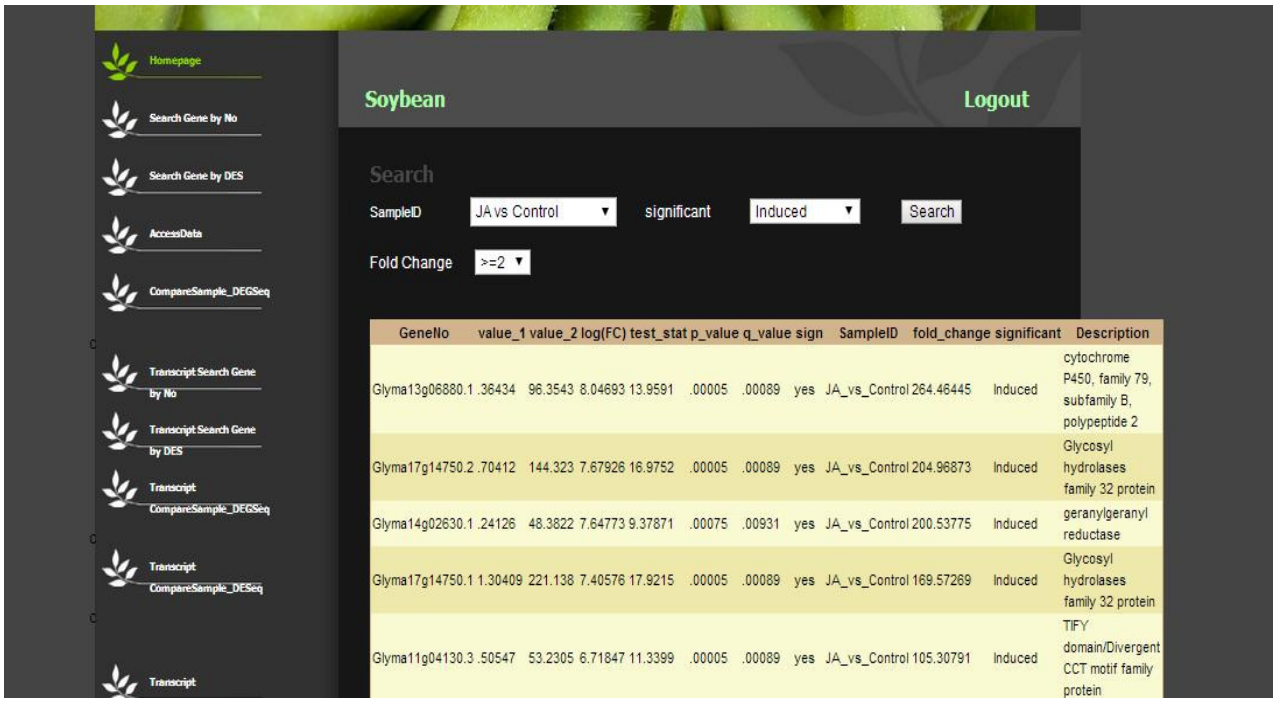
- <cuff_in>.refmap: If cufflinks transcripts either fully or partially match, the tab delimited file lists for each reference transcript would be saved. There is one row per reference transcript, and the columns are reference gene name, reference transcript id, class code and cufflinks matches. Reference gene name is the gene_name attribute of the reference GTF record for transcript. Reference transcript id is the transcript_id attribute of the reference GTF record for this transcript. Class code is the type of match between the Cufflinks transcripts and the reference transcript. One of either 'c' for partial match, or '=' for full match. Cufflinks matches, is a comma separated list of Cufflinks transcript ids matching the reference transcript.
- <cuff_in>.tmap: This tab delimited file lists the most closely matching reference transcript for each Cufflinks transcript. There is one row per Cufflinks transcript, and the columns are reference gene name, reference transcript id, class code, cufflinks gene id, cufflinks transcript id, FPKM. Reference gene name is the gene_name attribute of the reference GTF record for this transcript. Reference transcript id is the transcript_id attribute of the reference GTF record for this transcript. Class code is the type of relationship between the Cufflinks transcripts and the reference transcript. Cufflinks gene id is the Cufflinks internal gene id. Cufflinks transcript id is the Cufflinks internal transcript id. FPKM is the expression of this transcript expressed in FPKM.

Cufflinks includes a program cuffdiff to find significant changes in transcript expression. Cuffdiff takes a GTF file of transcripts as input, along with two or more SAM files containing the fragment alignments for two or more samples. It is not necessary to have the same number of

replicates for each sample. However, we have the same number replicates (two replicates per sample) in our study.

It tracks changes in the relative abundance of transcripts sharing a common transcription start site, and in the relative abundances of the primary transcripts of each gene. It produces a number of output files that contain results for changes in expression at the level of transcripts, primary transcripts, and genes.

Cuffdiff calculates the FPKM of each transcript, primary transcript, and gene in each sample. Primary transcript and gene FPKMs are computed by summing the FPKMs of transcripts in each primary transcript group or gene group. The results are output in FPKM tracking files. Positive fold change determines up-regulated and negative fold change classifies down-regulated gene. Figure 6.4 shows the result by cuffdiff program.



GeneNo	value_1	value_2	log(FC)	test_stat	p_value	q_value	sign	SampleID	fold_change	significant	Description
Glyma13g06880.1	.36434	96.3543	8.04693	13.9591	.00005	.00089	yes	JA_vs_Control	264.46445	Induced	cytochrome P450, family 79, subfamily B, polypeptide 2
Glyma17g14750.2	.70412	144.323	7.67926	16.9752	.00005	.00089	yes	JA_vs_Control	204.96873	Induced	Glycosyl hydrolases family 32 protein
Glyma14g02630.1	.24126	48.3822	7.64773	9.37871	.00075	.00931	yes	JA_vs_Control	200.53775	Induced	geranylgeranyl reductase
Glyma17g14750.1	1.30409	221.138	7.40576	17.9215	.00005	.00089	yes	JA_vs_Control	169.57269	Induced	Glycosyl hydrolases family 32 protein
Glyma11g04130.3	.50547	53.2305	6.71847	11.3399	.00005	.00089	yes	JA_vs_Control	105.30791	Induced	TIFY domain/Divergent CCT motif family protein

Figure 6.4 Shows the result by cuffdiff program

6.6 The comparison result of all five analysis tools on soybean sequence data

DEGSeq, DESeq, NOISeq, edgeR, and cuffdiff are tools which used to process and compare our data set. The common data set among these five tools is the most accurate data. The common differentially expressed genes among all five tools are as follows: 2,997 genes were identified as induced and 2,045 genes as suppressed by Salicylic acid (SA) treatment. 1,507 genes were identified as induced and 1,490 genes as suppressed by Jasmonic acid (JA) treatment. 2,240 genes were identified as induced and 3,302 genes as suppressed by auxin (IAA) treatment. 9,151 genes were identified as induced and 6,315 genes as suppressed by ethylene (Eth) treatment. Functional analysis of differential transcripts revealed several distinct transcriptomic profiles. Table 6.1 shows the result of all five programs.

	SA		JA		IAA		Eth	
	Suppressed	Induced	Suppressed	Induced	Suppressed	Induced	Suppressed	Induced
DEGSeq	23909	10385	21644	6798	16156	15758	24204	10776
DESeq	5219	6220	5059	3934	7380	8297	11824	12339
EdgeR	5731	7805	4460	5243	7906	9158	14373	14033
NOISeq	3600	2347	1998	1717	2398	3480	14724	6551
Cuffdiff	4154	5813	2495	3665	5760	6873	10198	10985
All_5	2997	2045	1507	1490	2240	3302	9151	6315
DEGSeq_vs_DESeq	5142	6220	3015	3934	7331	8297	11737	10371
DEGSeq_vs_EdgeR	5731	7456	4460	4980	7903	9146	14372	10083
Cuffdiff_vs_DEGSeq	4111	5752	2486	3646	5705	6833	10114	9574
Cuffdiff_vs_DESeq	3772	4789	2053	3028	5194	6305	9270	10490
Cuffdiff_vs_edgeR	4025	5743	2475	3639	5660	6816	10066	10894
DESeq_vs_EdgeR	4340	5539	2653	3585	6201	7355	10547	11707
NOISeq_vs_DEGSeq	3600	2347	1998	1717	2398	3480	14724	6551
NOISeq_vs_DESeq	3366	2139	1679	1544	2268	3343	10558	6511
NOISeq_vs_Cuffdiff	3253	2261	1778	1661	2374	3449	9885	6374
NOISeq_vs_EdgeR	3319	2337	1946	1717	2397	3478	12625	6530

Table 6.1 Shows the results of all five programs

6.7 Conclusion

In this study, five tools were used to compute differential expression values. DEGSeq, DESeq, NOISeq, edgeR, and cuffdiff are tools which are used to process and compare these huge data sets. Each tool has its own protocol, method, inputs and outputs. The common data set among these five tools is the most accurate data set. The result of NOISeq tool is more similar to the common result of all five analysis tools. Since, NOISeq has the most conservative algorithm. The package takes read counts from RNA-Seq and normalized read counts based on RPKM method. NOISeq filters out low expression data and noise prior to differential expression analysis. Removing the noises is very important in order to make the samples comparable and accurate.

DESeq, edgeR and cuffdiff tools are more similar to each other. Therefore, they results are mostly close. However, DEGseq tool provides the huge amount of genes as differentially expressed. This tool is not as conservative as other tools. Up-regulated and down-regulated genes have been categorized whenever the signature is “TRUE”. Positive fold change determines up-regulated and negative fold change classifies down-regulated gene.

Chapter 7: Visualizationof Biological Pathways

7.1 Introduction

Analysis results of gene expression data are huge; visualization can make it much traceable for scientists. We used Cytoscape software to map the results of RNA-Seq to pathways. The tools for visualization were specifically designed to map EC accessions onto KEGG pathways.

KEGG stands for Kyoto Encyclopedia of Genes and Genomes. KEGG is a major resource for pathway analysis. It contains genes, genomes, chemical compounds and reaction information. All protein sequences are annotated with KEGG Orthology (KO) IDs. Sequences that are similar and interacting with the same pathway are categorized as the same group [46, 47, 48].

In KEGG pathways each figure represents a different concept. Each has three different shapes: rectangular, small circle and oval. First, rectangular box represents a gene product usually it shows a protein but sometimes an RNA molecule. Second, a small circle represents a compound. Third a large oval represents a link to another pathway map. Various arrows represent the relationships between genes products for molecular interactions. The pathway database consists of various biological processes that are divided into five groups: metabolism, genetic information processing, environmental information processing, cellular processes and human diseases.

The KEGG pathways are based on existing knowledge and are drawn manually. Pathways are not colorized so users are able to connect the pathways to their own data and colorize it by their own values. This visualization helps user to extract the information of

pathways easier. One of the reasons that KEGG pathways are really useful is that by clicking on any of the gene names users are able to see the corresponding KEGG genes entry. The gene entries are hyperlinked to the corresponding enzyme entry pages and annotated by an enzyme commission (EC) number. EC numbers are needed for mapping KEGG pathways to one's own data. EC accessions are derived through BLASTP soybean proteins against UniProt and get the EC accession for the respective UniProt ID [46, 47].

In our case, there are few gene numbers which mapped to EC numbers; however all of them mapped to their corresponding UniProt ID. UniProt is high-quality and freely accessible database of protein sequence and functional information. It has a huge amount of information about the biological function of proteins derived from the researches.

We used Cytoscape to map ourresults to pathways, and it requires Entrez Gene ID. Entrez Gene is the gene-specific database at the NCBI (National Center for Biotechnology Information). Entrez Gene generates unique integers GeneID as identifiers for genes. It tracks those identifiers and uses them to integrate multiple types of information including summary descriptions, accessions of gene-specific, gene product-specific sequences, chromosomal localization, reports of pathways and protein interactions. When a record in Entrez Gene is established, it is assigned a category such as protein coding, rRNA or unknown. The term 'unknown' is used when the category is under review. The category can change without changing the GeneID [47,48].

Converting UniProt ID to Entrez Gene ID can be done by BioMart software. The BioMart provides free software and data services. BioMart is an open source, independent platform database system that provides unified access to distributed data sources. First, we need to go to ensemble plants website to be able to use BioMart tools and choose our data base

(Glycine Max genes). Figure 7.1 shows the database for mapping IDs. Therefore, filter conditions need to be selected. Figure 7.2 shows the filtrations conditions. The result is shown in the new page. Figure 7.3 shows the result of converting UniProt ID to Entrez Gene ID.

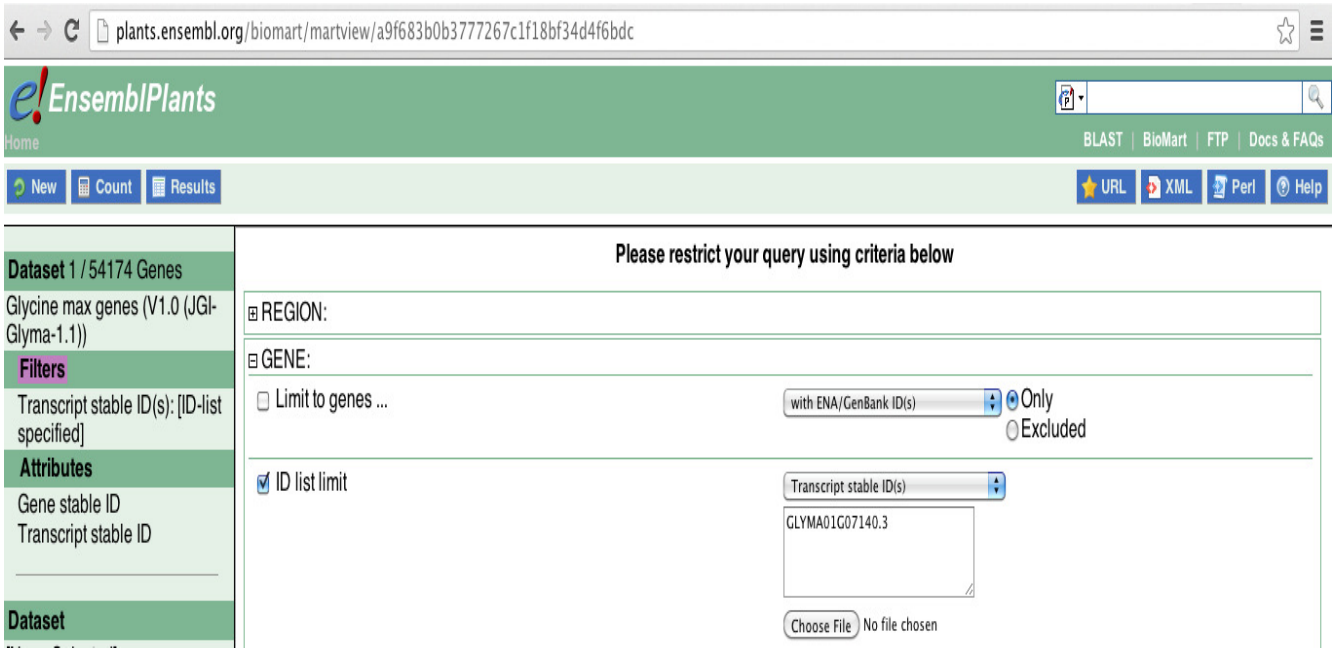


Figure 7.1 Shows the database for mapping IDs

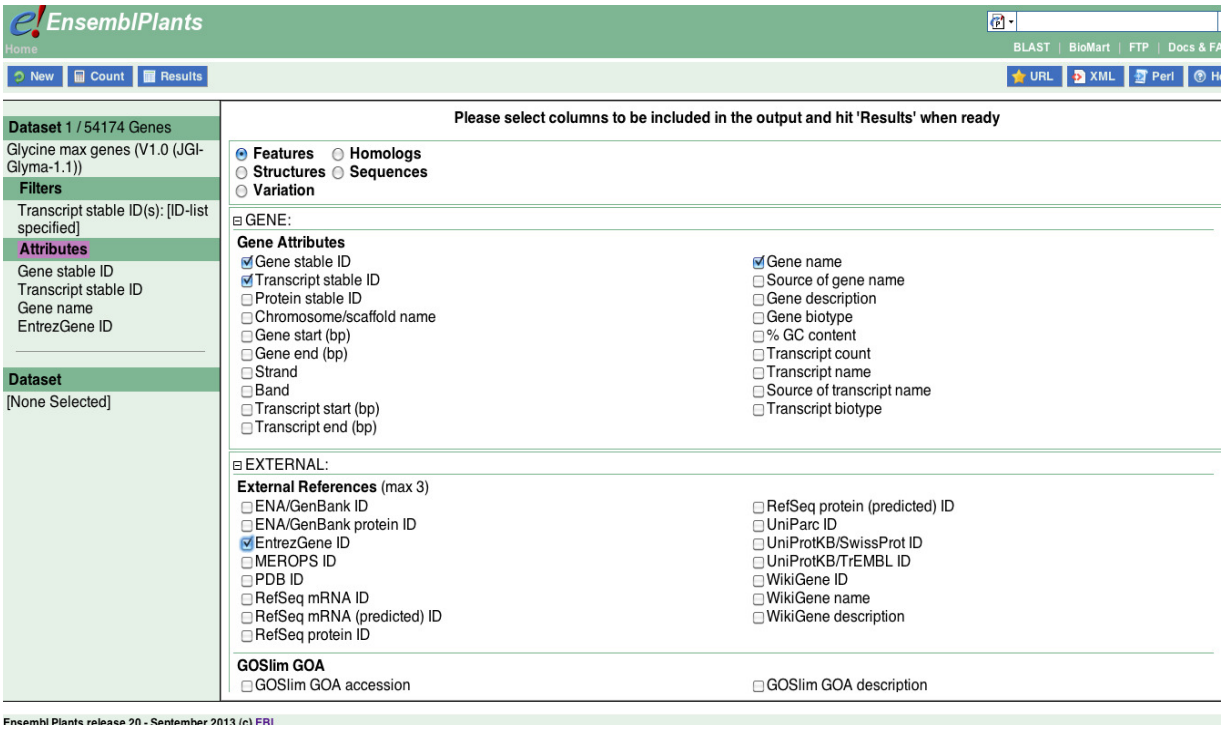


Figure 7.2 Shows the filtrations conditions

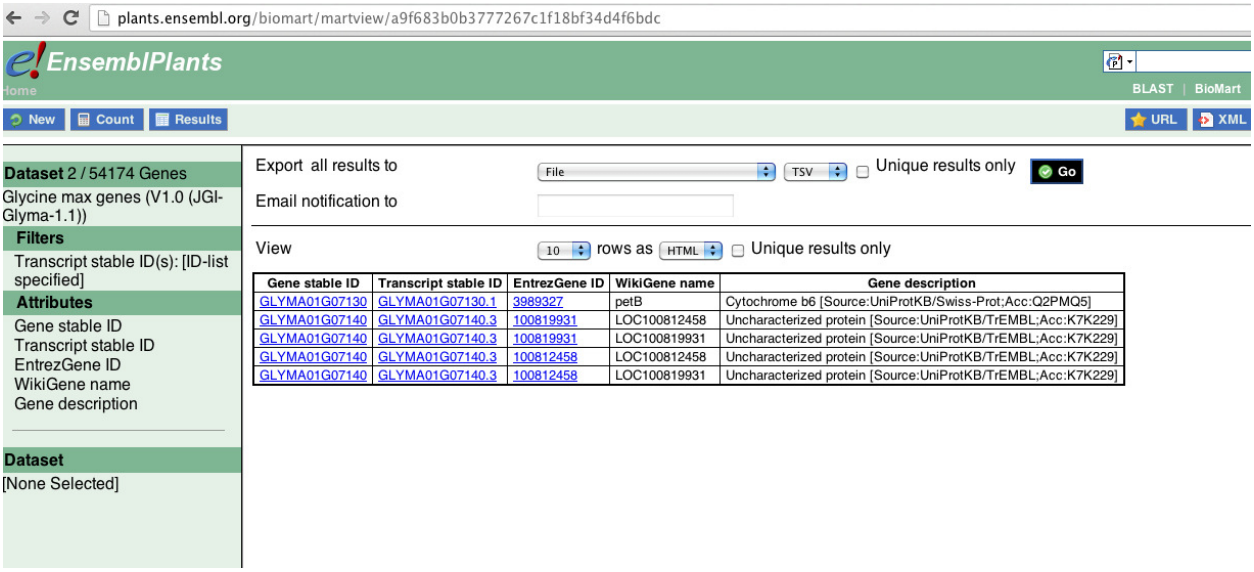


Figure 7.3 Shows the result of converting UniProt ID to Entrez Gene ID

7.2 TCA pathway

One of the fundamental metabolic pathways is the circle acid cycle (TCA). It involves eight enzymes which are necessary for energy production. It begins with the transfer of a two-carbon acetyl group to the four-carbon acceptor compound to form a six-carbon compound. Therefore, a series of chemical transformations causes losing two carboxyl groups as CO2. The colorized TCA pathway based on all four phytohormonesdata makes the changes traceable.

We used Cytoscape software to map the results to TCA pathways for time series transcriptomics analysis. It serves the result as visual pathways for transcriptomics knowledge discovery and functional annotation. Figure 7.4 shows the colorized TCA pathway based on treated soybean roots with SA phytohormones.

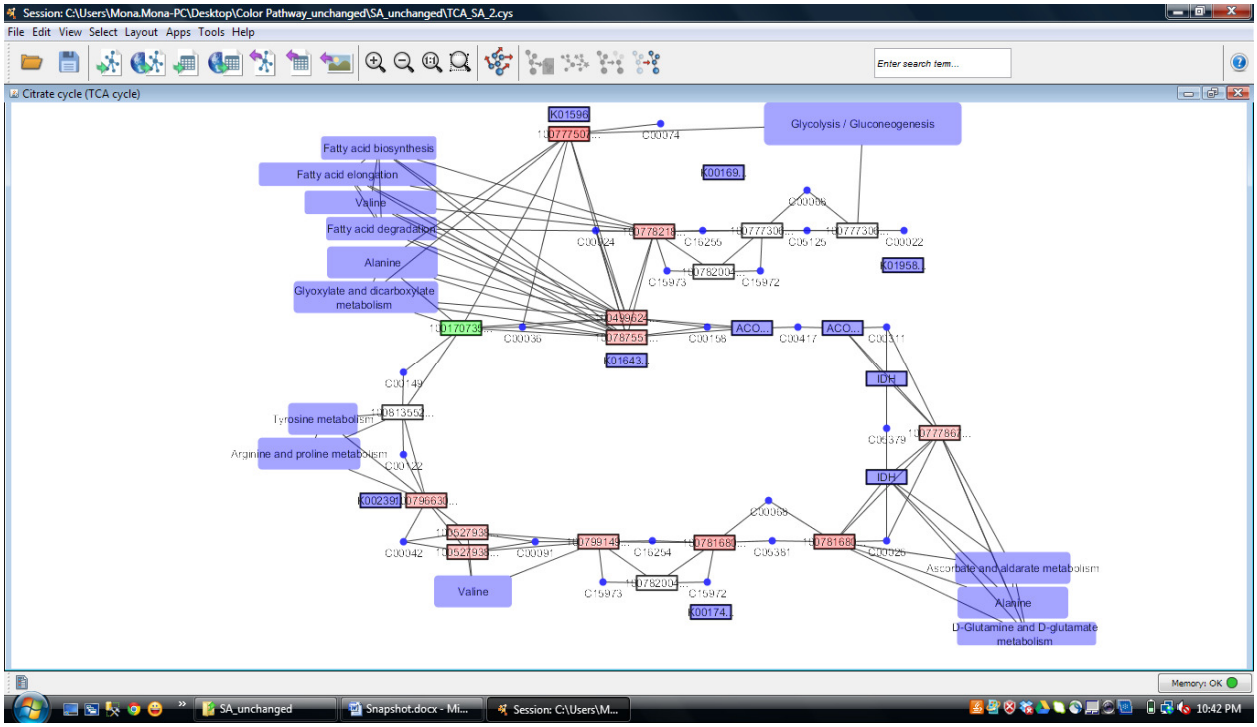


Figure 7.4 Shows the colorized TCA pathway based on treated soybean roots with SA phytohormones

According to importance of TCA in both energy production and biosynthesis, it is crucial for the cell to regulate concentrations of TCA Cycle metabolites in the mitochondria. This pathway is also an important source of biosynthetic building blocks used in gluconeogenesis, amino acid biosynthesis, and fatty acid biosynthesis. Figure 7.5 shows the colorized TCA pathway based on treated soybean roots with JA phytohormones.

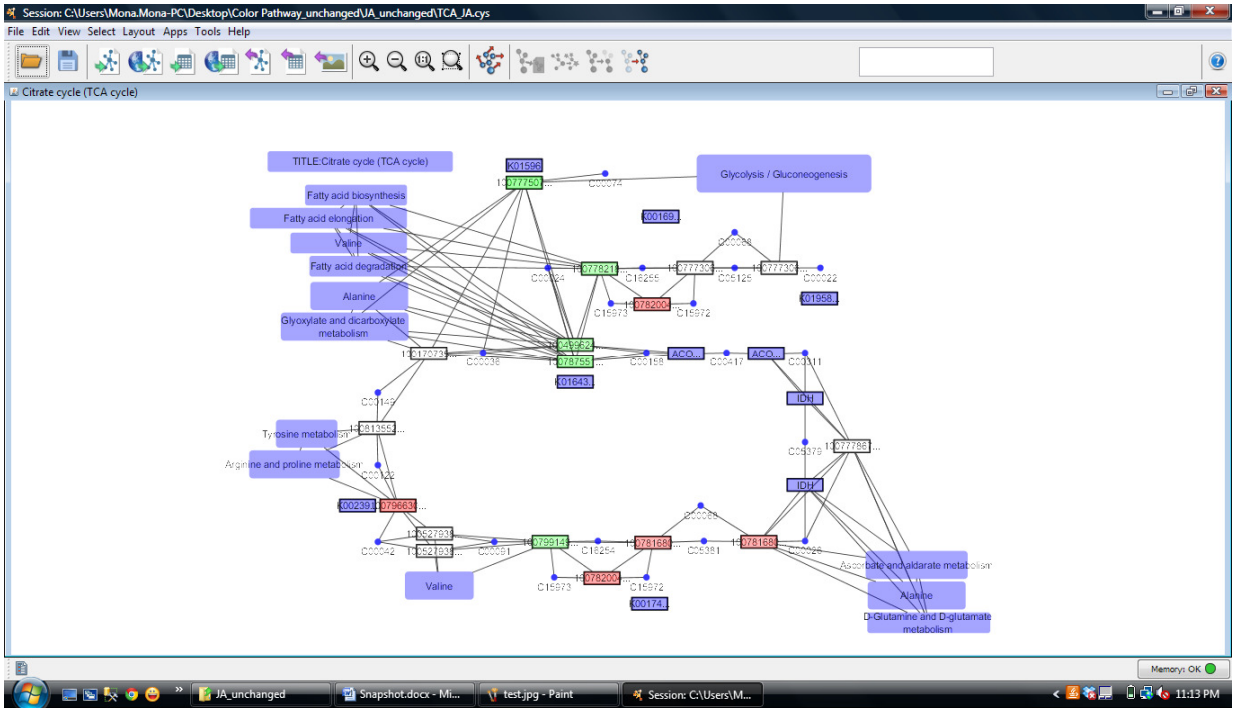


Figure 7.5 Shows the colorized TCA pathway based on treated soybean roots with JA phytohormones

The citrate cycle is an important aerobic pathway for the final steps of the oxidation of carbohydrates and fatty acids. The cycle starts with acetyl-CoA, the activated form of acetate, derived from glycolysis and pyruvate oxidation for carbohydrates and from beta oxidation of fatty acids. The pathway also supplies important precursor metabolites including 2-oxoglutarate.

Figure 7.6 shows the colorized TCA pathway based on treated soybean roots with IAA phytohormones.

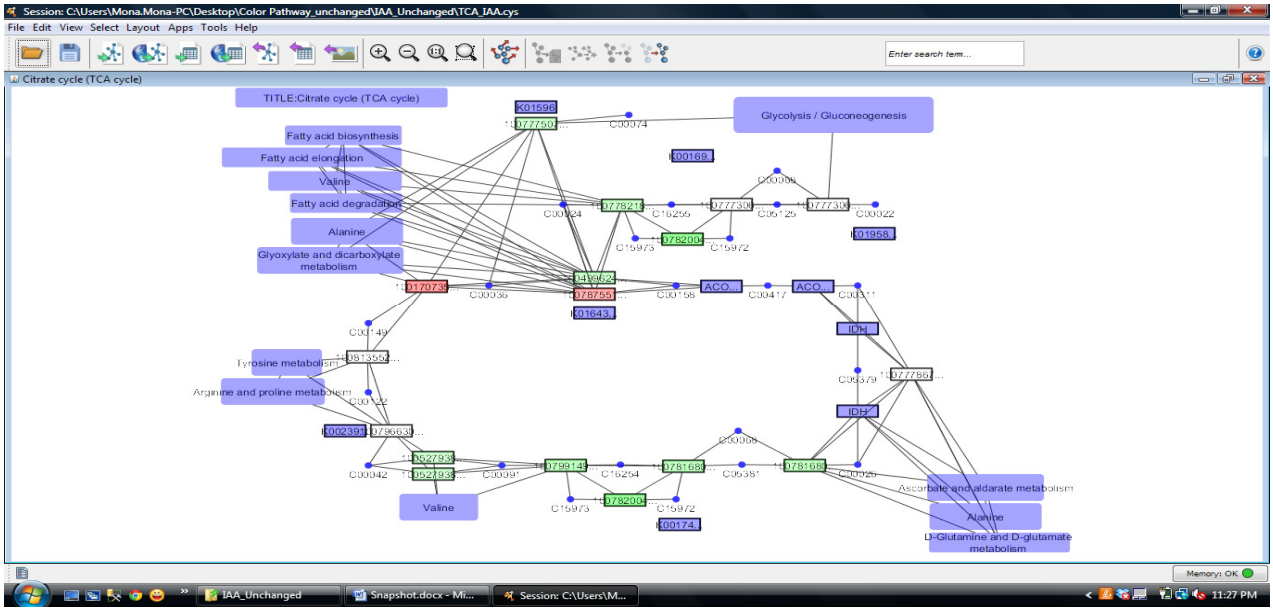


Figure 7.6 Shows the colorized TCA pathway based on treated soybean roots with IAA phytohormones

Figure 7.7 shows the colorized TCA pathway based on treated soybean roots with Eth phytohormones. Visualization of TCA pathway base on all four phytohormones makes the comparison process easy and traceable.

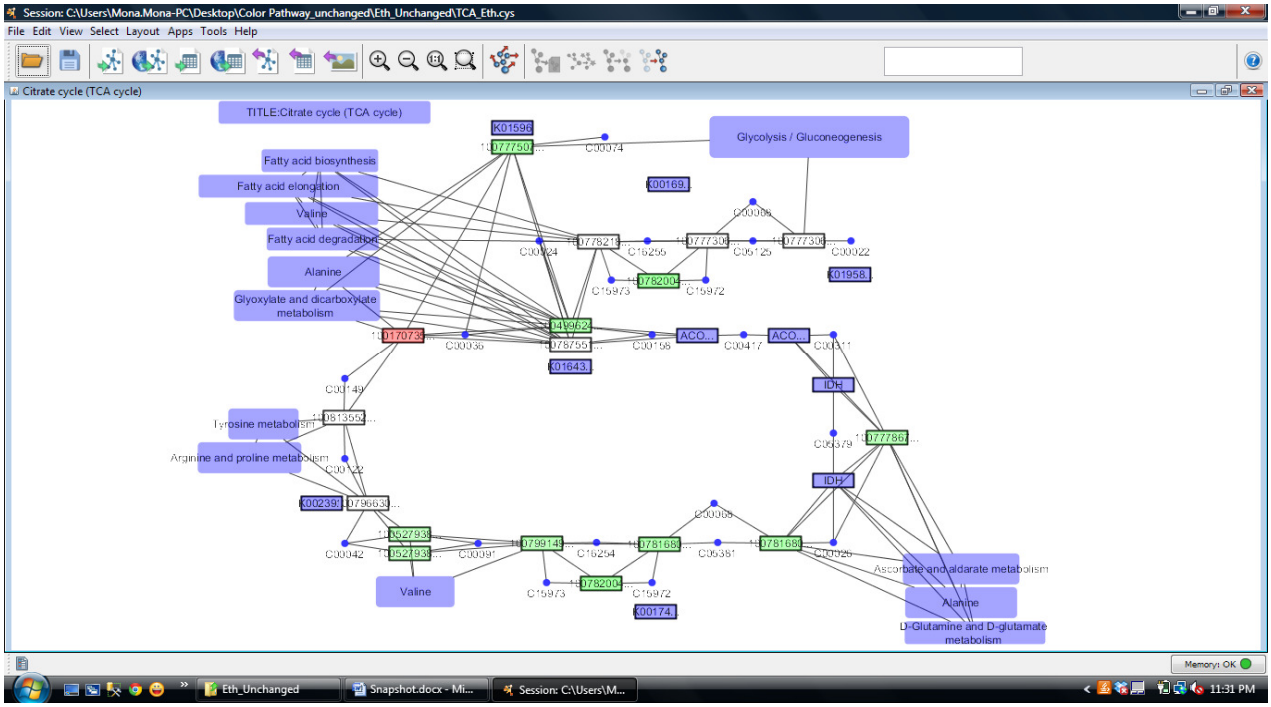


Figure 7.7 Shows the colorized TCA pathway based on treated soybean roots with Eth phytohormones

7.3 Glycolysis/Gluconeogenesis pathway

Glycolysis is the process of converting glucose into pyruvate and generating small amounts of energy. It is a fundamental pathway that produces important precursor metabolites. The reaction steps of three-carbon compounds from glycerone-P to pyruvate form a conserved core module, which is found in almost all organisms, when the enzyme genes of this pathway are examined in completely sequenced genomes. It is essentially a reversal of glycolysis with minor variations of alternative paths. Figure 7.8 shows the colorized Glycolysis/Gluconeogenesis pathway treated with SA phytohormones.

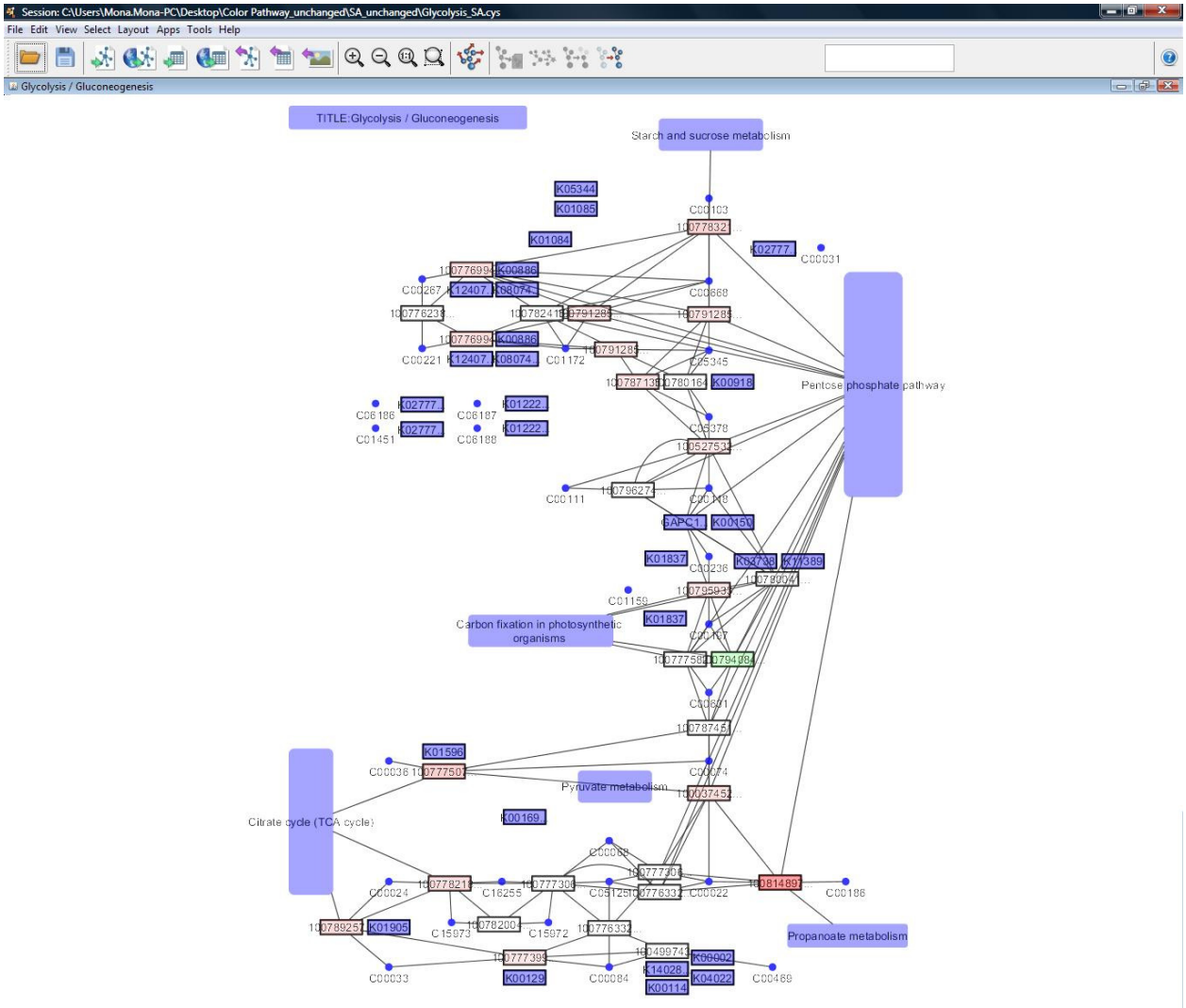


Figure 7.8 Shows the colorized Glycolysis/Gluconeogenesis pathway treated with SA phytohormones

Although glycolysis and gluconeogenesis have some of the same enzymes in common, the two pathways are not simply the reverse of each other. In fact, the highly exergonic, irreversible steps of glycolysis are bypassed in gluconeogenesis. Both pathways are controlled by intercellular and intracellular signals, and they are reciprocally regulated so that glycolysis and gluconeogenesis do not take place simultaneously in the same cell to a significant extent. Glycolysis is the sequence of reactions that metabolizes one molecule of glucose to two

molecules of pyruvate with the concomitant net production of two molecules. Glucose can be synthesized from noncarbohydrate precursors, such as pyruvate and lactic acid, in the process of gluconeogenesis. Figure 7.9 shows the colorized Glycolysis/Gluconeogenesis pathway treated with JA phytohormones.

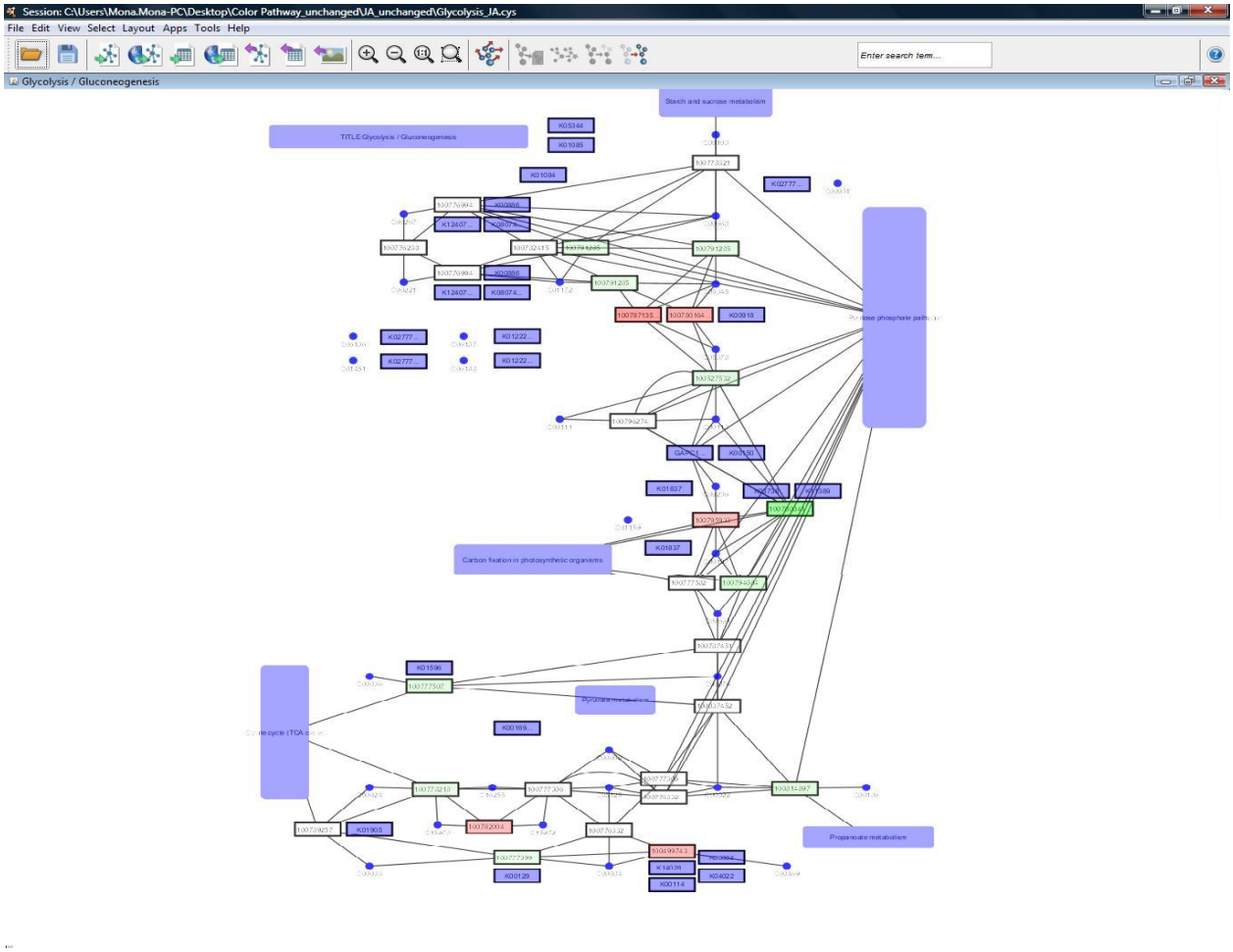


Figure 7.9 Shows the colorized Glycolysis/Gluconeogenesis pathway treated with JA phytohormones

Glucose is an important and common fuel. Glucose is one of the monosaccharides formed from formaldehyde under prebiotic conditions, so it may have been available as a fuel source for

primitive biochemical systems. Glucose has a low tendency, relative to other monosaccharides, to nonenzymatically glycosylate proteins. Glucose has a strong tendency to exist in the ring formation and, consequently, relatively little tendency to modify proteins. Figure 7.10 shows the colorized Glycolysis/Gluconeogenesis pathway treated with IAA phytohormones.

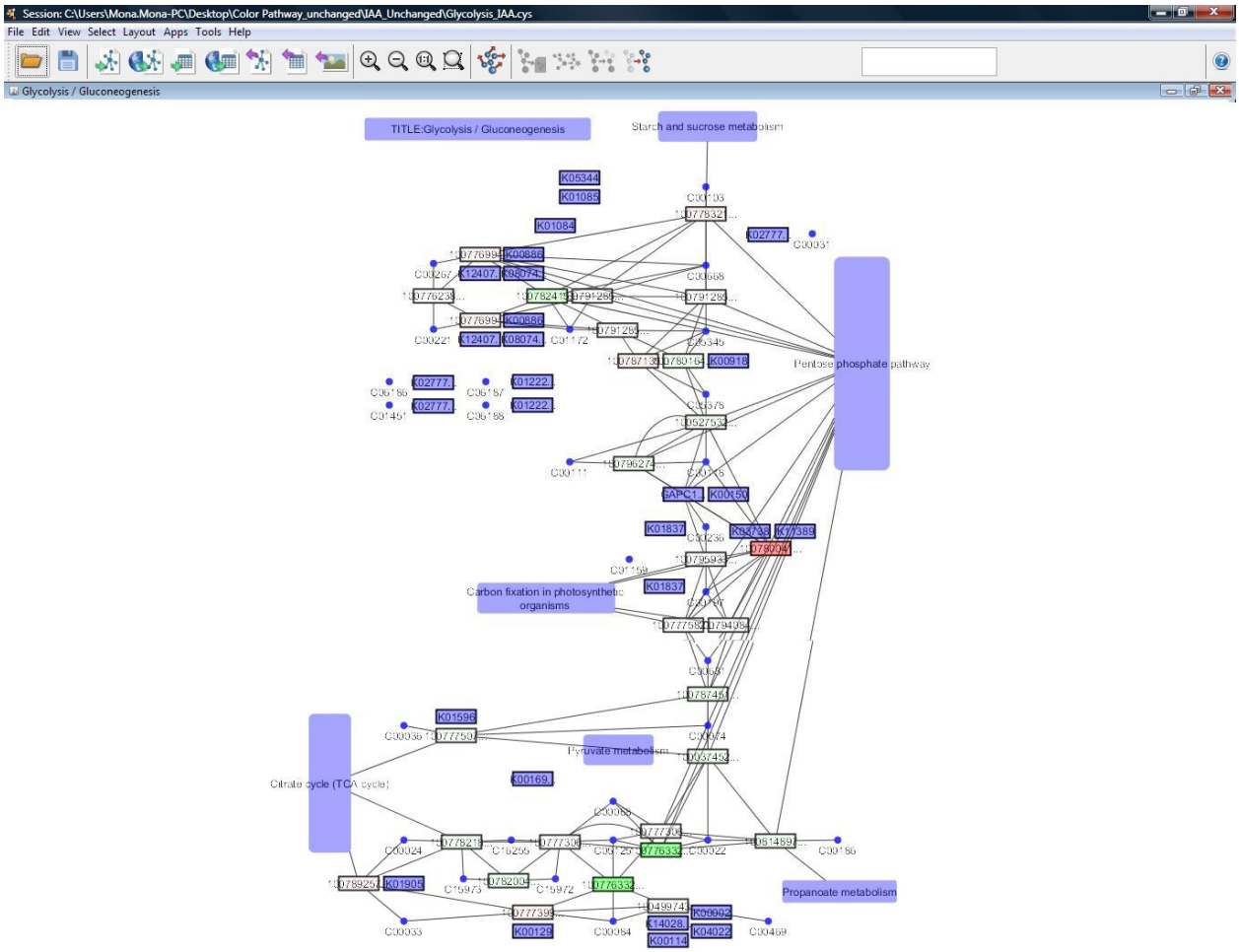


Figure 7.10 Shows the colorized Glycolysis/Gluconeogenesis pathway treated with IAA phytohormones

Gluconeogenesis is the process of synthesizing glucose from non-carbohydrate sources. The starting point of gluconeogenesis is pyruvic acid, although oxaloacetic acid and dihydroxyacetone phosphate also provide entry points. Lactic acid, some amino acids from protein and glycerol from fat can be converted into glucose. Figure 7.11 shows the colorized Glycolysis/Gluconeogenesis pathway treated with Eth phytohormones.

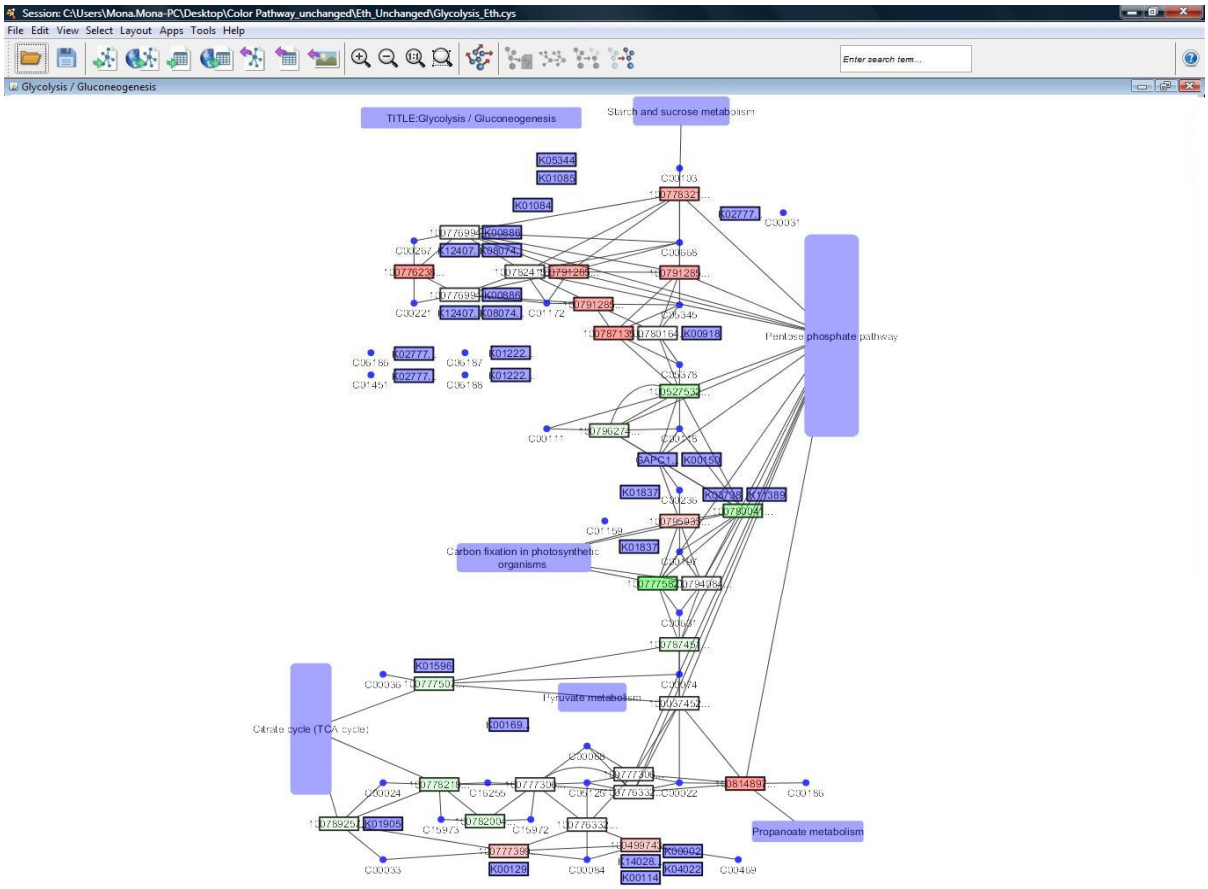


Figure 7.11 Shows the colorized Glycolysis/Gluconeogenesis pathway treated with Eth phytohormones

7.4 Fatty acid biosynthesis

Fatty acid (FA) is a central, conserved process by which acyl chains are produced for utilization in a number of end-products. Central to FA synthesis, the acyl carrier protein represents the cofactor protein that covalently binds all fatty acyl intermediates via a phosphopantetheine linker during the synthesis process. It has two categories. First, it is categorized by being composed of large multifunctional polypeptides. Second, it is categorized consisting of discretely expressed mono-functional proteins. Based on this difference in architecture, the FA system is a good target for the discovery of expression genes. Figure 7.12 shows the colorized Fatty acid biosynthesis pathway treated with SA phytohormones.

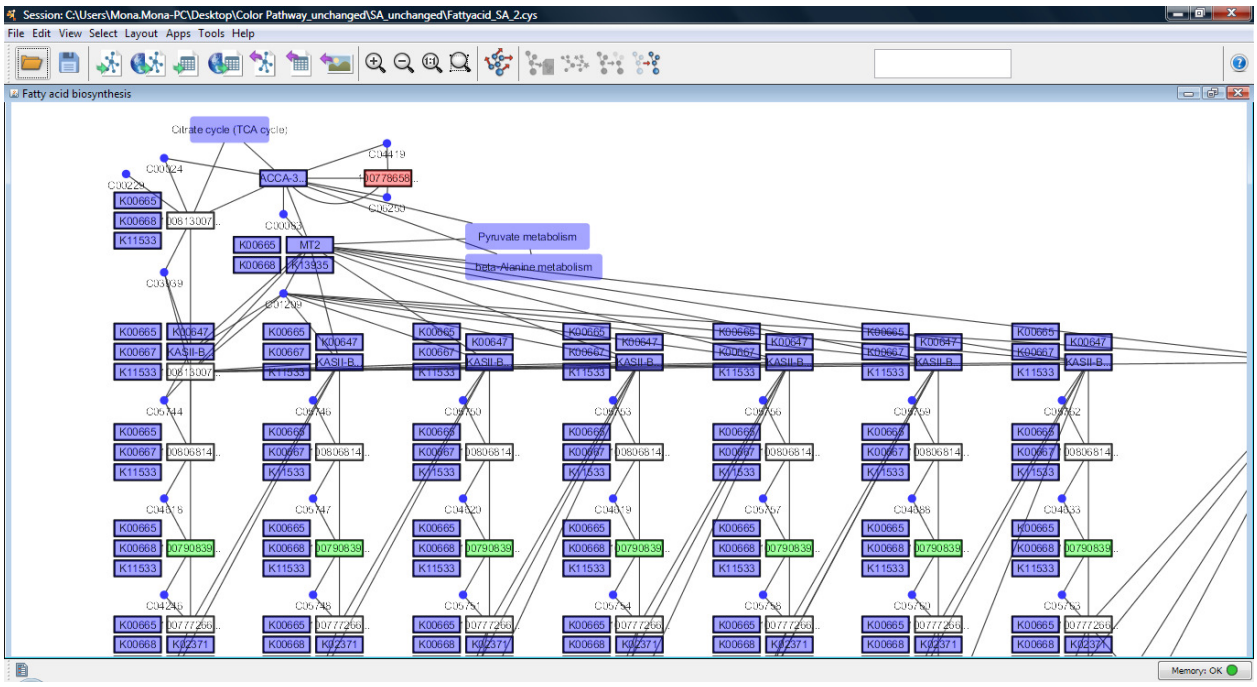


Figure 7.12 Shows the colorized Fatty acid biosynthesis pathway treated with SA phytohormones

FA is a large multienzyme complex. The biosynthetic reaction pathway to a compound is usually not a simple opposite of its breakdown. In fatty acid synthesis, acetylCoA is the direct

precursor only of the methyl end of the growing fatty acid chain. All the other carbons come from the acetyl group of acetylCoA.MalonylCoA contains a 3-carbon dicarboxylic acid, malonate; bound to Coenzyme A. Malonate is formed from acetylCoA by the addition of CO 2 using the biotin cofactor of the enzyme acetylCoA carboxylase. Formation of malonylCoA is the commitment step for fatty acid synthesis, because malonylCoA has no metabolic role other than serving as a precursor to fatty acids. Figure 7.13 shows the colorized Fatty acid biosynthesis pathway treated with JA phytohormones.

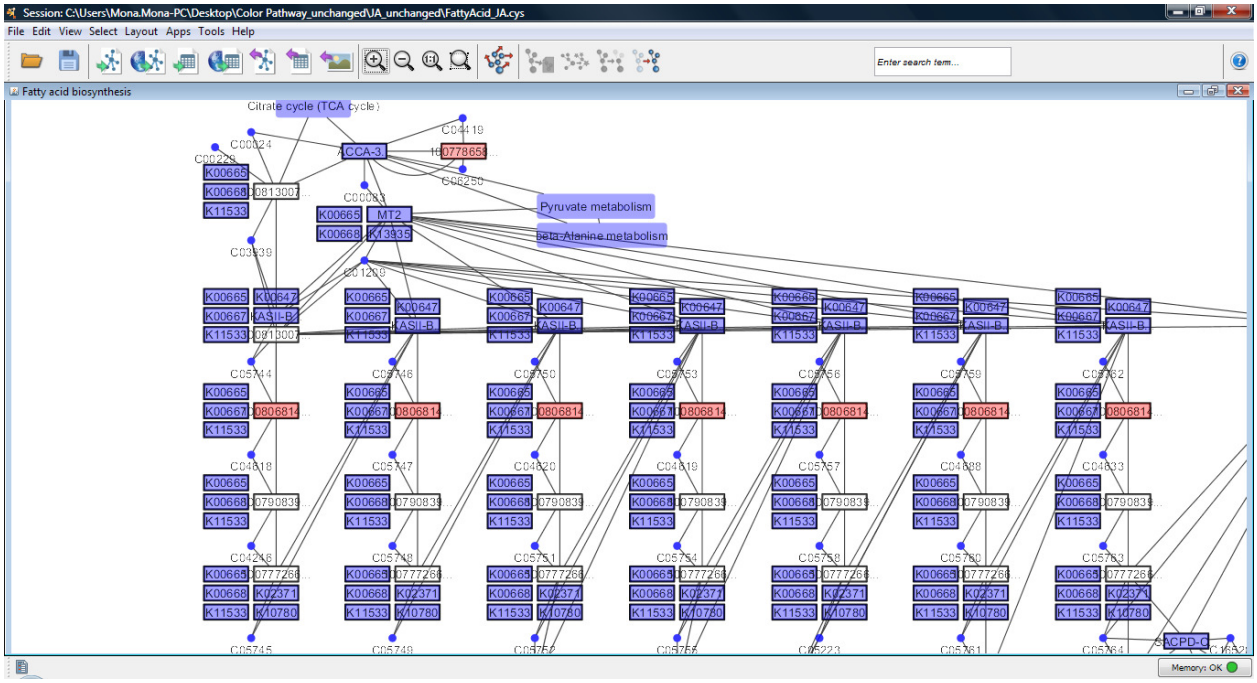


Figure 7.13 Shows the colorized Fatty acid biosynthesis pathway treated with JA phytohormones

Fatty acid synthesis consist of a new set of reactions, Fatty acid synthesis is not simply a reversal of the degradative pathway. For example, the enzymes of fatty acid synthesis in higher organisms are joined in a single polypeptide chain called fatty acid synthase. However, the

degradative enzymes do not seem to be associated. Synthesis takes place in the cytosol, in contrast with degradation, which takes place primarily in the mitochondrial matrix. Figure 7.14 shows the colorized Fatty acid biosynthesis pathway treated with IAA phytohormones.

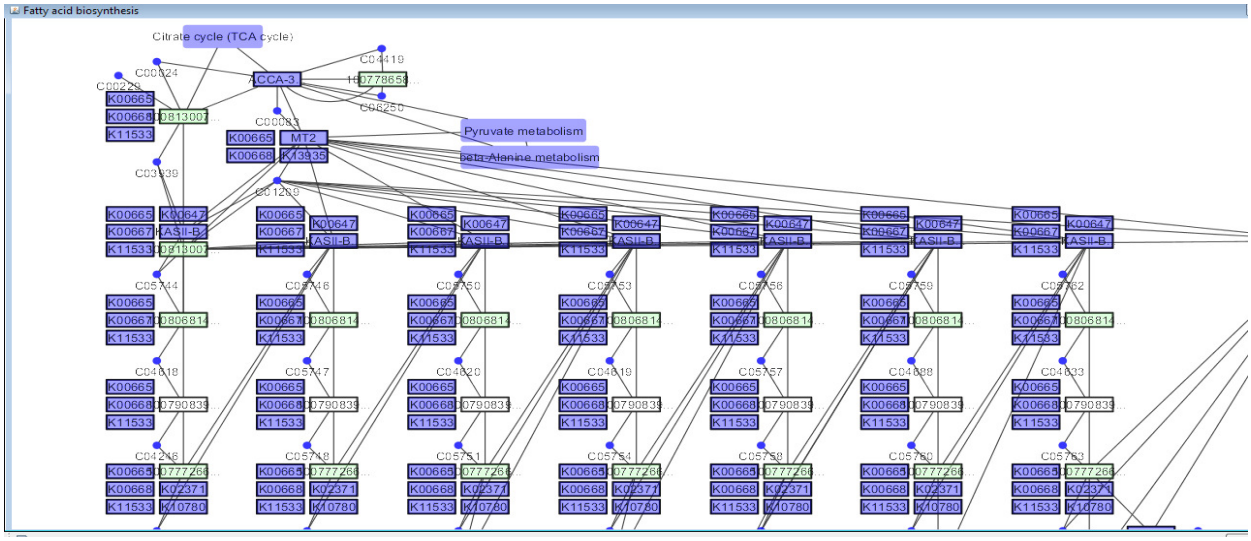


Figure 7.14 Shows the colorized Fatty acid biosynthesis pathway treated with IAA phytohormones

The major unsaturated fatty acids in soybean oil triglycerides are the poly-unsaturates, alpha-linolenic acid, linoleic acid and the mono-unsaturate, oleic acid. Figure 7.15 shows the colorized Fatty acid biosynthesis pathway treated with Eth phytohormones.

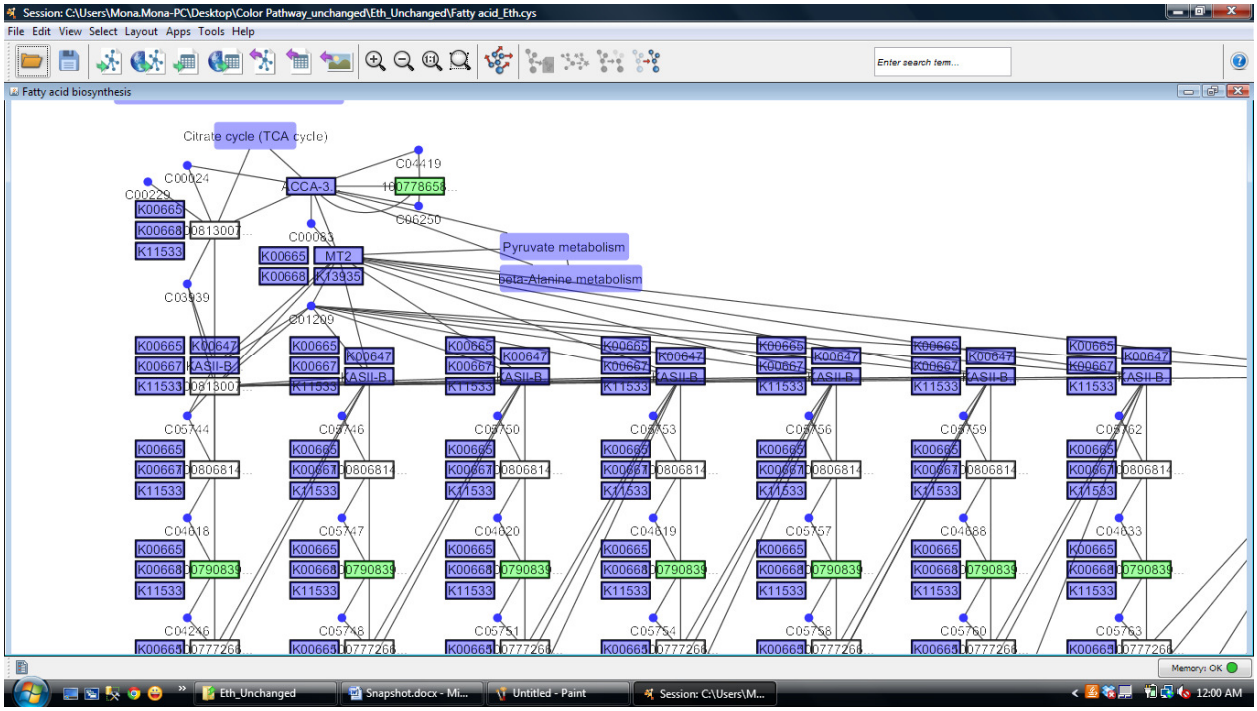


Figure 7.15 Shows the colorized Fatty acid biosynthesis pathway treated with Eth phytohormones

7.5 Conclusion

Gene expression experiments (such as RNA-Seq) have experienced increases in data throughput and the results are always huge. Processing and comparing these huge data sets is difficult and time consuming, however visualization can make it much easier. Pathway visualization means to map quantified transcript expression onto biochemical pathways and makes these huge datasets traceable.

We used Cytoscape software to map the results to pathways for time series and single time point transcriptomics analysis. It can show lightly expressed, moderately expressed, and heavily expressed by different colors whereby green and red represent induced and suppressed respectively. Users are able to pass a user-defined fold-change cutoff, and specify the colors. It will be mapped to the whole pathway automatically.

The Cytoscape tool uses Entrez Gene ID to colorize pathways based on expression values. Genes are mapped to pathways, and the range of data values levels in a given experimental dataset is mapped to colors. Therefore, they are colored according to the corresponding data value. These colorized pathways enable the user to see instantly which pathways are active or inactive under some set of experimental conditions. It serves the result as visual pathways for transcriptomics knowledge discovery and functional annotation.We mapped a number of data sets that we obtained from our collaborators at the USDA, and we are working on a publication.

Chapter 8: Conclusions and Future Perspectives

In this dissertation, we developed methods to identify and analyze soybean proteins and their functions. We also created the necessary databases and web interfaces to make our data and analysis accessible to all researchers around the world.

The results have been used to uncover the composition of storage, allergen and anti-nutritional proteins in cultivated soybean using a combined proteomics approach. We developed a number of databases for 2D gel analysis; they include the identification of several proteins and protein information which is important for developing SCN resistant varieties.

In the second part of the dissertation, we analyzed next generation sequence data related to soybean treated with four different hormones. Salicylic acid (SA), Jasmonic acid (JA), auxin (IAA), and ethylene (Eth) are the four hormones that were used to treat soybean. Statistical analysis of soybean transcription factor binding sites identified differentially expressed genes.

Functional analysis of differential transcripts revealed numerous distinct transcriptomic profiles. Our results provide a potentially novel catalog of over-expressed and under-expressed genes relevant to SA, JA, IAA, and Eth. Therefore, RNA-Seq analysis is stored in SDEG database to be publicly available. SDEG stores are relevant data relating to every RNA_Seq experiment.

In the third part of the dissertation, five tools have been used to compute differential expression values. DEGSeq, DESeq, NOISeq, edgeR, and cuffdiff are tools which we used to process and compare these huge data sets. Each tool has its own protocol, method, inputs and outputs. Therefore, the common set between all tools are the most conservative and accurate data set. In the fourth part of the dissertation, we collected the common result between five tools and

visualize it to make it easy for biologists. Pathway visualization maps quantified transcript expression onto biochemical pathways. We used Cytoscape software to serve the result as visual pathways for transcriptomics knowledge discovery and functional annotation. To date, this dissertation work has resulted in 4 peer reviewed publications, with one of them still in preparation.

In the future, we can analyze transcript expression of other datasets. We can then map them to pathways and visualize them given the pipelines that have already been developed. Processing and comparing the new datasets can help scientists in transcriptomics knowledge discovery and functional annotation. Our database collection also can be extended with new datasets. In addition, we can create tools that facilitate generating new pathways that could be mapped to differential expression analysis result and automatically colorized.

Chapter 9: Appendix

Search gene by No:

Search gene by No provides users to search genes by their numbers. It shows the gene number, the description for the gene and the RPKM values in all the samples

```
<%@ Page Language="C#" MasterPageFile="MasterPage.master" AutoEventWireup="true" CodeFile="Transcript_SearchGeneByNo.aspx.cs"
Inherits="Test" Title="Untitled Page" %>

<asp:Content ID="Content1" ContentPlaceHolderID="head" Runat="Server">

</asp:Content>

<asp:Content ID="Content2" ContentPlaceHolderID="ContentPlaceHolder1" Runat="Server">

<asp:Label ID="Label1" runat="server" Text="Gene NO" BorderColor="White"

   ForeColor="White"></asp:Label>

<asp:TextBox ID="TextBox1" runat="server"></asp:TextBox>

<asp:Button ID="Button1" runat="server" style="margin-left: 37px" Text="Search"

   Width="99px" />

<asp:SqlDataSource ID="SqlDataSource1" runat="server"

      ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"


   SelectCommand="SELECT SUBSTRING(Transcript_Final_RPKM.GeneNo,1,15) as GeneNo,GeneDescription.Description ,
Sample1,Sample2, Sample3, Sample4, Sample5, Sample6, Sample7, Sample8, Sample9 FROM [Transcript_Final_RPKM] INNER JOIN
GeneDescription ON SUBSTRING(Transcript_Final_RPKM.GeneNo, 1, 15) = GeneDescription.GeneNo WHERE
(Transcript_Final_RPKM.GeneNo LIKE '%' + @GeneNo + '%')">

<SelectParameters>

<asp:ControlParameter ControlID="TextBox1" Name="GeneNo" PropertyName="Text"

            Type="String" />

</SelectParameters>

</asp:SqlDataSource>

<asp:GridView ID="GridView1" runat="server" AllowPaging="True"

      AllowSorting="True" AutoGenerateColumns="False"

      BackColor="LightGoldenrodYellow" BorderColor="Tan" BorderWidth="1px"

      CellPadding="2" DataSourceID="SqlDataSource1" ForeColor="Black"

      GridLines="None"

onselectedindexchanged="GridView1_SelectedIndexChanged1"
```

```
   DataKeyNames="GeneNo">
<Columns>
<asp:CommandField HeaderText="Details" ShowSelectButton="True">
<ItemStyle Font-Bold="True" />
</asp:CommandField>
<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"
            SortExpression="GeneNo" ReadOnly="True" />
<asp:BoundField DataField="Description" HeaderText="Description"
            SortExpression="Description" />
<asp:BoundField DataField="Sample1" HeaderText="Sample1" DataFormatString="{0:###.#####}"
            SortExpression="Sample1" />
<asp:BoundField DataField="Sample2" HeaderText="Sample2" DataFormatString="{0:###.#####}"
            SortExpression="Sample2" />
<asp:BoundField DataField="Sample3" HeaderText="Sample3" DataFormatString="{0:###.#####}"
            SortExpression="Sample3" />
<asp:BoundField DataField="Sample4" HeaderText="Sample4" DataFormatString="{0:###.#####}"
            SortExpression="Sample4" />
<asp:BoundField DataField="Sample5" HeaderText="Sample5" DataFormatString="{0:###.#####}"
            SortExpression="Sample5" />
<asp:BoundField DataField="Sample6" HeaderText="Sample6" DataFormatString="{0:###.#####}"
            SortExpression="Sample6" />
<asp:BoundField DataField="Sample7" HeaderText="Sample7" DataFormatString="{0:###.#####}"
            SortExpression="Sample7" />
<asp:BoundField DataField="Sample8" HeaderText="Sample8" DataFormatString="{0:###.#####}"
            SortExpression="Sample8" />
<asp:BoundField DataField="Sample9" HeaderText="Sample9" DataFormatString="{0:###.#####}"
            SortExpression="Sample9" />
</Columns>
<FooterStyle BackColor="Tan" />
<PagerStyle BackColor="PaleGoldenrod" ForeColor="DarkSlateBlue"
          HorizontalAlign="Center" />
<SelectedRowStyle BackColor="DarkSlateBlue" ForeColor="GhostWhite" />
```

```
<HeaderStyle BackColor="Tan" Font-Bold="True" />
<AlternatingRowStyle BackColor="PaleGoldenrod" />
</asp:GridView>
<asp:DetailsView
        ID="DetailsView1" runat="server"
        AutoGenerateRows="False" BackColor="LightGoldenrodYellow" BorderColor="Tan"
        BorderWidth="1px" CellPadding="2" DataSourceID="SqlDataSource2"
        ForeColor="Black" GridLines="None" Height="50px" Width="125px"
      Visible="False">
<FooterStyle BackColor="Tan" />
<PagerStyle BackColor="PaleGoldenrod" ForeColor="DarkSlateBlue"
            HorizontalAlign="Center" />
<Fields>
<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"
                SortExpression="GeneNo" />
<asp:BoundField DataField="GeneLength" HeaderText="GeneLength"
                SortExpression="GeneLength" />
<asp:BoundField DataField="Description" HeaderText="Description"
                SortExpression="Description" />
<asp:TemplateField HeaderText="Seq" SortExpression="Seq">
<EditItemTemplate>
<asp:TextBox ID="TextBox1" runat="server" Text='<%# Bind("Seq") %>'></asp:TextBox>
</EditItemTemplate>
<InsertItemTemplate>
<asp:TextBox ID="TextBox1" runat="server" Text='<%# Bind("Seq") %>'></asp:TextBox>
</InsertItemTemplate>
<ItemTemplate>
<asp:TextBox ID="TextBox2" runat="server" Height="183px" ReadOnly="True"
                    Text='<%# Bind("Seq") %>' TextMode="MultiLine" Width="576px"></asp:TextBox>
</ItemTemplate>
</asp:TemplateField>
```

```
</Fields>

<HeaderStyle BackColor="Tan" Font-Bold="True" />

<EditRowStyle BackColor="DarkSlateBlue" ForeColor="GhostWhite" />

<AlternatingRowStyle BackColor="PaleGoldenrod" />

</asp:DetailsView>

<asp:SqlDataSource ID="SqlDataSource2" runat="server"

      ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"

      SelectCommand="SELECT Transcript_GeneLength.GeneNo, Transcript_GeneLength.Seq,Transcript_GeneLength.GeneLength,
GeneDescription.Description FROM Transcript_GeneLength INNER JOIN GeneDescription ON
SUBSTRING(Transcript_GeneLength.GeneNo, 1, 15) = GeneDescription.GeneNo WHERE (Transcript_GeneLength.GeneNo like '%' +
@GeneNo + '%')">

<SelectParameters>

<asp:ControlParameter ControlID="GridView1" Name="GeneNo"

            PropertyName="SelectedValue" Type="String" />

</SelectParameters>

</asp:SqlDataSource>

</asp:Content>
```

Search gene by DES:

Search gene by DES provides users to search genes by their descriptions. It shows the gene description and the number of the gene and the RPKM values in all the samples

```
<%@ Page Language="C#" MasterPageFile="MasterPage.master" AutoEventWireup="true" CodeFile="Transcript_SearchGeneByDES.aspx.cs"
Inherits="SearchGeneByDES" Title="Untitled Page" %>


<asp:Content ID="Content1" ContentPlaceHolderID="head" Runat="Server">

</asp:Content>

<asp:Content ID="Content2" ContentPlaceHolderID="ContentPlaceHolder1" Runat="Server">

<asp:Label ID="Label1" runat="server" ForeColor="White" Text="GeneDescription"></asp:Label>

<asp:TextBox ID="TextBox1" runat="server"></asp:TextBox>

<asp:Button ID="Button1" runat="server" Text="Search" />

<asp:GridView ID="GridView1" runat="server" AllowPaging="True"

        AllowSorting="True" BackColor="LightGoldenrodYellow" BorderColor="Tan"

        BorderWidth="1px" CellPadding="2" DataSourceID="SqlDataSource1"

        ForeColor="Black" GridLines="None" AutoGenerateColumns="False"

      DataKeyNames="Column1"

onselectedindexchanged="GridView1_SelectedIndexChanged">

<FooterStyle BackColor="Tan" />

<Columns>

<asp:CommandField ShowSelectButton="True">

<ItemStyle Font-Bold="True" />

</asp:CommandField>

<asp:BoundField DataField="Column1" HeaderText="GeneNo"

              SortExpression="Column1" ReadOnly="True" />

<asp:BoundField DataField="Description" HeaderText="Description"

              SortExpression="Description"/>

<asp:BoundField DataField="Sample1" HeaderText="Sample1" DataFormatString="{0:###.#####}"

              SortExpression="Sample1"/>
```

```
<asp:BoundField DataField="Sample2" HeaderText="Sample2" DataFormatString="{0:###.#####}"
                SortExpression="Sample2" />
<asp:BoundField DataField="Sample3" HeaderText="Sample3" DataFormatString="{0:###.#####}"
                SortExpression="Sample3"/>
<asp:BoundField DataField="Sample4" HeaderText="Sample4" DataFormatString="{0:###.#####}"
                SortExpression="Sample4" />
<asp:BoundField DataField="Sample5" HeaderText="Sample5" DataFormatString="{0:###.#####}"
                SortExpression="Sample5"/>
<asp:BoundField DataField="Sample6" HeaderText="Sample6" DataFormatString="{0:###.#####}"
                SortExpression="Sample6" />
<asp:BoundField DataField="Sample7" HeaderText="Sample7" DataFormatString="{0:###.#####}"
                SortExpression="Sample7"/>
<asp:BoundField DataField="Sample8" HeaderText="Sample8" DataFormatString="{0:###.#####}"
                SortExpression="Sample8" />
<asp:BoundField DataField="Sample9" HeaderText="Sample9" DataFormatString="{0:###.#####}"
                SortExpression="Sample9" />
</Columns>
<PagerStyle BackColor="PaleGoldenrod" ForeColor="DarkSlateBlue"
            HorizontalAlign="Center" />
<SelectedRowStyle BackColor="DarkSlateBlue" ForeColor="GhostWhite" />
<HeaderStyle BackColor="Tan" Font-Bold="True" />
<AlternatingRowStyle BackColor="PaleGoldenrod" />
</asp:GridView>
<asp:SqlDataSource ID="SqlDataSource1" runat="server"
        ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"
        SelectCommand="SELECT SUBSTRING(Transcript_Final_RPKM.GeneNo,1,15),GeneDescription.Description , Sample1,Sample2,
Sample3, Sample4, Sample5, Sample6, Sample7, Sample8, Sample9 FROM [Transcript_Final_RPKM] INNER JOIN GeneDescription ON
SUBSTRING(Transcript_Final_RPKM.GeneNo, 1, 15) = GeneDescription.GeneNo where substring (Transcript_Final_RPKM.GeneNo,1,15) in (
select GeneDescription.GeneNo from GeneDescription  WHERE (GeneDescription.Description LIKE N'%' +  @genedesc  + N'%'))  ORDER BY
Transcript_Final_RPKM.GeneNo ">

<SelectParameters>
<asp:ControlParameter ControlID="TextBox1" Name="genedesc"
```

```
            PropertyName="Text" />
</SelectParameters>
</asp:SqlDataSource>
<asp:DetailsView ID="DetailsView1" runat="server" AutoGenerateRows="False"
        BackColor="LightGoldenrodYellow" BorderColor="Tan" BorderWidth="1px"
        CellPadding="2" DataSourceID="SqlDataSource2" ForeColor="Black"
        GridLines="None" Height="50px" Width="125px" Visible="False">
<FooterStyle BackColor="Tan" />
<PagerStyle BackColor="PaleGoldenrod" ForeColor="DarkSlateBlue"
          HorizontalAlign="Center" />
<Fields>
<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"
            SortExpression="GeneNo" />
<asp:BoundField DataField="GeneLength" HeaderText="GeneLength"
            SortExpression="GeneLength" />
<asp:BoundField DataField="Description" HeaderText="Description"
            SortExpression="Description" />
<asp:TemplateField HeaderText="Seq">
<ItemTemplate>
<<asp:TextBox ID="TextBox2" runat="server" Height="183px" ReadOnly="True"
                Text='<%# Bind("Seq") %>' TextMode="MultiLine" Width="576px"></asp:TextBox>
</ItemTemplate>
</asp:TemplateField>
</Fields>
<HeaderStyle BackColor="Tan" Font-Bold="True" />
<EditRowStyle BackColor="DarkSlateBlue" ForeColor="GhostWhite" />
<AlternatingRowStyle BackColor="PaleGoldenrod" />
</asp:DetailsView>
<asp:SqlDataSource ID="SqlDataSource2" runat="server"
        ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"
        SelectCommand="SELECT Transcript_GeneLength.GeneNo, Transcript_GeneLength.GeneLength, Transcript_GeneLength.Seq,
GeneDescription.Description FROM Transcript_GeneLength INNER JOIN GeneDescription ON
```

```
SUBSTRING(Transcript_GeneLength.GeneNo, 1, 15) = GeneDescription.GeneNo WHERE (Transcript_GeneLength.GeneNo like '%' +
@GeneNo + '%')">

<SelectParameters>

<asp:ControlParameter ControlID="GridView1" Name="GeneNo"

        PropertyName="SelectedValue" Type="String" />

</SelectParameters>

</asp:SqlDataSource>

</asp:Content>
```

Access data:

Access data shows number of reads, number of reads which aligned zero time, number of reads which aligned exactly one time, number of reads which aligned more than one times for each experiment

```
<%@ Page Language="C#" MasterPageFile="MasterPage.master" AutoEventWireup="true" CodeFile="Browse.aspx.cs" Inherits="Browse"
Title="Untitled Page" %>

<asp:Content ID="Content1" ContentPlaceHolderID="head" Runat="Server">

</asp:Content>

<asp:Content ID="Content2" ContentPlaceHolderID="ContentPlaceHolder1" Runat="Server">

<p>

<asp:GridView ID="GridView3" runat="server" AllowPaging="True"

        AllowSorting="True" AutoGenerateColumns="False"

        DataSourceID="SqlDataSource1" BackColor="LightGoldenrodYellow"

        BorderColor="Tan" BorderWidth="1px" CellPadding="2" DataKeyNames="SampleID"

        ForeColor="Black" GridLines="None"

onselectedindexchanged="GridView3_SelectedIndexChanged">

<Columns>

<asp:CommandField HeaderText="Details" ShowSelectButton="True">

<ItemStyle Font-Bold="True" />

</asp:CommandField>

<asp:BoundField DataField="SampleID" HeaderText="SampleID"

                SortExpression="SampleID" />

<asp:BoundField DataField="Number_of_Reads" HeaderText="Number_of_Reads"

                SortExpression="Number_of_Reads" />

<asp:BoundField DataField="aligned_0_times" HeaderText="aligned_0_times"

                SortExpression="aligned_0_times" />

<asp:BoundField DataField="column1" HeaderText="aligned_0_times%"

                SortExpression="column1" />

<asp:BoundField DataField="aligned_exactly_1_time"

                HeaderText="aligned_exactly_1_time" SortExpression="aligned_exactly_1_time" />

<asp:BoundField DataField="column2" HeaderText="aligned_exactly_1_time%"
```

```
            SortExpression="column2" />
<asp:BoundField DataField="column3" HeaderText="aligned&gt;1times"
            SortExpression="column3" />
<asp:BoundField DataField="column4" HeaderText="aligned&gt;1times%"
            SortExpression="column4" />
<asp:BoundField DataField="overall_alignment_rate"
            HeaderText="overall_alignment_rate"
            SortExpression="overall_alignment_rate" />
</Columns>
<FooterStyle BackColor="Tan" />
<PagerStyle BackColor="PaleGoldenrod" ForeColor="DarkSlateBlue"
          HorizontalAlign="Center" />
<SelectedRowStyle BackColor="DarkSlateBlue" ForeColor="GhostWhite" />
<HeaderStyle BackColor="Tan" Font-Bold="True" />
<AlternatingRowStyle BackColor="PaleGoldenrod" />
</asp:GridView>
<br />
<asp:SqlDataSource ID="SqlDataSource1" runat="server"
      ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"
      SelectCommand="SELECT [SampleID], [Number_of_Reads], [aligned_0_times], [%aligned_0_times] AS column1,
[aligned_exactly_1_time], [%aligned_exactly_1_time] AS column2, [aligned &gt;1_times] AS column3, [%aligned &gt;1_times] AS column4,
[overall_alignment_rate] FROM [Transcript_Alignment_result]"></asp:SqlDataSource>
<asp:GridView ID="GridView2" runat="server" AllowPaging="True"
      AllowSorting="True" AutoGenerateColumns="False"
      BackColor="LightGoldenrodYellow" BorderColor="Tan" BorderWidth="1px"
      CellPadding="2" DataSourceID="SqlDataSource2" ForeColor="Black"
      GridLines="None" DataKeyNames="GeneNo"
onselectedindexchanged="GridView2_SelectedIndexChanged" Visible="False">
<Columns>
<asp:CommandField ShowSelectButton="True" />
<asp:BoundField DataField="SampleID" HeaderText="SampleID"
            SortExpression="SampleID" />
<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"
```

```
            SortExpression="GeneNo" />
<asp:BoundField DataField="Description" HeaderText="Description"
            SortExpression="Description" />
<asp:BoundField DataField="ReadsCount" HeaderText="ReadsCount"
            SortExpression="ReadsCount" />
<asp:BoundField DataField="RPKM" HeaderText="RPKM"
            SortExpression="RPKM">
</asp:BoundField>
</Columns>
<FooterStyle BackColor="Tan" />
<PagerStyle BackColor="PaleGoldenrod" ForeColor="DarkSlateBlue"
          HorizontalAlign="Center" />
<SelectedRowStyle BackColor="DarkSlateBlue" ForeColor="GhostWhite" />
<HeaderStyle BackColor="Tan" Font-Bold="True" />
<AlternatingRowStyle BackColor="PaleGoldenrod" />
</asp:GridView>
<asp:SqlDataSource ID="SqlDataSource2" runat="server"
      ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"

      SelectCommand="SELECT Transcript_RPKM_ReadsCount.GeneNo, Transcript_RPKM_ReadsCount.ReadsCount,
Transcript_RPKM_ReadsCount.SampleID, Transcript_RPKM_ReadsCount.RPKM, GeneDescription.Description FROM
Transcript_RPKM_ReadsCount INNER JOIN GeneDescription ON SUBSTRING(Transcript_RPKM_ReadsCount.GeneNo, 1, 15) =
GeneDescription.GeneNo WHERE (Transcript_RPKM_ReadsCount.SampleID = @SampleID)">
<SelectParameters>
<asp:ControlParameter ControlID="GridView3" Name="SampleID"
            PropertyName="SelectedValue" Type="String" />
</SelectParameters>
</asp:SqlDataSource>
<asp:DetailsView
      ID="DetailsView1" runat="server"
      AutoGenerateRows="False" BackColor="LightGoldenrodYellow" BorderColor="Tan"
      BorderWidth="1px" CellPadding="2" DataSourceID="SqlDataSource4"
      ForeColor="Black" GridLines="None" Height="50px" Width="125px"
```

```
        Visible="False">
<FooterStyle BackColor="Tan" />
<PagerStyle BackColor="PaleGoldenrod" ForeColor="DarkSlateBlue"
            HorizontalAlign="Center" />
<Fields>
<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"
                SortExpression="GeneNo" />
<asp:BoundField DataField="GeneLength" HeaderText="GeneLength"
                SortExpression="GeneLength" />
<asp:BoundField DataField="Description" HeaderText="Description"
                SortExpression="Description" />
<asp:TemplateField HeaderText="Seq" SortExpression="Seq">
<EditItemTemplate>
<asp:TextBox ID="TextBox2" runat="server" Height="183px" ReadOnly="True"
                    Text='<%# Bind("Seq") %>' TextMode="MultiLine" Width="576px"></asp:TextBox>
</EditItemTemplate>
<InsertItemTemplate>
<asp:TextBox ID="TextBox1" runat="server" Text='<%# Bind("GeneSeq") %>'></asp:TextBox>
</InsertItemTemplate>
<ItemTemplate>
<asp:TextBox ID="TextBox2" runat="server" Height="183px" ReadOnly="True"
                    Text='<%# Bind("Seq") %>' TextMode="MultiLine" Width="576px"></asp:TextBox>
</ItemTemplate>
</asp:TemplateField>
</Fields>
<HeaderStyle BackColor="Tan" Font-Bold="True" />
<EditRowStyle BackColor="DarkSlateBlue" ForeColor="GhostWhite" />
<AlternatingRowStyle BackColor="PaleGoldenrod" />
</asp:DetailsView>
<asp:SqlDataSource ID="SqlDataSource4" runat="server"
        ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"
```

```
        SelectCommand="SELECT Transcript_GeneLength.GeneNo, GeneDescription.Description, Transcript_GeneLength.GeneLength,
Transcript_GeneLength.Seq FROM Transcript_GeneLength INNER JOIN GeneDescription ON Transcript_GeneLength.GeneNo =
GeneDescription.GeneNo WHERE (Transcript_GeneLength.GeneNo LIKE '%' + @GeneNo + '%')">

<SelectParameters>

<asp:ControlParameter ControlID="GridView2" Name="GeneNo"

                PropertyName="SelectedValue" Type="String" />

</SelectParameters>

</asp:SqlDataSource>

<asp:SqlDataSource ID="SqlDataSource3" runat="server"

        ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"

        SelectCommand="SELECT * FROM [GeneLength] WHERE ([GeneNo] = @GeneNo)">

<SelectParameters>

<asp:ControlParameter ControlID="GridView2" Name="GeneNo"

                PropertyName="SelectedValue" Type="String" />

</SelectParameters>

</asp:SqlDataSource>

</asp:Content>
```

DEGSeq:

Compared sample DEGSeq needs Samples, significance and fold-change to be clarified as inputs. It shows gene No, gene description, number of reads for the experiment, number of reads for the control, fold-change, significance.

```
<%@ Page Language="C#" MasterPageFile="MasterPage.master" AutoEventWireup="true" CodeFile="Transcript_DEGSeq.aspx.cs"
Inherits="Test" Title="Untitled Page" %>

<asp:Content ID="Content2" ContentPlaceHolderID="ContentPlaceHolder1" Runat="Server">

<asp:Label ID="Label6" runat="server" style="color: #FFFFFF" Text="SampleID"></asp:Label>

<asp:DropDownList ID="DropDownList2" runat="server" Height="25px" Width="140px">

<asp:ListItem Value="Eth_vs_Control"> Eth vs Control </asp:ListItem>

<asp:ListItem Value="IAA_vs_Control"> IAA vs Control </asp:ListItem>

<asp:ListItem Value="JA_vs_Control"> JA vs Control </asp:ListItem>

<asp:ListItem Value="SA_vs_Control"> SA vs Control </asp:ListItem>

</asp:DropDownList>

            <asp:Label ID="Label4" runat="server" style="color: #FFFFFF; font-size: small"

      Text="significant"></asp:Label>

            <asp:DropDownList ID="DropDownList1" runat="server">

<asp:ListItem Value="0">All  </asp:ListItem>

<asp:ListItem Value="Induced">Induced  </asp:ListItem>

<asp:ListItem Value="Suppressed">Suppressed  </asp:ListItem>

<asp:ListItem Value="Unchanged">Unchanged</asp:ListItem>

</asp:DropDownList>

<asp:Button

      ID="Button1" runat="server" Text="Search" onclick="Button1_Click" />

<asp:Label ID="Label2" runat="server" style="color: #FFFFFF; font-size: small"

      Text="Fold Change"></asp:Label>

<asp:DropDownList ID="DropDownList3" runat="server">

<asp:ListItem Value="0">>>=0  </asp:ListItem>

<asp:ListItem Value="1">>>=1 </asp:ListItem>

<asp:ListItem Value="2">>>=2 </asp:ListItem>

<asp:ListItem Value="3">>>=3 </asp:ListItem>

<asp:ListItem Value="4">>>=4 </asp:ListItem>
```

```
<asp:ListItem Value="5">>=5 </asp:ListItem>

</asp:DropDownList>

<asp:SqlDataSource ID="SqlDataSource1" runat="server"

    ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"

    SelectCommand = "SELECT Transcript_DEGseq.GeneNo , Transcript_DEGseq.value1,
Transcript_DEGseq.value2,Transcript_DEGseq.[log2(FC)], Transcript_DEGseq.[log], Transcript_DEGseq.z_score,
Transcript_DEGseq.p_value,Transcript_DEGseq.q_value_Benjamini, Transcript_DEGseq.q_value_Storey,
Transcript_DEGseq.Signature,Transcript_DEGseq.SampleID, Transcript_DEGseq.Fold_Change,
Transcript_DEGseq.significant,GeneDescription.GeneNo AS Expr1, GeneDescription.Description,GeneDescription.Description AS Expr2
FROM Transcript_DEGseq INNER JOIN GeneDescription ON SUBSTRING(Transcript_DEGseq.GeneNo, 1, 15) =GeneDescription.GeneNo
WHERE (Transcript_DEGseq.SampleID = @SampleID2) AND(ABS(Transcript_DEGseq.Fold_Change) &gt;= @Fold_Change) AND
(Transcript_DEGseq.significant= @significant) OR (Transcript_DEGseq.SampleID = @SampleID2)
AND(ABS(Transcript_DEGseq.Fold_Change) &gt;= @Fold_Change) AND ('0' = @significant)ORDER BY Transcript_DEGseq.Fold_Change
DESC">

<SelectParameters>

<asp:ControlParameter ControlID="DropDownList2" Name="SampleID2"

        PropertyName="SelectedValue" Type="String" />

<asp:ControlParameter ControlID="DropDownList3" Name="Fold_Change"

        PropertyName="SelectedValue" />

<asp:ControlParameter ControlID="DropDownList1" Name="significant"

            PropertyName="SelectedValue" Type="String" />

</SelectParameters>

</asp:SqlDataSource>

<asp:GridView ID="GridView1" runat="server" AllowPaging="True"

    AllowSorting="True" AutoGenerateColumns="False"

    BackColor="LightGoldenrodYellow" BorderColor="Tan" BorderWidth="1px"

    CellPadding="2" DataSourceID="SqlDataSource1" ForeColor="Black"

    GridLines="None" onselectedindexchanged="GridView1_SelectedIndexChanged1" DataKeyNames="GeneNo,SampleID"  >

<Columns>

<asp:CommandField ShowSelectButton="True">

<ItemStyle Font-Bold="True" />

</asp:CommandField>

<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"

        SortExpression="GeneNo" />

<asp:BoundField DataField="Description" HeaderText="Description"

        SortExpression="Description" />
```

```
<asp:BoundField DataField="value1" HeaderText="EXP"
        SortExpression="value1" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="value2" HeaderText="Control"
        SortExpression="value2" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="SampleID" HeaderText="SampleID"
        SortExpression="SampleID" />
<asp:BoundField DataField="Fold_Change" HeaderText="Fold_Change"
        SortExpression="Fold_Change" />
<asp:BoundField DataField="significant" HeaderText="significant"
        SortExpression="significant" />
</Columns>
<FooterStyle BackColor="Tan" />
<PagerStyle BackColor="PaleGoldenrod" ForeColor="DarkSlateBlue"         HorizontalAlign="Center" />
<pagersettings mode="NumericFirstLast"
firstpagetext="First"
lastpagetext="Last"
pagebuttoncount="5"
position="Bottom"/>
<SelectedRowStyle BackColor="DarkSlateBlue" ForeColor="GhostWhite" />
<HeaderStyle BackColor="Tan" Font-Bold="True" />
<AlternatingRowStyle BackColor="PaleGoldenrod" />
</asp:GridView>
<br />
<asp:SqlDataSource ID="SqlDataSource2" runat="server"
    ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"
    SelectCommand="SELECT [GeneNo], [value1], [value2], [log2(FC)] AS column1, [log], [z_score], [p_value], [q_value_Benjamini],
[q_value_Storey], [Signature], [SampleID], [Fold_Change], [significant] FROM [Transcript_DEGseq] WHERE (([GeneNo] = @GeneNo) AND
([SampleID] = @SampleID))">
<SelectParameters>
```

```
<asp:ControlParameter ControlID="GridView1" Name="GeneNo"
        PropertyName="SelectedValue" Type="String" />
<asp:ControlParameter ControlID="Label7" Name="SampleID" PropertyName="Text"
        Type="String" />
</SelectParameters>
</asp:SqlDataSource>
<asp:Label ID="Label7" runat="server" style="color: #FFFFFF" Visible="False"></asp:Label>
<asp:DetailsView ID="DetailsView1" runat="server" AutoGenerateRows="False"
    CellPadding="4" DataSourceID="SqlDataSource2" ForeColor="#333333"
    GridLines="None" Height="50px" Width="125px" Visible="False">
<FooterStyle BackColor="#990000" Font-Bold="True" ForeColor="White" />
<CommandRowStyle BackColor="#FFFFC0" Font-Bold="True" />
<RowStyle BackColor="#FFFBD6" ForeColor="#333333" />
<FieldHeaderStyle BackColor="#FFFF99" Font-Bold="True" />
<PagerStyle BackColor="#FFCC66" ForeColor="#333333" HorizontalAlign="Center" />
<Fields>
<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"
        SortExpression="GeneNo" />
<asp:BoundField DataField="value1" HeaderText="value1"
        SortExpression="value1" />
<asp:BoundField DataField="value2" HeaderText="value2"
        SortExpression="value2" />
<asp:BoundField DataField="column1" HeaderText="column1"
        SortExpression="column1" />
<asp:BoundField DataField="log" HeaderText="log" SortExpression="log" />
<asp:BoundField DataField="z_score" HeaderText="z_score"
        SortExpression="z_score" />
<asp:BoundField DataField="p_value" HeaderText="p_value"
        SortExpression="p_value" />
<asp:BoundField DataField="q_value_Benjamini" HeaderText="q_value_Benjamini"
        SortExpression="q_value_Benjamini" />
<asp:BoundField DataField="q_value_Storey" HeaderText="q_value_Storey"
```

```
            SortExpression="q_value_Storey" />
<asp:BoundField DataField="Signature" HeaderText="Signature"
            SortExpression="Signature" />
<asp:BoundField DataField="SampleID" HeaderText="SampleID"
            SortExpression="SampleID" />
<asp:BoundField DataField="Fold_Change" HeaderText="Fold_Change"
            SortExpression="Fold_Change" />
<asp:BoundField DataField="significant" HeaderText="significant"
            SortExpression="significant" />
</Fields>
<HeaderStyle BackColor="#990000" Font-Bold="True" ForeColor="White" />
<AlternatingRowStyle BackColor="White" />
</asp:DetailsView>
</asp:Content>
```

DESeq:

Compared sample DESeq shows GeneNo, gene description, baseMean A, baseMean B, fold-change, pval, sampleID and significante.

```
<%@ Page Language="C#" MasterPageFile="MasterPage.master" AutoEventWireup="true" CodeFile="Transcript_DESeq.aspx.cs"
Inherits="Test" Title="Untitled Page" %>

<asp:Content ID="Content2" ContentPlaceHolderID="ContentPlaceHolder1" Runat="Server">

<asp:Label ID="Label6" runat="server" style="color: #FFFFFF" Text="SampleID"></asp:Label>

<asp:DropDownList ID="DropDownList2" runat="server" Height="25px" Width="140px">

<asp:ListItem Value="Eth_vs_Control"> Eth vs Control </asp:ListItem>

<asp:ListItem Value="IAA_vs_Control"> IAA vs Control </asp:ListItem>

<asp:ListItem Value="JA_vs_Control"> JA vs Control </asp:ListItem>

<asp:ListItem Value="SA_vs_Control"> SA vs Control </asp:ListItem>

</asp:DropDownList>

            <asp:Label ID="Label4" runat="server" style="color: #FFFFFF; font-size: small"

      Text="significant"></asp:Label>

            <asp:DropDownList ID="DropDownList1" runat="server">

<asp:ListItem Value="0">All  </asp:ListItem>

<asp:ListItem Value="Induced">Induced  </asp:ListItem>

<asp:ListItem Value="Suppressed">Suppressed  </asp:ListItem>

<asp:ListItem Value="Unchanged">Unchanged</asp:ListItem>

</asp:DropDownList>

          <asp:Button

      ID="Button1" runat="server" Text="Search" onclick="Button1_Click" />

<asp:Label ID="Label2" runat="server" style="color: #FFFFFF; font-size: small"

      Text="Fold Change"></asp:Label>

<asp:DropDownList ID="DropDownList3" runat="server">

<asp:ListItem Value="0">>=0  </asp:ListItem>

<asp:ListItem Value="1">>=1 </asp:ListItem>

<asp:ListItem Value="2">>=2 </asp:ListItem>

<asp:ListItem Value="3">>=3 </asp:ListItem>
```

```
<asp:ListItem Value="4">>=4 </asp:ListItem>

<asp:ListItem Value="5">>=5 </asp:ListItem>

</asp:DropDownList>

<asp:SqlDataSource ID="SqlDataSource1" runat="server"

      ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"

   SelectCommand = "SELECT [Soybean].[dbo].[Transcript_DESeq]. [GeneNo]

,[Soybean].[dbo].[Transcript_DESeq].[baseMean]

,[Soybean].[dbo].[Transcript_DESeq].[baseMeanA]

,[Soybean].[dbo].[Transcript_DESeq].[baseMeanB]

,[Soybean].[dbo].[Transcript_DESeq].[Fold_Change]

,[Soybean].[dbo].[Transcript_DESeq].[log2FoldChange]

,[Soybean].[dbo].[Transcript_DESeq].[pval]

,[Soybean].[dbo].[Transcript_DESeq].[padj]

,[Soybean].[dbo].[Transcript_DESeq].[SampleID]

,[Soybean].[dbo].[Transcript_DESeq].[significant]

,GeneDescription.Description

  FROM [Soybean].[dbo].[Transcript_DESeq]  INNER JOIN GeneDescription ON SUBSTRING([Soybean].[dbo].[Transcript_DESeq].GeneNo,
1, 15) =GeneDescription.GeneNo WHERE ([Soybean].[dbo].[Transcript_DESeq].SampleID = @SampleID2)
AND(ABS([Soybean].[dbo].[Transcript_DESeq].Fold_Change) &gt;= @Fold_Change) AND ([Soybean].[dbo].[Transcript_DESeq].significant=
@significant) OR ([Soybean].[dbo].[Transcript_DESeq].SampleID = @SampleID2)
AND(ABS([Soybean].[dbo].[Transcript_DESeq].Fold_Change) &gt;= @Fold_Change) AND ('0' = @significant)ORDER BY
[Soybean].[dbo].[Transcript_DESeq].Fold_Change DESC">

<SelectParameters>

<asp:ControlParameter ControlID="DropDownList2" Name="SampleID2"

                  PropertyName="SelectedValue" Type="String" />

<asp:ControlParameter ControlID="DropDownList3" Name="Fold_Change"

                  PropertyName="SelectedValue" />

<asp:ControlParameter ControlID="DropDownList1" Name="significant"

                        PropertyName="SelectedValue" Type="String" />

</SelectParameters>

</asp:SqlDataSource>

<asp:GridView ID="GridView1" runat="server" AllowPaging="True"

      AllowSorting="True" AutoGenerateColumns="False"

      BackColor="LightGoldenrodYellow" BorderColor="Tan" BorderWidth="1px"
```

```
    CellPadding="2" DataSourceID="SqlDataSource1" ForeColor="Black"
    GridLines="None" onselectedindexchanged="GridView1_SelectedIndexChanged1" DataKeyNames="SampleID,GeneNo"   >
<Columns>
<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"
        SortExpression="GeneNo" />
<asp:BoundField DataField="Description" HeaderText="Description"
        SortExpression="Description" />
<asp:BoundField DataField="baseMean" HeaderText="baseMean"
        DataFormatString="{0:###.#####}" SortExpression="baseMean" />
<asp:BoundField DataField="baseMeanA" HeaderText="baseMeanA"
        DataFormatString="{0:###.#####}" SortExpression="baseMeanA" >
</asp:BoundField>
<asp:BoundField DataField="baseMeanB" HeaderText="baseMeanB"
        DataFormatString="{0:###.#####}" SortExpression="baseMeanB" >
</asp:BoundField>
<asp:BoundField DataField="Fold_Change" HeaderText="Fold_Change"
        DataFormatString="{0:###.#####}" SortExpression="Fold_Change" />
<asp:BoundField DataField="log2FoldChange" HeaderText="log2FoldChange"
        DataFormatString="{0:###.#####}" SortExpression="log2FoldChange" />
<asp:BoundField DataField="pval" HeaderText="pval"
        DataFormatString="{0:###.#####}" SortExpression="pval" />
<asp:BoundField DataField="padj" HeaderText="padj" DataFormatString="{0:###.#####}" SortExpression="padj" />
<asp:BoundField DataField="SampleID" HeaderText="SampleID"
        SortExpression="SampleID" />
<asp:BoundField DataField="significant" HeaderText="significant"
        SortExpression="significant" />
</Columns>
<FooterStyle BackColor="Tan" />
<PagerStyle BackColor="PaleGoldenrod" ForeColor="DarkSlateBlue"         HorizontalAlign="Center" />
<pagersettings mode="NumericFirstLast"
firstpagetext="First"
lastpagetext="Last"
```

```
pagebuttoncount="5"
position="Bottom"/>
<SelectedRowStyle BackColor="DarkSlateBlue" ForeColor="GhostWhite" />
<HeaderStyle BackColor="Tan" Font-Bold="True" />
<AlternatingRowStyle BackColor="PaleGoldenrod" />
</asp:GridView>
<asp:SqlDataSource ID="SqlDataSource2" runat="server"
      ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"
   SelectCommand="SELECT  [GeneNo] ,[baseMean] ,[baseMeanA] ,[baseMeanB] ,[Fold_Change] ,[log2FoldChange]
,[pval] , [padj] ,[SampleID]  ,[significant]
 FROM [Soybean].[dbo].[Transcript_DESeq] WHERE (([GeneNo] = @GeneNo) AND ([SampleID] = @SampleID))">
<SelectParameters>
<asp:ControlParameter ControlID="GridView1" Name="GeneNo"
            PropertyName="SelectedValue" Type="String" />
<asp:ControlParameter ControlID="Label7" Name="SampleID" PropertyName="Text"
            Type="String" />
</SelectParameters>
</asp:SqlDataSource>
<asp:Label ID="Label7" runat="server" style="color: #FFFFFF" Visible="False"></asp:Label>
<asp:DetailsView ID="DetailsView1" runat="server" AutoGenerateRows="False"
      CellPadding="4" DataSourceID="SqlDataSource2" ForeColor="#333333"
      GridLines="None" Height="50px" Width="125px" Visible="False">
<FooterStyle BackColor="#990000" Font-Bold="True" ForeColor="White" />
<CommandRowStyle BackColor="#FFFFC0" Font-Bold="True" />
<RowStyle BackColor="#FFFBD6" ForeColor="#333333" />
<FieldHeaderStyle BackColor="#FFFF99" Font-Bold="True" />
<PagerStyle BackColor="#FFCC66" ForeColor="#333333" HorizontalAlign="Center" />
<Fields>
<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"
            SortExpression="GeneNo" />
<asp:BoundField DataField="baseMean" HeaderText="baseMean"
            SortExpression="baseMean" />
```

```
<asp:BoundField DataField="baseMeanA" HeaderText="baseMeanA"
        SortExpression="baseMeanA" />
<asp:BoundField DataField="baseMeanB" HeaderText="baseMeanB"
        SortExpression="baseMeanB" />
<asp:BoundField DataField="Fold_Change" HeaderText="Fold_Change"
        SortExpression="Fold_Change" />
<asp:BoundField DataField="log2FoldChange" HeaderText="log2FoldChange"
        SortExpression="log2FoldChange" />
<asp:BoundField DataField="pval" HeaderText="pval"
        SortExpression="pval" />
<asp:BoundField DataField="padj" HeaderText="padj"
        SortExpression="padj" />
<asp:BoundField DataField="SampleID" HeaderText="SampleID"
        SortExpression="SampleID" />
<asp:BoundField DataField="significant" HeaderText="significant"
        SortExpression="significant" />
</Fields>
<HeaderStyle BackColor="#990000" Font-Bold="True" ForeColor="White" />
<AlternatingRowStyle BackColor="White" />
</asp:DetailsView>
</asp:Content>
```

edgR:

It shows GeneNo, gene description, logFC, false discovery rate (FDR), sampleID, and significant.

```
<%@ Page Language="C#" MasterPageFile="MasterPage.master" AutoEventWireup="true" CodeFile="Transcript_edgeR.aspx.cs"
Inherits="Test" Title="Untitled Page" %>

<asp:Content ID="Content2" ContentPlaceHolderID="ContentPlaceHolder1" Runat="Server">

<asp:Label ID="Label6" runat="server" style="color: #FFFFFF" Text="SampleID"></asp:Label>

<asp:DropDownList ID="DropDownList2" runat="server" Height="25px" Width="140px">

<asp:ListItem Value="Eth_vs_Control"> Eth vs Control </asp:ListItem>

<asp:ListItem Value="IAA_vs_Control"> IAA vs Control </asp:ListItem>

<asp:ListItem Value="JA_vs_Control"> JA vs Control </asp:ListItem>

<asp:ListItem Value="SA_vs_Control"> SA vs Control </asp:ListItem>

</asp:DropDownList>

            <asp:Label ID="Label4" runat="server" style="color: #FFFFFF; font-size: small"

      Text="significant"></asp:Label>

            <asp:DropDownList ID="DropDownList1" runat="server">

<asp:ListItem Value="0">All  </asp:ListItem>

<asp:ListItem Value="Induced">Induced  </asp:ListItem>

<asp:ListItem Value="Suppressed">Suppressed  </asp:ListItem>

<asp:ListItem Value="Unchanged">Unchanged</asp:ListItem>

</asp:DropDownList>

<asp:Button

      ID="Button1" runat="server" Text="Search" onclick="Button1_Click" />

<asp:Label ID="Label2" runat="server" style="color: #FFFFFF; font-size: small"

      Text="Fold Change"></asp:Label>

<asp:DropDownList ID="DropDownList3" runat="server">

<asp:ListItem Value="0">>=0  </asp:ListItem>

<asp:ListItem Value="1">>=1 </asp:ListItem>

<asp:ListItem Value="2">>=2 </asp:ListItem>

<asp:ListItem Value="3">>=3 </asp:ListItem>

<asp:ListItem Value="4">>=4 </asp:ListItem>
```

```
<asp:ListItem Value="5">>=5 </asp:ListItem>
</asp:DropDownList>
<asp:GridView ID="GridView1" runat="server" AllowPaging="True"
    AllowSorting="True" AutoGenerateColumns="False"
    BackColor="LightGoldenrodYellow" BorderColor="Tan" BorderWidth="1px"
    CellPadding="2" DataSourceID="SqlDataSource1" ForeColor="Black"
    GridLines="None" onselectedindexchanged="GridView1_SelectedIndexChanged1" DataKeyNames="GeneNo" >
<Columns>
<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"
        DataFormatString="{0:###.#####}"  SortExpression="GeneNo" >
</asp:BoundField>
<asp:BoundField DataField="Description" HeaderText="Description"
       DataFormatString="{0:###.#####}"   SortExpression="Description" />
<asp:BoundField DataField="logFC" HeaderText="logFC"
        DataFormatString="{0:###.#####}" SortExpression="logFC" />
<asp:BoundField DataField="logCPM" HeaderText="logCPM"
        DataFormatString="{0:###.#####}"  SortExpression="logCPM" />
<asp:BoundField DataField="PValue" HeaderText="PValue"
         SortExpression="PValue" />
<asp:BoundField DataField="FDR" HeaderText="FDR"    SortExpression="FDR" />
<asp:BoundField DataField="SampleID" HeaderText="SampleID"
        DataFormatString="{0:###.#####}"  SortExpression="SampleID" />
<asp:BoundField DataField="Significant" HeaderText="Significant"
       DataFormatString="{0:###.#####}"  SortExpression="Significant" />
</Columns>
<FooterStyle BackColor="Tan" />
<PagerStyle BackColor="PaleGoldenrod" ForeColor="DarkSlateBlue"         HorizontalAlign="Center" />
<pagersettings mode="NumericFirstLast"
firstpagetext="First"
lastpagetext="Last"
pagebuttoncount="5"
position="Bottom"/>
```

```
<SelectedRowStyle BackColor="DarkSlateBlue" ForeColor="GhostWhite" />

<HeaderStyle BackColor="Tan" Font-Bold="True" />

<AlternatingRowStyle BackColor="PaleGoldenrod" />

</asp:GridView>

<asp:SqlDataSource ID="SqlDataSource1" runat="server"

      ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"

   SelectCommand = "SELECT GeneDescription.GeneNo AS Expr1, GeneDescription.Description, GeneDescription.Description AS Expr2,
Transcript_EdgeR.GeneNo, Transcript_EdgeR.logFC, Transcript_EdgeR.logCPM, Transcript_EdgeR.PValue, Transcript_EdgeR.FDR,
Transcript_EdgeR.SampleID, Transcript_EdgeR.Significant FROM GeneDescription INNER JOIN Transcript_EdgeR ON
GeneDescription.GeneNo = Transcript_EdgeR.GeneNo WHERE (Transcript_EdgeR.SampleID = @SampleID2) AND
(ABS(Transcript_EdgeR.logFC) &gt;= @Fold_Change) AND (Transcript_EdgeR.Significant = @significant) OR ('0' = @significant) AND
(Transcript_EdgeR.SampleID = @SampleID2) AND (ABS(Transcript_EdgeR.logFC) &gt;= @Fold_Change) ORDER BY
Transcript_EdgeR.logFC DESC">



<SelectParameters>

<asp:ControlParameter ControlID="DropDownList2" Name="SampleID2"

                  PropertyName="SelectedValue" Type="String" />

<asp:ControlParameter ControlID="DropDownList3" Name="Fold_Change"

                  PropertyName="SelectedValue" />

<asp:ControlParameter ControlID="DropDownList1" Name="significant"

                        PropertyName="SelectedValue" Type="String" />

</SelectParameters>

</asp:SqlDataSource>

<asp:SqlDataSource ID="SqlDataSource2" runat="server"

      ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"

      SelectCommand="SELECT [GeneNo], [value1], [value2], [log2(FC)] AS column1, [log], [z_score], [p_value], [q_value_Benjamini],
[q_value_Storey], [Signature], [SampleID], [Fold_Change], [significant] FROM [Transcript_DEGseq] WHERE (([GeneNo] = @GeneNo) AND
([SampleID] = @SampleID))">

<SelectParameters>

<asp:ControlParameter ControlID="GridView1" Name="GeneNo"

            PropertyName="SelectedValue" Type="String" />

<asp:ControlParameter ControlID="Label7" Name="SampleID" PropertyName="Text"

            Type="String" />

</SelectParameters>

</asp:SqlDataSource>
```

```
<asp:Label ID="Label7" runat="server" style="color: #FFFFFF" Visible="False"></asp:Label>
<asp:DetailsView ID="DetailsView1" runat="server" AutoGenerateRows="False"
    CellPadding="4" DataSourceID="SqlDataSource2" ForeColor="#333333"
    GridLines="None" Height="50px" Width="125px" Visible="False"
    DataKeyNames="GeneNo">
<FooterStyle BackColor="#990000" Font-Bold="True" ForeColor="White" />
<CommandRowStyle BackColor="#FFFFC0" Font-Bold="True" />
<RowStyle BackColor="#FFFBD6" ForeColor="#333333" />
<FieldHeaderStyle BackColor="#FFFF99" Font-Bold="True" />
<PagerStyle BackColor="#FFCC66" ForeColor="#333333" HorizontalAlign="Center" />
<Fields>
<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"
        SortExpression="GeneNo" />
<asp:BoundField DataField="value1" HeaderText="value1"
        SortExpression="value1" />
<asp:BoundField DataField="value2" HeaderText="value2"
        SortExpression="value2" />
<asp:BoundField DataField="column1" HeaderText="column1"
        SortExpression="column1" />
<asp:BoundField DataField="log" HeaderText="log" SortExpression="log" />
<asp:BoundField DataField="z_score" HeaderText="z_score"
        SortExpression="z_score" />
<asp:BoundField DataField="p_value" HeaderText="p_value"
        SortExpression="p_value" />
<asp:BoundField DataField="q_value_Benjamini" HeaderText="q_value_Benjamini"
        SortExpression="q_value_Benjamini" />
<asp:BoundField DataField="q_value_Storey" HeaderText="q_value_Storey"
        SortExpression="q_value_Storey" />
<asp:BoundField DataField="Signature" HeaderText="Signature"
        SortExpression="Signature" />
<asp:BoundField DataField="SampleID" HeaderText="SampleID"
        SortExpression="SampleID" />
```

```
<asp:BoundField DataField="Fold_Change" HeaderText="Fold_Change"
        SortExpression="Fold_Change" />
<asp:BoundField DataField="significant" HeaderText="significant"
        SortExpression="significant" />
</Fields>
<HeaderStyle BackColor="#990000" Font-Bold="True" ForeColor="White" />
<AlternatingRowStyle BackColor="White" />
</asp:DetailsView>
</asp:Content>
```

NOISeq:

It shows GeneNo, gene description, control_mean, sample_mean, M, D, prob and ranking.

```
<%@ Page Language="C#" MasterPageFile="MasterPage.master" AutoEventWireup="true" CodeFile="Transcript_NOISeq.aspx.cs"
Inherits="Test" Title="Untitled Page" %>

<asp:Content ID="Content2" ContentPlaceHolderID="ContentPlaceHolder1" Runat="Server">

<asp:Label ID="Label6" runat="server" style="color: #FFFFFF" Text="SampleID"></asp:Label>

<asp:DropDownList ID="DropDownList2" runat="server" Height="25px" Width="140px">

<asp:ListItem Value="Eth_vs_Control"> Eth vs Control </asp:ListItem>

<asp:ListItem Value="IAA_vs_Control"> IAA vs Control </asp:ListItem>

<asp:ListItem Value="JA_vs_Control"> JA vs Control </asp:ListItem>

<asp:ListItem Value="SA_vs_Control"> SA vs Control </asp:ListItem>

</asp:DropDownList>

            <asp:Label ID="Label4" runat="server" style="color: #FFFFFF; font-size: small"

      Text="significant"></asp:Label>

            <asp:DropDownList ID="DropDownList1" runat="server">

<asp:ListItem Value="0">All  </asp:ListItem>

<asp:ListItem Value="Induced">Induced  </asp:ListItem>

<asp:ListItem Value="Suppressed">Suppressed  </asp:ListItem>

<asp:ListItem Value="Unchanged">Unchanged</asp:ListItem>

</asp:DropDownList>

          <asp:Button

      ID="Button1" runat="server" Text="Search" onclick="Button1_Click" />

<asp:Label ID="Label2" runat="server" style="color: #FFFFFF; font-size: small"

      Text="Fold Change"></asp:Label>

<asp:DropDownList ID="DropDownList3" runat="server">

<asp:ListItem Value="0">>=0  </asp:ListItem>

<asp:ListItem Value="1">>=1 </asp:ListItem>

<asp:ListItem Value="2">>=2 </asp:ListItem>

<asp:ListItem Value="3">>=3 </asp:ListItem>

<asp:ListItem Value="4">>=4 </asp:ListItem>

<asp:ListItem Value="5">>=5 </asp:ListItem>
```

```
</asp:DropDownList>

<asp:SqlDataSource ID="SqlDataSource1" runat="server"

      ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"

   SelectCommand = "SELECT [Soybean].[dbo].[Transcript_NOISeq_2013].[GeneNo]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[Control_mean]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[SA_mean]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[M]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[D]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[prob]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[ranking]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[SampleID]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[Significance]

,GeneDescription.Description

  FROM [Soybean].[dbo].[Transcript_NOISeq_2013] INNER JOIN GeneDescription ON
SUBSTRING([Soybean].[dbo].[Transcript_NOISeq_2013].GeneNo, 1, 15) =GeneDescription.GeneNo WHERE
([Soybean].[dbo].[Transcript_NOISeq_2013].SampleID = @SampleID2) AND(ABS([Soybean].[dbo].[Transcript_NOISeq_2013].[M] ) &gt;=
@Fold_Change) AND ([Soybean].[dbo].[Transcript_NOISeq_2013].[Significance]= @significant) OR
([Soybean].[dbo].[Transcript_NOISeq_2013].SampleID = @SampleID2) AND(ABS([Soybean].[dbo].[Transcript_NOISeq_2013].[M]) &gt;=
@Fold_Change) AND ('0' = @significant)ORDER BY [Soybean].[dbo].[Transcript_NOISeq_2013].[M] DESC">

<SelectParameters>

<asp:ControlParameter ControlID="DropDownList2" Name="SampleID2"

                  PropertyName="SelectedValue" Type="String" />

<asp:ControlParameter ControlID="DropDownList3" Name="Fold_Change"

                  PropertyName="SelectedValue" />

<asp:ControlParameter ControlID="DropDownList1" Name="significant"

                        PropertyName="SelectedValue" Type="String" />

</SelectParameters>

</asp:SqlDataSource>

<asp:GridView ID="GridView1" runat="server" AllowPaging="True"

      AllowSorting="True" AutoGenerateColumns="False"

      BackColor="LightGoldenrodYellow" BorderColor="Tan" BorderWidth="1px"

      CellPadding="2" DataSourceID="SqlDataSource1" ForeColor="Black"

      GridLines="None" onselectedindexchanged="GridView1_SelectedIndexChanged1" DataKeyNames="GeneNo"  >
```

```
<Columns>
<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"
            DataFormatString="{0:###.#####}" SortExpression="GeneNo" />
<asp:BoundField DataField="Control_mean" HeaderText="Control_mean"
             DataFormatString="{0:###.#####}" SortExpression="Control_mean" />
<asp:BoundField DataField="SA_mean" HeaderText="SA_mean"
             DataFormatString="{0:###.#####}" SortExpression="SA_mean" />
<asp:BoundField DataField="M" HeaderText="M" DataFormatString="{0:###.#####}" SortExpression="M" >
</asp:BoundField>
<asp:BoundField DataField="D" HeaderText="D"  DataFormatString="{0:###.#####}" SortExpression="D" >
</asp:BoundField>
<asp:BoundField DataField="prob" HeaderText="prob" DataFormatString="{0:###.#####}" SortExpression="prob" />
<asp:BoundField DataField="ranking" HeaderText="ranking"
            DataFormatString="{0:###.#####}" SortExpression="ranking" />
<asp:BoundField DataField="SampleID" HeaderText="SampleID"
            DataFormatString="{0:###.#####}" SortExpression="SampleID" />
<asp:BoundField DataField="Significance" HeaderText="Significance"
             DataFormatString="{0:###.#####}" SortExpression="Significance" />
<asp:BoundField DataField="Description" HeaderText="Description"
            DataFormatString="{0:###.#####}"  SortExpression="Description" />
</Columns>
<FooterStyle BackColor="Tan" />
<PagerStyle BackColor="PaleGoldenrod" ForeColor="DarkSlateBlue"          HorizontalAlign="Center" />
<pagersettings mode="NumericFirstLast"
firstpagetext="First"
lastpagetext="Last"
pagebuttoncount="5"
position="Bottom"/>
<SelectedRowStyle BackColor="DarkSlateBlue" ForeColor="GhostWhite" />
<HeaderStyle BackColor="Tan" Font-Bold="True" />
<AlternatingRowStyle BackColor="PaleGoldenrod" />
</asp:GridView>
```

```
<asp:SqlDataSource ID="SqlDataSource2" runat="server"
      ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"
   SelectCommand="SELECT  [GeneNo] ,[baseMean] ,[baseMeanA] ,[baseMeanB] ,[Fold_Change] ,[log2FoldChange]
,[pval] , [padj] ,[SampleID]  ,[significant]
  FROM [Soybean].[dbo].[Transcript_DESeq] WHERE (([GeneNo] = @GeneNo) AND ([SampleID] = @SampleID))">
<SelectParameters>
<asp:ControlParameter ControlID="GridView1" Name="GeneNo"
            PropertyName="SelectedValue" Type="String" />
<asp:ControlParameter ControlID="Label7" Name="SampleID" PropertyName="Text"
            Type="String" />
</SelectParameters>
</asp:SqlDataSource>
<asp:Label ID="Label7" runat="server" style="color: #FFFFFF" Visible="False"></asp:Label>
<asp:DetailsView ID="DetailsView1" runat="server" AutoGenerateRows="False"
      CellPadding="4" DataSourceID="SqlDataSource2" ForeColor="#333333"
      GridLines="None" Height="50px" Width="125px" Visible="False">
<FooterStyle BackColor="#990000" Font-Bold="True" ForeColor="White" />
<CommandRowStyle BackColor="#FFFFC0" Font-Bold="True" />
<RowStyle BackColor="#FFFBD6" ForeColor="#333333" />
<FieldHeaderStyle BackColor="#FFFF99" Font-Bold="True" />
<PagerStyle BackColor="#FFCC66" ForeColor="#333333" HorizontalAlign="Center" />
<Fields>
<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"
            SortExpression="GeneNo" />
<asp:BoundField DataField="baseMean" HeaderText="baseMean"
            SortExpression="baseMean" />
<asp:BoundField DataField="baseMeanA" HeaderText="baseMeanA"
            SortExpression="baseMeanA" />
<asp:BoundField DataField="baseMeanB" HeaderText="baseMeanB"
            SortExpression="baseMeanB" />
<asp:BoundField DataField="Fold_Change" HeaderText="Fold_Change"
            SortExpression="Fold_Change" />
```

```
<asp:BoundField DataField="log2FoldChange" HeaderText="log2FoldChange"
        SortExpression="log2FoldChange" />
<asp:BoundField DataField="pval" HeaderText="pval"
        SortExpression="pval" />
<asp:BoundField DataField="padj" HeaderText="padj"
        SortExpression="padj" />
<asp:BoundField DataField="SampleID" HeaderText="SampleID"
        SortExpression="SampleID" />
<asp:BoundField DataField="significant" HeaderText="significant"
        SortExpression="significant" />
</Fields>
<HeaderStyle BackColor="#990000" Font-Bold="True" ForeColor="White" />
<AlternatingRowStyle BackColor="White" />
</asp:DetailsView>
</asp:Content>
```

Cuffdiff:

It shows geneDesc, p-value, q-value, SampleID, Fold-Change, Significant.

```
<%@ Page Language="C#" MasterPageFile="MasterPage.master" AutoEventWireup="true" CodeFile="Transcript_Cuffdiff.aspx.cs"
Inherits="Test" Title="Untitled Page" %>

<asp:Content ID="Content2" ContentPlaceHolderID="ContentPlaceHolder1" Runat="Server">

<asp:Label ID="Label6" runat="server" style="color: #FFFFFF" Text="SampleID"></asp:Label>

<asp:DropDownList ID="DropDownList2" runat="server" Height="25px" Width="140px">

<asp:ListItem Value="Eth_vs_Control"> Eth vs Control </asp:ListItem>

<asp:ListItem Value="IAA_vs_Control"> IAA vs Control </asp:ListItem>

<asp:ListItem Value="JA_vs_Control"> JA vs Control </asp:ListItem>

<asp:ListItem Value="SA_vs_Control"> SA vs Control </asp:ListItem>

</asp:DropDownList>

            <asp:Label ID="Label4" runat="server" style="color: #FFFFFF; font-size: small"

      Text="significant"></asp:Label>

            <asp:DropDownList ID="DropDownList1" runat="server">

<asp:ListItem Value="0">All  </asp:ListItem>

<asp:ListItem Value="Induced">Induced  </asp:ListItem>

<asp:ListItem Value="Suppressed">Suppressed  </asp:ListItem>

<asp:ListItem Value="Unchanged">Unchanged</asp:ListItem>

</asp:DropDownList>

<asp:Button ID="Button1" runat="server" Text="Search" onclick="Button1_Click" />

<asp:Label ID="Label2" runat="server" style="color: #FFFFFF; font-size: small"

      Text="Fold Change"></asp:Label>

<asp:DropDownList ID="DropDownList3" runat="server">

<asp:ListItem Value="0">>=0  </asp:ListItem>

<asp:ListItem Value="1">>=1 </asp:ListItem>

<asp:ListItem Value="2">>=2 </asp:ListItem>

<asp:ListItem Value="3">>=3 </asp:ListItem>

<asp:ListItem Value="4">>=4 </asp:ListItem>

<asp:ListItem Value="5">>=5 </asp:ListItem>

</asp:DropDownList>
```

```
<asp:SqlDataSource ID="SqlDataSource1" runat="server"

      ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"

      SelectCommand = "SELECT SUBSTRING([Soybean].[dbo].[Transcript_Cuffdiff].GeneNo, 1, 15) AS GeneNo

     ,[value_1],[value_2],[log2(fold_change),[test_stat] ,[p_value] ,[q_value],[sign],[SampleID],[fold_change] ,[significant]

,GeneDescription.Description

  FROM [Soybean].[dbo].[Transcript_Cuffdiff]INNER JOIN GeneDescription ON SUBSTRING([Soybean].[dbo].[Transcript_Cuffdiff].GeneNo,
1, 15) =GeneDescription.GeneNo WHERE ([Soybean].[dbo].[Transcript_Cuffdiff].SampleID = @SampleID2)
AND(ABS([Soybean].[dbo].[Transcript_Cuffdiff].fold_change) &gt;= @Fold_Change) AND
([Soybean].[dbo].[Transcript_Cuffdiff].significant= @significant) OR ([Soybean].[dbo].[Transcript_Cuffdiff].SampleID = @SampleID2)
AND(ABS([Soybean].[dbo].[Transcript_Cuffdiff].fold_change) &gt;= @Fold_Change) AND ('0' = @significant)ORDER BY
[Soybean].[dbo].[Transcript_Cuffdiff].fold_change DESC">

<SelectParameters>

<asp:ControlParameter ControlID="DropDownList2" Name="SampleID2"

                  PropertyName="SelectedValue" Type="String" />

<asp:ControlParameter ControlID="DropDownList3" Name="Fold_Change"

                  PropertyName="SelectedValue" />

<asp:ControlParameter ControlID="DropDownList1" Name="significant"

                        PropertyName="SelectedValue" Type="String" />

</SelectParameters>

</asp:SqlDataSource>

<asp:GridView ID="GridView1" runat="server" AllowPaging="True"

      AllowSorting="True" AutoGenerateColumns="False"

      BackColor="LightGoldenrodYellow" BorderColor="Tan" BorderWidth="1px"

      CellPadding="2" DataSourceID="SqlDataSource1" ForeColor="Black"

      GridLines="None" onselectedindexchanged="GridView1_SelectedIndexChanged1" DataKeyNames="GeneNo">

<Columns>

<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"

            DataFormatString="{0:###.#####}" SortExpression="GeneNo" ReadOnly="True" />

<asp:BoundField DataField="value_1" HeaderText="value_1"

            DataFormatString="{0:###.#####}" SortExpression="value_1" />

<asp:BoundField DataField="value_2" HeaderText="value_2"

           DataFormatString="{0:###.#####}" SortExpression="value_2" />

<asp:BoundField DataField="log2(fold_change)" HeaderText="log(FC)"

           DataFormatString="{0:###.#####}" SortExpression="log2(fold_change)" >
```

```
</asp:BoundField>
<asp:BoundField DataField="test_stat" HeaderText="test_stat"
        DataFormatString="{0:###.#####}"  SortExpression="test_stat" >
</asp:BoundField>
<asp:BoundField DataField="p_value" HeaderText="p_value"
        DataFormatString="{0:###.#####}"  SortExpression="p_value" />
<asp:BoundField DataField="q_value" HeaderText="q_value"
         DataFormatString="{0:###.#####}" SortExpression="q_value" />
<asp:BoundField DataField="sign" HeaderText="sign" DataFormatString="{0:###.#####}" SortExpression="sign" />
<asp:BoundField DataField="SampleID" HeaderText="SampleID"
         DataFormatString="{0:###.#####}" SortExpression="SampleID" />
<asp:BoundField DataField="fold_change" HeaderText="fold_change"
         DataFormatString="{0:###.#####}" SortExpression="fold_change" />
<asp:BoundField DataField="significant" HeaderText="significant"
         DataFormatString="{0:###.#####}" SortExpression="significant" />
<asp:BoundField DataField="Description" HeaderText="Description"
         DataFormatString="{0:###.#####}" SortExpression="Description" />
</Columns>
<FooterStyle BackColor="Tan" />
<PagerStyle BackColor="PaleGoldenrod" ForeColor="DarkSlateBlue"       HorizontalAlign="Center"/>
<pagersettings mode="NumericFirstLast"
firstpagetext="First"
lastpagetext="Last"
pagebuttoncount="5"
position="Bottom"/>
<SelectedRowStyle BackColor="DarkSlateBlue" ForeColor="GhostWhite" />
<HeaderStyle BackColor="Tan" Font-Bold="True" />
<AlternatingRowStyle BackColor="PaleGoldenrod" />
</asp:GridView>
<asp:SqlDataSource ID="SqlDataSource2" runat="server"
    ConnectionString="<%$ ConnectionStrings:SoybeanConnectionString %>"
```

```
    SelectCommand="SELECT  [GeneNo] ,[baseMean] ,[baseMeanA] ,[baseMeanB] ,[Fold_Change] ,[log2FoldChange]
,[pval] , [padj] ,[SampleID]  ,[significant]
  FROM [Soybean].[dbo].[Transcript_DESeq] WHERE (([GeneNo] = @GeneNo) AND ([SampleID] = @SampleID))">
<SelectParameters>
<asp:ControlParameter ControlID="GridView1" Name="GeneNo"
            PropertyName="SelectedValue" Type="String" />
<asp:ControlParameter ControlID="Label7" Name="SampleID" PropertyName="Text"
            Type="String" />
</SelectParameters>
</asp:SqlDataSource>
<asp:Label ID="Label7" runat="server" style="color: #FFFFFF" Visible="False"></asp:Label>
<asp:DetailsView ID="DetailsView1" runat="server" AutoGenerateRows="False"
      CellPadding="4" DataSourceID="SqlDataSource2" ForeColor="#333333"
      GridLines="None" Height="50px" Width="125px" Visible="False">
<FooterStyle BackColor="#990000" Font-Bold="True" ForeColor="White" />
<CommandRowStyle BackColor="#FFFFC0" Font-Bold="True" />
<RowStyle BackColor="#FFFBD6" ForeColor="#333333" />
<FieldHeaderStyle BackColor="#FFFF99" Font-Bold="True" />
<PagerStyle BackColor="#FFCC66" ForeColor="#333333" HorizontalAlign="Center" />
<Fields>
<asp:BoundField DataField="GeneNo" HeaderText="GeneNo"
            SortExpression="GeneNo" />
<asp:BoundField DataField="baseMean" HeaderText="baseMean"
            SortExpression="baseMean" />
<asp:BoundField DataField="baseMeanA" HeaderText="baseMeanA"
            SortExpression="baseMeanA" />
<asp:BoundField DataField="baseMeanB" HeaderText="baseMeanB"
            SortExpression="baseMeanB" />
<asp:BoundField DataField="Fold_Change" HeaderText="Fold_Change"
            SortExpression="Fold_Change" />
<asp:BoundField DataField="log2FoldChange" HeaderText="log2FoldChange"
            SortExpression="log2FoldChange" />
```

```
<asp:BoundField DataField="pval" HeaderText="pval"
        SortExpression="pval" />
<asp:BoundField DataField="padj" HeaderText="padj"
        SortExpression="padj" />
<asp:BoundField DataField="SampleID" HeaderText="SampleID"
        SortExpression="SampleID" />
<asp:BoundField DataField="significant" HeaderText="significant"
        SortExpression="significant" />
</Fields>
<HeaderStyle BackColor="#990000" Font-Bold="True" ForeColor="White" />
<AlternatingRowStyle BackColor="White" />
</asp:DetailsView>
</asp:Content>
```

Soybean_Seed_Proteins_2D_Gel_DB:

It houses several 2D Gel images showing isolated soybean seed proteins.

```
<%@ Master Language="C#" AutoEventWireup="true" CodeFile="MasterPage2.master.cs" Inherits="MasterPage" %>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<html xmlns="http://www.w3.org/1999/xhtml" >

<head>

<meta http-equiv="content-type" content="text/html; charset=utf-8" />

<meta name="description" content="" />

<meta name="keywords" content="" />

<title>Soybean Seed Proteins 2D Gel DB</title>

<link href="http://fonts.googleapis.com/css?family=Abel" rel="stylesheet" type="text/css" />

<link rel="stylesheet" type="text/css" href="style.css" />

<style type="text/css">

.style1 {

font-size: medium;

}

</style>

</head>

<body>

<div id="outer">

<div id="wrapper">

<div id="menu">

<ul>

<li class="first"><a href="Home.aspx">Home</a></li>

<li><a href="Gel1.aspx" >Gel1</a></li>

<li><a href="Gel2.aspx" >Gel2</a></li>

<li><a href="Gel3.aspx" >Gel3</a></li>

</ul>

<br class="clearfix" />

</div>

<div id="header">
```

```
<div id="logo">
<h1 class="style1">
 SoyProDB: Soybean Seed Proteins 2D Gel DB</h1>
</div>
<div id="search">
<form action="" method="post">
</form>
</div>
</div>
<div id="page">
<form id="form1" runat="server">
<div>
<asp:contentplaceholder id="ContentPlaceHolder1" runat="server">
<br />
</asp:contentplaceholder>
</div>
</form>
<br class="clearfix" />
<div id="footer">
                </body>
</html>
```

```
<%@ Page Language="C#" MasterPageFile="MasterPage2.master" AutoEventWireup="true" CodeFile="Gel1.aspx.cs" Inherits="Default3"
Title="Untitled Page" %>

<script runat="server">

protected void TextBox1_TextChanged(object sender, EventArgs e)

   {

Page.Validate();

      GridView1.Visible=Page.IsValid;

   }

</script>

<asp:Content ID="Content1" ContentPlaceHolderID="ContentPlaceHolder1" Runat="Server">

<b>Please type in the number of the protein you wish to view information on, and click on submit </b>

<asp:Label ID="Label1" runat="server"

      Font-Bold="True" Text="SpotId" Width="57px"></asp:Label>

<asp:TextBox ID="TextBox1" runat="server" ontextchanged="TextBox1_TextChanged"></asp:TextBox>

<asp:Button ID="Button1" runat="server" Text="Submit" Font-Bold="True"

      ForeColor="#009933" Width="83px" />

<asp:RangeValidator ID="RangeValidator1" runat="server"

      ControlToValidate="TextBox1"

      ErrorMessage="Please enter a number in the range 1 to 76" MaximumValue="76"

      MinimumValue="1" Type="Integer" EnableClientScript="False"></asp:RangeValidator>

<br />

<asp:GridView ID="GridView1" runat="server" AutoGenerateColumns="False" CellPadding="4"

      DataSourceID="AccessDataSource1" ForeColor="#333333" GridLines="None"

      EmptyDataText=""

      Font-Size="10pt" Width="1124px">

<RowStyle BackColor="#E3EAEB" />

<Columns>

<asp:BoundField DataField="Spot_ID" HeaderText="Spot_ID"

            SortExpression="Spot_ID" >

<ItemStyle HorizontalAlign="Center" />
```

```
</asp:BoundField>
<asp:BoundField DataField="column1" HeaderText="Theoretical _pI _Mr"
        SortExpression="column1" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Protein_Identity" HeaderText="Protein_Identity"
        SortExpression="Protein_Identity" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Peptides_Matched" HeaderText="Peptides_Matched" SortExpression="Peptides_Matched" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="column2" HeaderText="Sequence_Coverage"
        SortExpression="column2" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="MOWSE_score" HeaderText="MOWSE_score" SortExpression="MOWSE_score" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Accession_no" HeaderText="Accession_no"
        SortExpression="Accession_no" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="ID_method" HeaderText="ID_method"
        SortExpression="ID_method" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="column3" HeaderText="Protein Type(ST_AL_AN)"
        SortExpression="column3" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Expect_value" HeaderText="Expect_value" SortExpression="Expect_value" >
```

```
<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>

<asp:BoundField DataField="Theoretical_pI" HeaderText="Theoretical_pI" SortExpression="Theoretical_pI" >

<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>

<asp:BoundField DataField="Theoretical_Mr" HeaderText="Theoretical_Mr" SortExpression="Theoretical_Mr" >

<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>

</Columns>

<FooterStyle BackColor="#1C5E55" Font-Bold="True" ForeColor="White" />

<PagerStyle BackColor="#666666" ForeColor="White" HorizontalAlign="Center" />

<SelectedRowStyle BackColor="#C5BBAF" Font-Bold="True" ForeColor="#333333" />

<HeaderStyle BackColor="#1C5E55" Font-Bold="False" ForeColor="White"

        Font-Size="X-Small" />

<EditRowStyle BackColor="#7C6F57" />

<AlternatingRowStyle BackColor="White" />

</asp:GridView>

<img src="pH(3-10)_M.jpg" width="900" /><br />

<asp:AccessDataSource ID="AccessDataSource1" runat="server" DataFile="App_Data/Database1.mdb"

      SelectCommand="SELECT [Spot ID] AS Spot_ID, [Theoretical pI / Mr] AS column1, [Protein Identity] AS Protein_Identity, [Peptides
Matched] AS Peptides_Matched, [% Sequence Coverage] AS column2, [MOWSE score] AS MOWSE_score, [Accession no] AS Accession_no,
[ID method] AS ID_method, [Protein Type (ST, AL, AN)] AS column3, [Expect value] AS Expect_value, [Theoretical pI] AS Theoretical_pI,
[Theoretical Mr] AS Theoretical_Mr FROM [Sheet2] WHERE ([Spot ID] = ?)">

<SelectParameters>

<asp:ControlParameter ControlID="TextBox1" Name="Spot_ID" PropertyName="Text" Type="Double" />

</SelectParameters>

</asp:AccessDataSource>

</asp:Content>
```

```
<%@ Page Language="C#" MasterPageFile="MasterPage2.master" AutoEventWireup="true" CodeFile="Gel2.aspx.cs" Inherits="Default3"
Title="Untitled Page" %>

<script runat="server">


protected void TextBox1_TextChanged(object sender, EventArgs e)

   {

Page.Validate();

      GridView1.Visible=Page.IsValid;

   }

</script>

<asp:Content ID="Content1" ContentPlaceHolderID="ContentPlaceHolder1" Runat="Server">

<b>Please type in the number of the protein you wish to view information on, and click on submit<br />

<asp:Label ID="Label1" runat="server" Font-Bold="True" Text="SpotId" Width="72px"></asp:Label> <asp:TextBox

      ID="TextBox1" runat="server" ontextchanged="TextBox1_TextChanged"></asp:TextBox>

<asp:Button ID="Button1" runat="server" Text="Submit" Font-Bold="True"

      ForeColor="#009933" Width="83px" />

<asp:RangeValidator ID="RangeValidator1" runat="server"

      ControlToValidate="TextBox1"

      ErrorMessage="Please enter a number in the range 1 to 76" MaximumValue="76"

      MinimumValue="1" Type="Integer" EnableClientScript="False"></asp:RangeValidator>

<asp:GridView ID="GridView1" runat="server" AutoGenerateColumns="False" CellPadding="4"

      DataSourceID="AccessDataSource1" ForeColor="#333333" GridLines="None"

      EmptyDataText=""

      Font-Size="10pt" Width="1121px">

<RowStyle BackColor="#E3EAEB" />

<Columns>

<asp:BoundField DataField="Spot_ID" HeaderText="Spot_ID"

            SortExpression="Spot_ID" >

<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>
```

```
<asp:BoundField DataField="column1" HeaderText="Theoretical _pI _Mr"
        SortExpression="column1" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Protein_Identity" HeaderText="Protein_Identity"
        SortExpression="Protein_Identity" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Peptides_Matched" HeaderText="Peptides_Matched"
        SortExpression="Peptides_Matched" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="column2" HeaderText="Sequence Coverage"
        SortExpression="column2" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="MOWSE_score" HeaderText="MOWSE_score"
        SortExpression="MOWSE_score" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Accession_no" HeaderText="Accession_no"
        SortExpression="Accession_no" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="ID_method" HeaderText="ID_method"
        SortExpression="ID_method" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="column3" HeaderText="Protein Type (ST_AL_AN)"
        SortExpression="column3" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
```

```
<asp:BoundField DataField="Expect_value" HeaderText="Expect_value"
            SortExpression="Expect_value" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Theoretical_pI" HeaderText="Theoretical_pI"
            SortExpression="Theoretical_pI" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Theoretical_Mr" HeaderText="Theoretical_Mr"
            SortExpression="Theoretical_Mr" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
</Columns>
<FooterStyle BackColor="#1C5E55" Font-Bold="True" ForeColor="White" />
<PagerStyle BackColor="#666666" ForeColor="White" HorizontalAlign="Center" />
<SelectedRowStyle BackColor="#C5BBAF" Font-Bold="True" ForeColor="#333333" />
<HeaderStyle BackColor="#1C5E55" Font-Bold="False" ForeColor="White"
          Font-Size="X-Small" />
<EditRowStyle BackColor="#7C6F57" />
<AlternatingRowStyle BackColor="White" />
</asp:GridView>
<img src="pH(4-7)_M.jpg" width="900" /><br />
<asp:AccessDataSource ID="AccessDataSource1" runat="server" DataFile="App_Data/Database1.mdb"
      SelectCommand="SELECT [Spot ID] AS Spot_ID, [Theoretical pI / Mr] AS column1, [Protein Identity] AS Protein_Identity, [Peptides
Matched] AS Peptides_Matched, [% Sequence Coverage] AS column2, [MOWSE score] AS MOWSE_score, [Accession no] AS Accession_no,
[ID method] AS ID_method, [Protein Type (ST, AL, AN)] AS column3, [Expect value] AS Expect_value, [Theoretical pI] AS Theoretical_pI,
[Theoretical Mr] AS Theoretical_Mr FROM [Sheet2] WHERE ([Spot ID] = ?)">
<SelectParameters>
<asp:ControlParameter ControlID="TextBox1" Name="Spot_ID" PropertyName="Text" Type="Double" />
</SelectParameters>
</asp:AccessDataSource>
</asp:Content>
```

```
<%@ Page Language="C#" MasterPageFile="MasterPage2.master" AutoEventWireup="true" CodeFile="Gel3.aspx.cs" Inherits="Default3"
Title="Untitled Page" %>

<script runat="server">

protected void TextBox1_TextChanged(object sender, EventArgs e)

   {

Page.Validate();

      GridView1.Visible=Page.IsValid;

   }

</script>

<asp:Content ID="Content1" ContentPlaceHolderID="ContentPlaceHolder1" Runat="Server">

<b>Please type in the number of the protein you wish to view information on, and click on submit<br />

<asp:Label ID="Label1" runat="server" Font-Bold="True" Text="SpotId" Width="73px"></asp:Label>

<asp:TextBox ID="TextBox1" runat="server"

ontextchanged="TextBox1_TextChanged"></asp:TextBox>

<asp:Button ID="Button1" runat="server" Text="Submit" Font-Bold="True"

      ForeColor="#009933" Width="83px"  />

<asp:RangeValidator ID="RangeValidator1" runat="server"

      ErrorMessage="Please enter a number in the range 1 to 76"

   EnableClientScript="False" MaximumValue="76" MinimumValue="1" Type="Integer"

      ControlToValidate="TextBox1"></asp:RangeValidator>

<asp:GridView ID="GridView1" runat="server" AutoGenerateColumns="False" CellPadding="4"

      DataSourceID="AccessDataSource1" ForeColor="#333333" GridLines="None"

      EmptyDataText=""  Font-Size="10pt" Width="1130px">

<RowStyle BackColor="#E3EAEB" />

<Columns>

<asp:BoundField DataField="Spot_ID" HeaderText="Spot_ID"

            SortExpression="Spot_ID" >

<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>

<asp:BoundField DataField="column1" HeaderText="Theoretical _pI _Mr"

            SortExpression="column1" >

<ItemStyle HorizontalAlign="Center" />
```

```
</asp:BoundField>
<asp:BoundField DataField="Protein_Identity" HeaderText="Protein_Identity"
        SortExpression="Protein_Identity" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Peptides_Matched" HeaderText="Peptides_Matched"
        SortExpression="Peptides_Matched" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="column2" HeaderText="Sequence Coverage"
        SortExpression="column2" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="MOWSE_score" HeaderText="MOWSE_score"
        SortExpression="MOWSE_score" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Accession_no" HeaderText="Accession_no"
        SortExpression="Accession_no" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="ID_method" HeaderText="ID_method"
        SortExpression="ID_method" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="column3" HeaderText="Protein Type (ST_AL_AN)"
        SortExpression="column3" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Expect_value" HeaderText="Expect_value"
        SortExpression="Expect_value" >
<ItemStyle HorizontalAlign="Center" />
```

```
</asp:BoundField>

<asp:BoundField DataField="Theoretical_pI" HeaderText="Theoretical_pI"

        SortExpression="Theoretical_pI" >

<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>

<asp:BoundField DataField="Theoretical_Mr" HeaderText="Theoretical_Mr"

        SortExpression="Theoretical_Mr" >

<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>

</Columns>

<FooterStyle BackColor="#1C5E55" Font-Bold="True" ForeColor="White" />

<PagerStyle BackColor="#666666" ForeColor="White" HorizontalAlign="Center" />

<SelectedRowStyle BackColor="#C5BBAF" Font-Bold="True" ForeColor="#333333" />

<HeaderStyle BackColor="#1C5E55" Font-Bold="True" ForeColor="White"

      Font-Size="X-Small" />

<EditRowStyle BackColor="#7C6F57" />

<AlternatingRowStyle BackColor="White" />

</asp:GridView>

<img src="pH(6-11)_M.jpg" width="900" /><br />

<asp:AccessDataSource ID="AccessDataSource1" runat="server" DataFile="App_Data/Database1.mdb"

    SelectCommand="SELECT [Spot ID] AS Spot_ID, [Theoretical pI / Mr] AS column1, [Protein Identity] AS Protein_Identity, [Peptides
Matched] AS Peptides_Matched, [% Sequence Coverage] AS column2, [MOWSE score] AS MOWSE_score, [Accession no] AS Accession_no,
[ID method] AS ID_method, [Protein Type (ST, AL, AN)] AS column3, [Expect value] AS Expect_value, [Theoretical pI] AS Theoretical_pI,
[Theoretical Mr] AS Theoretical_Mr FROM [Sheet2] WHERE ([Spot ID] = ?)">

<SelectParameters>

<asp:ControlParameter ControlID="TextBox1" Name="Spot_ID" PropertyName="Text" Type="Double" />

</SelectParameters>

</asp:AccessDataSource>

</asp:Content>
```

Soybean_SCN_proteins_2D_Gel_DB:

It contains information about a 2D Gel image that represents isolated SCN proteins.

```
<%@ Master Language="C#" AutoEventWireup="true" CodeFile="MasterPage2.master.cs" Inherits="MasterPage" %>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">


<html xmlns="http://www.w3.org/1999/xhtml" >

<head>

<meta http-equiv="content-type" content="text/html; charset=utf-8" />

<meta name="description" content="" />

<meta name="keywords" content="" />

<title>Soybean Seed Proteins 2D Gel DB</title>

<link href="http://fonts.googleapis.com/css?family=Abel" rel="stylesheet" type="text/css" />

<link rel="stylesheet" type="text/css" href="style.css" />

<style type="text/css">

.style1 {

font-size: medium;

}

</style>

</head>

<body>

<div id="outer">

<div id="wrapper">

<div id="menu">

<ul>

<li class="first"><a href="Home.aspx">Home</a></li>

<li><a href="Gel1.aspx" >Gel1</a></li>

<li><a href="Gel2.aspx" >Gel2</a></li>

<li><a href="Gel3.aspx" >Gel3</a></li>

</ul>

<br class="clearfix" />

</div>
```

```
<div id="header">
<div id="logo">
<h1 class="style1">
 SoyProDB: Soybean Seed Proteins 2D Gel DB</h1>
</div>
<div id="search">
<form action="" method="post">
</form>
<div id="page">
<form id="form1" runat="server">
<div>
<asp:contentplaceholder id="ContentPlaceHolder1" runat="server">
</asp:contentplaceholder>
</div>
</form>
<br class="clearfix" />
<div id="footer">
</body>
</html>
```

```
<%@ Page Language="C#" MasterPageFile="MasterPage2.master" AutoEventWireup="true" CodeFile="Gel1_1.aspx.cs" Inherits="Default3"
Title="Untitled Page" %>

<script runat="server">

protected void TextBox1_TextChanged(object sender, EventArgs e)

   {

Page.Validate();

      GridView1.Visible=Page.IsValid;

   }

</script>

<asp:Content ID="Content1" ContentPlaceHolderID="ContentPlaceHolder1" Runat="Server">

<b>Please type in the number of the protein you wish to view information on, and click on submit </b>


<asp:Label ID="Label1" runat="server"

      Font-Bold="True" Text="SpotId" Width="57px"></asp:Label>

<asp:TextBox ID="TextBox1" runat="server" ontextchanged="TextBox1_TextChanged"></asp:TextBox>

<asp:Button ID="Button1" runat="server" Text="Submit" Font-Bold="True"

      ForeColor="#009933" Width="83px" />

<asp:RangeValidator ID="RangeValidator1" runat="server"

      ControlToValidate="TextBox1"

      ErrorMessage="Please enter a number in the range 1 to 76" MaximumValue="76"

      MinimumValue="1" Type="Integer" EnableClientScript="False"></asp:RangeValidator>

<asp:GridView ID="GridView1" runat="server" AutoGenerateColumns="False" CellPadding="4"

      DataSourceID="AccessDataSource1" ForeColor="#333333" GridLines="None"

      EmptyDataText=""

      Font-Size="10pt" Width="1124px">

<RowStyle BackColor="#E3EAEB" />

<Columns>

<asp:BoundField DataField="Spot_ID" HeaderText="Spot_ID"

            SortExpression="Spot_ID" >

<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>
```

```
<asp:BoundField DataField="column1" HeaderText="Theoretical _pI _Mr"
        SortExpression="column1" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Protein_Identity" HeaderText="Protein_Identity"
        SortExpression="Protein_Identity" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Peptides_Matched" HeaderText="Peptides_Matched" SortExpression="Peptides_Matched" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="column2" HeaderText="Sequence_Coverage"
        SortExpression="column2" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="MOWSE_score" HeaderText="MOWSE_score" SortExpression="MOWSE_score" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Accession_no" HeaderText="Accession_no"
        SortExpression="Accession_no" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="ID_method" HeaderText="ID_method"
        SortExpression="ID_method" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="column3" HeaderText="Protein Type(ST_AL_AN)"
        SortExpression="column3" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Expect_value" HeaderText="Expect_value" SortExpression="Expect_value" >
<ItemStyle HorizontalAlign="Center" />
```

```
</asp:BoundField>

<asp:BoundField DataField="Theoretical_pI" HeaderText="Theoretical_pI" SortExpression="Theoretical_pI" >

<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>

<asp:BoundField DataField="Theoretical_Mr" HeaderText="Theoretical_Mr" SortExpression="Theoretical_Mr" >

<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>

</Columns>

<FooterStyle BackColor="#1C5E55" Font-Bold="True" ForeColor="White" />

<PagerStyle BackColor="#666666" ForeColor="White" HorizontalAlign="Center" />

<SelectedRowStyle BackColor="#C5BBAF" Font-Bold="True" ForeColor="#333333" />

<HeaderStyle BackColor="#1C5E55" Font-Bold="False" ForeColor="White"

        Font-Size="X-Small" />

<EditRowStyle BackColor="#7C6F57" />

<AlternatingRowStyle BackColor="White" />

</asp:GridView>

<img src="pH(3-10)_M.jpg" width="900" /><br />

<asp:AccessDataSource ID="AccessDataSource1" runat="server" DataFile="App_Data/Database1.mdb"

      SelectCommand="SELECT [Spot ID] AS Spot_ID, [Theoretical pI / Mr] AS column1, [Protein Identity] AS Protein_Identity, [Peptides
Matched] AS Peptides_Matched, [% Sequence Coverage] AS column2, [MOWSE score] AS MOWSE_score, [Accession no] AS Accession_no,
[ID method] AS ID_method, [Protein Type (ST, AL, AN)] AS column3, [Expect value] AS Expect_value, [Theoretical pI] AS Theoretical_pI,
[Theoretical Mr] AS Theoretical_Mr FROM [Sheet2] WHERE ([Spot ID] = ?)">

<SelectParameters>

<asp:ControlParameter ControlID="TextBox1" Name="Spot_ID" PropertyName="Text" Type="Double" />

</SelectParameters>

</asp:AccessDataSource>

</asp:Content>
```

```
<%@ Page Language="C#" MasterPageFile="MasterPage2.master" AutoEventWireup="true" CodeFile="Gel2_1.aspx.cs" Inherits="Default3"
Title="Untitled Page" %>

<script runat="server">

protected void TextBox1_TextChanged(object sender, EventArgs e)

   {

Page.Validate();

      GridView1.Visible=Page.IsValid;

   }

</script>

<asp:Content ID="Content1" ContentPlaceHolderID="ContentPlaceHolder1" Runat="Server">

<b>Please type in the number of the protein you wish to view information on, and click on submit<br />

<asp:Label ID="Label1" runat="server" Font-Bold="True" Text="SpotId" Width="72px"></asp:Label> <asp:TextBox

      ID="TextBox1" runat="server" ontextchanged="TextBox1_TextChanged"></asp:TextBox>

<asp:Button ID="Button1" runat="server" Text="Submit" Font-Bold="True"

      ForeColor="#009933" Width="83px" />

<asp:RangeValidator ID="RangeValidator1" runat="server"

      ControlToValidate="TextBox1"

      ErrorMessage="Please enter a number in the range 1 to 76" MaximumValue="76"

      MinimumValue="1" Type="Integer" EnableClientScript="False"></asp:RangeValidator>

<asp:GridView ID="GridView1" runat="server" AutoGenerateColumns="False" CellPadding="4"

      DataSourceID="AccessDataSource1" ForeColor="#333333" GridLines="None"

      EmptyDataText=""

      Font-Size="10pt" Width="1121px">

<RowStyle BackColor="#E3EAEB" />

<Columns>

<asp:BoundField DataField="Spot_ID" HeaderText="Spot_ID"

            SortExpression="Spot_ID" >

<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>

<asp:BoundField DataField="column1" HeaderText="Theoretical _pI _Mr"

            SortExpression="column1" >

<ItemStyle HorizontalAlign="Center" />
```

```
</asp:BoundField>
<asp:BoundField DataField="Protein_Identity" HeaderText="Protein_Identity"
        SortExpression="Protein_Identity" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Peptides_Matched" HeaderText="Peptides_Matched"
        SortExpression="Peptides_Matched" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="column2" HeaderText="Sequence Coverage"
        SortExpression="column2" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="MOWSE_score" HeaderText="MOWSE_score"
        SortExpression="MOWSE_score" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Accession_no" HeaderText="Accession_no"
        SortExpression="Accession_no" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="ID_method" HeaderText="ID_method"
        SortExpression="ID_method" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="column3" HeaderText="Protein Type (ST_AL_AN)"
        SortExpression="column3" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Expect_value" HeaderText="Expect_value"
        SortExpression="Expect_value" >
<ItemStyle HorizontalAlign="Center" />
```

```
</asp:BoundField>

<asp:BoundField DataField="Theoretical_pI" HeaderText="Theoretical_pI"

        SortExpression="Theoretical_pI" >

<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>

<asp:BoundField DataField="Theoretical_Mr" HeaderText="Theoretical_Mr"

        SortExpression="Theoretical_Mr" >

<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>

</Columns>

<FooterStyle BackColor="#1C5E55" Font-Bold="True" ForeColor="White" />

<PagerStyle BackColor="#666666" ForeColor="White" HorizontalAlign="Center" />

<SelectedRowStyle BackColor="#C5BBAF" Font-Bold="True" ForeColor="#333333" />

<HeaderStyle BackColor="#1C5E55" Font-Bold="False" ForeColor="White"

      Font-Size="X-Small" />

<EditRowStyle BackColor="#7C6F57" />

<AlternatingRowStyle BackColor="White" />

</asp:GridView>

<img src="pH(4-7)_M.jpg" width="900" /><br />

<asp:AccessDataSource ID="AccessDataSource1" runat="server" DataFile="App_Data/Database1.mdb"

    SelectCommand="SELECT [Spot ID] AS Spot_ID, [Theoretical pI / Mr] AS column1, [Protein Identity] AS Protein_Identity, [Peptides
Matched] AS Peptides_Matched, [% Sequence Coverage] AS column2, [MOWSE score] AS MOWSE_score, [Accession no] AS Accession_no,
[ID method] AS ID_method, [Protein Type (ST, AL, AN)] AS column3, [Expect value] AS Expect_value, [Theoretical pI] AS Theoretical_pI,
[Theoretical Mr] AS Theoretical_Mr FROM [Sheet2] WHERE ([Spot ID] = ?)">

<SelectParameters>

<asp:ControlParameter ControlID="TextBox1" Name="Spot_ID" PropertyName="Text" Type="Double" />

</SelectParameters>

</asp:AccessDataSource>

</asp:Content>

<%@ Page Language="C#" MasterPageFile="MasterPage2.master" AutoEventWireup="true" CodeFile="Gel3_1.aspx.cs" Inherits="Default3"
Title="Untitled Page" %>

<script runat="server">
```

```
protected void TextBox1_TextChanged(object sender, EventArgs e)
   {
Page.Validate();
      GridView1.Visible=Page.IsValid;
   }
</script>
<asp:Content ID="Content1" ContentPlaceHolderID="ContentPlaceHolder1" Runat="Server">
<b>Please type in the number of the protein you wish to view information on, and click on submit<br />
<asp:Label ID="Label1" runat="server" Font-Bold="True" Text="SpotId" Width="73px"></asp:Label>
<asp:TextBox ID="TextBox1" runat="server"
ontextchanged="TextBox1_TextChanged"></asp:TextBox>
<asp:Button ID="Button1" runat="server" Text="Submit" Font-Bold="True"
      ForeColor="#009933" Width="83px"  />
        
<asp:RangeValidator ID="RangeValidator1" runat="server"
      ErrorMessage="Please enter a number in the range 1 to 76"
   EnableClientScript="False" MaximumValue="76" MinimumValue="1" Type="Integer"
      ControlToValidate="TextBox1"></asp:RangeValidator>
<br />
<asp:GridView ID="GridView1" runat="server" AutoGenerateColumns="False" CellPadding="4"
      DataSourceID="AccessDataSource1" ForeColor="#333333" GridLines="None"
      EmptyDataText=""
      Font-Size="10pt" Width="1130px">
<RowStyle BackColor="#E3EAEB" />
<Columns>
<asp:BoundField DataField="Spot_ID" HeaderText="Spot_ID"
            SortExpression="Spot_ID" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="column1" HeaderText="Theoretical _pI _Mr"
            SortExpression="column1" >
<ItemStyle HorizontalAlign="Center" />
```

```
</asp:BoundField>
<asp:BoundField DataField="Protein_Identity" HeaderText="Protein_Identity"
        SortExpression="Protein_Identity" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Peptides_Matched" HeaderText="Peptides_Matched"
        SortExpression="Peptides_Matched" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="column2" HeaderText="Sequence Coverage"
        SortExpression="column2" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="MOWSE_score" HeaderText="MOWSE_score"
        SortExpression="MOWSE_score" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Accession_no" HeaderText="Accession_no"
        SortExpression="Accession_no" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="ID_method" HeaderText="ID_method"
        SortExpression="ID_method" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="column3" HeaderText="Protein Type (ST_AL_AN)"
        SortExpression="column3" >
<ItemStyle HorizontalAlign="Center" />
</asp:BoundField>
<asp:BoundField DataField="Expect_value" HeaderText="Expect_value"
        SortExpression="Expect_value" >
<ItemStyle HorizontalAlign="Center" />
```

```
</asp:BoundField>

<asp:BoundField DataField="Theoretical_pI" HeaderText="Theoretical_pI"

        SortExpression="Theoretical_pI" >

<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>

<asp:BoundField DataField="Theoretical_Mr" HeaderText="Theoretical_Mr"

        SortExpression="Theoretical_Mr" >

<ItemStyle HorizontalAlign="Center" />

</asp:BoundField>

</Columns>

<FooterStyle BackColor="#1C5E55" Font-Bold="True" ForeColor="White" />

<PagerStyle BackColor="#666666" ForeColor="White" HorizontalAlign="Center" />

<SelectedRowStyle BackColor="#C5BBAF" Font-Bold="True" ForeColor="#333333" />

<HeaderStyle BackColor="#1C5E55" Font-Bold="True" ForeColor="White"

    Font-Size="X-Small" />

<EditRowStyle BackColor="#7C6F57" />

<AlternatingRowStyle BackColor="White" />

<img src="pH(6-11)_M.jpg" width="900" /><br />

<asp:AccessDataSource ID="AccessDataSource1" runat="server" DataFile="App_Data/Database1.mdb"

    SelectCommand="SELECT [Spot ID] AS Spot_ID, [Theoretical pI / Mr] AS column1, [Protein Identity] AS Protein_Identity, [Peptides
Matched] AS Peptides_Matched, [% Sequence Coverage] AS column2, [MOWSE score] AS MOWSE_score, [Accession no] AS Accession_no,
[ID method] AS ID_method, [Protein Type (ST, AL, AN)] AS column3, [Expect value] AS Expect_value, [Theoretical pI] AS Theoretical_pI,
[Theoretical Mr] AS Theoretical_Mr FROM [Sheet2] WHERE ([Spot ID] = ?)">

<SelectParameters>

<asp:ControlParameter ControlID="TextBox1" Name="Spot_ID" PropertyName="Text" Type="Double" />

</SelectParameters>

</asp:AccessDataSource>

</asp:Content>
```

Soybean_low_abundance_proteins_2D_Gel_DB:

It contains information about a 2D Gel image that represents low abundance proteins of soybean.

```
<%@ Master Language="C#" AutoEventWireup="true" CodeFile="MasterPage2.master.cs" Inherits="MasterPage" %>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<html xmlns="http://www.w3.org/1999/xhtml" >

<head>

<meta http-equiv="content-type" content="text/html; charset=utf-8" />

<meta name="description" content="" />

<meta name="keywords" content="" />

<title>Soybean Seed Proteins 2D Gel DB</title>

<link href="http://fonts.googleapis.com/css?family=Abel" rel="stylesheet" type="text/css" />

<link rel="stylesheet" type="text/css" href="style.css" />

<style type="text/css">

.style1 {

font-size: medium;

}

</style>

</head>

<body>

<div id="outer">

<div id="wrapper">

<div id="menu">

<ul>

<li class="first"><a href="Home.aspx">Home</a></li>

<li><a href="Gel1.aspx" >Gel1</a></li>

<li><a href="Gel2.aspx" >Gel2</a></li>

<li><a href="Gel3.aspx" >Gel3</a></li>

</ul>

<br class="clearfix" />
```

```
</div>
<div id="header">
<div id="logo">
<h1 class="style1">
 SoyProDB: Soybean Seed Proteins 2D Gel DB</h1>
</div>
<div id="search">
<form action="" method="post">
</form>
</div>
</div>
<div id="page">
<form id="form1" runat="server">
<div>
<asp:contentplaceholder id="ContentPlaceHolder1" runat="server">
<br />
</asp:contentplaceholder>
</div>
</form>
<br class="clearfix" />
<div id="footer">
                </body>
</html>
```

Selection of Queries:

Search gene by No:

Query 1: Select gene expression of all genes given the GeneNo

SelectCommand="SELECT SUBSTRING(Transcript_Final_RPKM.GeneNo,1,15) as GeneNo,GeneDescription.Description , Sample1,Sample2, Sample3, Sample4, Sample5, Sample6, Sample7, Sample8, Sample9 FROM [Transcript_Final_RPKM] INNER JOIN GeneDescription ON SUBSTRING(Transcript_Final_RPKM.GeneNo, 1, 15) = GeneDescription.GeneNo WHERE (Transcript_Final_RPKM.GeneNo LIKE '%' + @GeneNo + '%')">

Query 2: Select gene information by chosing a specific gene

SelectCommand="SELECT Transcript_GeneLength.GeneNo, Transcript_GeneLength.Seq,Transcript_GeneLength.GeneLength, GeneDescription.Description FROM Transcript_GeneLength INNER JOIN GeneDescription ON SUBSTRING(Transcript_GeneLength.GeneNo, 1, 15) = GeneDescription.GeneNo WHERE (Transcript_GeneLength.GeneNo like '%' + @GeneNo + '%')">

Search gene by DES:

Query 1: Select gene expression of a gene given a full/partial gene description

SelectCommand="SELECT SUBSTRING(Transcript_Final_RPKM.GeneNo,1,15),GeneDescription.Description , Sample1,Sample2, Sample3, Sample4, Sample5, Sample6, Sample7, Sample8, Sample9 FROM [Transcript_Final_RPKM] INNER JOIN GeneDescription ON SUBSTRING(Transcript_Final_RPKM.GeneNo, 1, 15) = GeneDescription.GeneNo where substring (Transcript_Final_RPKM.GeneNo,1,15) in (select GeneDescription.GeneNo from GeneDescription WHERE (GeneDescription.Description LIKE N'%' + @genedesc + N'%')) ORDER BY Transcript_Final_RPKM.GeneNo ">

Query 2: Select gene information by chosing a specific gene

SelectCommand="SELECT Transcript_GeneLength.GeneNo, Transcript_GeneLength.GeneLength, Transcript_GeneLength.Seq, GeneDescription.Description FROM Transcript_GeneLength INNER JOIN GeneDescription ON SUBSTRING(Transcript_GeneLength.GeneNo, 1, 15) = GeneDescription.GeneNo WHERE (Transcript_GeneLength.GeneNo like '%' + @GeneNo + '%')">

Access data:

Query 1: Select RNA-Seq samples information from a table named Transcript_Alignment_result

SelectCommand="SELECT [SampleID], [Number_of_Reads], [aligned_0_times], [%aligned_0_times] AS column1, [aligned_exactly_1_time], [%aligned_exactly_1_time] AS column2, [aligned >1_times] AS column3, [%aligned >1_times] AS column4, [overall_alignment_rate] FROM [Transcript_Alignment_result]

Query 2: Select gene expression of all genes given the sample name

SelectCommand="SELECT Transcript_RPKM_ReadsCount.GeneNo, Transcript_RPKM_ReadsCount.ReadsCount, Transcript_RPKM_ReadsCount.SampleID, Transcript_RPKM_ReadsCount.RPKM, GeneDescription.Description FROM Transcript_RPKM_ReadsCount INNER JOIN GeneDescription ON SUBSTRING(Transcript_RPKM_ReadsCount.GeneNo, 1, 15) = GeneDescription.GeneNo WHERE (Transcript_RPKM_ReadsCount.SampleID = @SampleID)"

Query 3: Select all gene information given the gene name

SelectCommand="SELECT Transcript_GeneLength.GeneNo, GeneDescription.Description, Transcript_GeneLength.GeneLength, Transcript_GeneLength.Seq FROM Transcript_GeneLength INNER JOIN GeneDescription ON Transcript_GeneLength.GeneNo = GeneDescription.GeneNo WHERE (Transcript_GeneLength.GeneNo LIKE '%' + @GeneNo + '%')">

DEGSeq:

Query 1: Select DEGSeq results given samples, significance and fold

SelectCommand = "SELECT Transcript_DEGseq.GeneNo , Transcript_DEGseq.value1, Transcript_DEGseq.value2,Transcript_DEGseq.[log2(FC)], Transcript_DEGseq.[log], Transcript_DEGseq.z_score, Transcript_DEGseq.p_value,Transcript_DEGseq.q_value_Benjamini, Transcript_DEGseq.q_value_Storey, Transcript_DEGseq.Signature,Transcript_DEGseq.SampleID, Transcript_DEGseq.Fold_Change, Transcript_DEGseq.significant,GeneDescription.GeneNo AS Expr1, GeneDescription.Description,GeneDescription.Description AS Expr2 FROM Transcript_DEGseq INNER JOIN GeneDescription ON SUBSTRING(Transcript_DEGseq.GeneNo, 1, 15) =GeneDescription.GeneNo WHERE (Transcript_DEGseq.SampleID = @SampleID2) AND(ABS(Transcript_DEGseq.Fold_Change) >= @Fold_Change) AND (Transcript_DEGseq.significant= @significant) OR (Transcript_DEGseq.SampleID = @SampleID2) AND(ABS(Transcript_DEGseq.Fold_Change) >= @Fold_Change) AND ('0' = @significant)ORDER BY Transcript_DEGseq.Fold_Change DESC">

Query 2: Select gene information by chosing a specific gene

SelectCommand="SELECT [GeneNo], [value1], [value2], [log2(FC)] AS column1, [log], [z_score], [p_value], [q_value_Benjamini], [q_value_Storey], [Signature], [SampleID], [Fold_Change], [significant] FROM [Transcript_DEGseq] WHERE (([GeneNo] = @GeneNo) AND ([SampleID] = @SampleID))">

DESeq:

Query 1: Select DESeq results given samples, significance and fold

SelectCommand = "SELECT [Soybean].[dbo].[Transcript_DESeq]. [GeneNo]

,[Soybean].[dbo].[Transcript_DESeq].[baseMean]

,[Soybean].[dbo].[Transcript_DESeq].[baseMeanA]

,[Soybean].[dbo].[Transcript_DESeq].[baseMeanB]

,[Soybean].[dbo].[Transcript_DESeq].[Fold_Change]

,[Soybean].[dbo].[Transcript_DESeq].[log2FoldChange]

,[Soybean].[dbo].[Transcript_DESeq].[pval]

,[Soybean].[dbo].[Transcript_DESeq].[padj]

,[Soybean].[dbo].[Transcript_DESeq].[SampleID]

,[Soybean].[dbo].[Transcript_DESeq].[significant]

,GeneDescription.Description

FROM [Soybean].[dbo].[Transcript_DESeq] INNER JOIN GeneDescription ON SUBSTRING([Soybean].[dbo].[Transcript_DESeq].GeneNo, 1, 15) =GeneDescription.GeneNo WHERE ([Soybean].[dbo].[Transcript_DESeq].SampleID = @SampleID2) AND(ABS([Soybean].[dbo].[Transcript_DESeq].Fold_Change) >= @Fold_Change) AND ([Soybean].[dbo].[Transcript_DESeq].significant=

@significant) OR ([Soybean].[dbo].[Transcript_DESeq].SampleID = @SampleID2)
AND(ABS([Soybean].[dbo].[Transcript_DESeq].Fold_Change) >= @Fold_Change) AND ('0' = @significant)ORDER BY
[Soybean].[dbo].[Transcript_DESeq].Fold_Change DESC">

Query 2: Select gene information by chosing a specific gene

SelectCommand="SELECT [GeneNo] ,[baseMean] ,[baseMeanA] ,[baseMeanB] ,[Fold_Change] ,[log2FoldChange]

,[pval] , [padj] ,[SampleID] ,[significant]

FROM [Soybean].[dbo].[Transcript_DESeq] WHERE (([GeneNo] = @GeneNo) AND ([SampleID] = @SampleID))">

edgeR:

Query 1: Select edgR results given samples, significance and fold

SelectCommand = "SELECT GeneDescription.GeneNo AS Expr1, GeneDescription.Description, GeneDescription.Description AS Expr2,
Transcript_EdgeR.GeneNo, Transcript_EdgeR.logFC, Transcript_EdgeR.logCPM, Transcript_EdgeR.PValue, Transcript_EdgeR.FDR,
Transcript_EdgeR.SampleID, Transcript_EdgeR.Significant FROM GeneDescription INNER JOIN Transcript_EdgeR ON
GeneDescription.GeneNo = Transcript_EdgeR.GeneNo WHERE (Transcript_EdgeR.SampleID = @SampleID2) AND
(ABS(Transcript_EdgeR.logFC) >= @Fold_Change) AND (Transcript_EdgeR.Significant = @significant) OR ('0' = @significant) AND
(Transcript_EdgeR.SampleID = @SampleID2) AND (ABS(Transcript_EdgeR.logFC) >= @Fold_Change) ORDER BY
Transcript_EdgeR.logFC DESC">

Query 2: Select gene information by chosing a specific gene

SelectCommand="SELECT [GeneNo], [value1], [value2], [log2(FC)] AS column1, [log], [z_score], [p_value], [q_value_Benjamini],
[q_value_Storey], [Signature], [SampleID], [Fold_Change], [significant] FROM [Transcript_DEGseq] WHERE (([GeneNo] = @GeneNo) AND
([SampleID] = @SampleID))">

NOISeq:

Query 1: Select NOISeq results given samples, significance and fold

SelectCommand = "SELECT [Soybean].[dbo].[Transcript_NOISeq_2013].[GeneNo]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[Control_mean]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[SA_mean]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[M]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[D]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[prob]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[ranking]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[SampleID]

,[Soybean].[dbo].[Transcript_NOISeq_2013].[Significance]

,GeneDescription.Description

FROM [Soybean].[dbo].[Transcript_NOISeq_2013] INNER JOIN GeneDescription ON SUBSTRING([Soybean].[dbo].[Transcript_NOISeq_2013].GeneNo, 1, 15) =GeneDescription.GeneNo WHERE ([Soybean].[dbo].[Transcript_NOISeq_2013].SampleID = @SampleID2) AND(ABS([Soybean].[dbo].[Transcript_NOISeq_2013].[M]) >= @Fold_Change) AND ([Soybean].[dbo].[Transcript_NOISeq_2013].[Significance]= @significant) OR ([Soybean].[dbo].[Transcript_NOISeq_2013].SampleID = @SampleID2) AND(ABS([Soybean].[dbo].[Transcript_NOISeq_2013].[M]) >= @Fold_Change) AND ('0' = @significant)ORDER BY [Soybean].[dbo].[Transcript_NOISeq_2013].[M] DESC">

Query 2: Select gene information by chosing a specific gene

SelectCommand="SELECT [GeneNo] ,[baseMean] ,[baseMeanA] ,[baseMeanB] ,[Fold_Change] ,[log2FoldChange]

,[pval] , [padj] ,[SampleID] ,[significant]

FROM [Soybean].[dbo].[Transcript_DESeq] WHERE (([GeneNo] = @GeneNo) AND ([SampleID] = @SampleID))">

Cuffdiff:

Query 1: Select Cuffdiff results given samples, significance and fold

SelectCommand = "SELECT SUBSTRING([Soybean].[dbo].[Transcript_Cuffdiff].GeneNo, 1, 15) AS GeneNo

,[value_1], [value_2], [log2(fold_change), [test_stat] , [p_value] , [q_value] , [sign] , [SampleID], [fold_change] , [significant]

,GeneDescription.Description FROM [Soybean].[dbo].[Transcript_Cuffdiff]INNER JOIN GeneDescription ON SUBSTRING([Soybean].[dbo].[Transcript_Cuffdiff].GeneNo, 1, 15) =GeneDescription.GeneNo WHERE ([Soybean].[dbo].[Transcript_Cuffdiff].SampleID = @SampleID2) AND(ABS([Soybean].[dbo].[Transcript_Cuffdiff].fold_change) >= @Fold_Change) AND ([Soybean].[dbo].[Transcript_Cuffdiff].significant= @significant) OR ([Soybean].[dbo].[Transcript_Cuffdiff].SampleID = @SampleID2) AND(ABS([Soybean].[dbo].[Transcript_Cuffdiff].fold_change) >= @Fold_Change) AND ('0' = @significant)ORDER BY [Soybean].[dbo].[Transcript_Cuffdiff].fold_change DESC">

Query 2: Select gene information by chosing a specific gene

SelectCommand="SELECT [GeneNo] ,[baseMean] ,[baseMeanA] ,[baseMeanB] ,[Fold_Change] ,[log2FoldChange]

,[pval] , [padj] ,[SampleID] ,[significant]

FROM [Soybean].[dbo].[Transcript_DESeq] WHERE (([GeneNo] = @GeneNo) AND ([SampleID] = @SampleID))">

Bibliography

1. Hajduch M et al. Plant Physiol. 2005 137: [PMID: 15824287]
2. Ohyanagi H et al. Front Plant Sci. 2012 3: [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]
11. A. Santner, M. Estelle, Recent advances and emerging trends in plant hormone signalling, Nature 459 (7250) (2009) 1071-1078.
12. V. N. Matiru, F. D. Dakora, The rhizosphere signal molecule lumichrome alters seedling development in both legumes and cereals, New Phytolo-gist 166 (2) (2005) 439-444.
13. G. Loake, M. Grant, Salicylic acid in plant defencethe players and pro-tagonists, Current Opinion in Plant Biology 10 (5) (2007) 466-472.

14. H. Gundlach, M. J. Muller, T. M. Kutchan, M. H. Zenk, Jasmonic acid is a signal transducer in elicitor-induced plant cell cultures, Proceedings of the National Academy of Sciences 89 (6) (1992) 2389-2393.

15. O. Lorenzo, R. Piqueras, J. J. Sanchez-Serrano, R. Solano, ETHYLENE RESPONSE FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense, The Plant Cell Online 15 (1) (2003) 165-178.

16. C. M. Pieterse, A. Leon-Reyes, S. Van der Ent, S. C. Van Wees, Net-working by small-molecule hormones in plant immunity, Nature Chem-ical Biology 5 (5) (2009) 308-316.

17. D. M. Goodstein, S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, D. S. Rokhsar, Phyto-zome: a comparative platform for green plant genomics., Nucleic Acids Research 40 (Database issue) (2012) D1178-D1186.

18. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform., Bioinformatics 25 (14) (2009) 1754-1760.

19. S. Anders, W. Huber, Di erential expression analysis for sequence count data., Genome Biology 11 (10) (2010) R106.

20. Z. Du, X. Zhou, Y. Ling, Z. Zhang, Z. Su, agriGO: a GO analysis toolkit for the agricultural community, Nucleic Acids Research 38 (Web-Server- Issue) (2010) 64-70.

21. N. O. Ste ens, C. Galuschka, M. Schindler, L. Bulow, R. Hehl, AthaMap: an online resource for in silico transcription factor bind-ing sites in the Arabidopsis thaliana genome, Nucleic Acids Research 32 (Suppl 1) (2004) D368-D372.

22. Hosseini, P. and Ovcharenko, I. and Matthews, B. F., Using an ensemble of statistical metrics to quantify large sets of plant transcription factor binding sites, Plant Methods 9 (1) (2013) 12.

23. J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for general-ized linear models via coordinate descent, Journal of Statistical Software 33 (1) (2010) 1.

24. J. Sun, V. Cardoza, D. M. Mitchell, L. Bright, G. Oldroyd, J. M. Har-ris, Crosstalk between jasmonic acid, ethylene and Nod factor signaling allows integration of diverse inputs for regulation of nodulation, The Plant Journal 46 (6) (2006) 961-970.

25. L. C. van Loon, B. P. Geraats, H. Linthorst, Ethylene as a modulator of disease resistance in plants, Trends in Plant Science 11 (4) (2006) 184-191.

26. Thomma, B. P.H.J. and Penninckx, I. A.M.A. and Cammue, B. and Broekaert, W. F., The complexity of disease signaling in Arabidopsis, Current Opinion in Immunology 13 (1) (2001) 63-68.

27. N. A. Eckardt, Oxylipin signaling in plant stress responses, The Plant Cell Online 20 (3) (2008) 495-497.

28. P. E. Staswick, I. Tiryaki, The oxylipin signal jasmonic acid is activated by an enzyme that conjugates it to isoleucine in Arabidopsis, The Plant Cell Online 16 (8) (2004) 2117-2127.

29. G. A. Howe, A. L. Schilmiller, Oxylipin metabolism in response to stress, Current Opinion in Plant Biology 5 (3) (2002) 230-236.

30. E. Bartholomeus Kuettner, R. Hilgenfeld, M. S. Weiss, Puri cation, characterization, and crystallization of alliinase from garlic, Archives of Biochemistry and Biophysics 402 (2) (2002) 192-200.

31. M. Jia, H. Wu, K. L. Clay, R. Jung, B. A. Larkins, B. C. Gibbon, Iden-ti cation and characterization of lysine-rich proteins and starch biosyn-thesis genes in the opaque2 mutant by transcriptional and proteomic analysis, BMC Plant Biology 13 (1) (2013) 60.

32. C.-Y. Yang, F.-C. Hsu, J.-P. Li, N.-N. Wang, M.-C. Shih, The AP2/ERF transcription factor AtERF73/HRE1 modulates ethylene responses dur-ing hypoxia in Arabidopsis, Plant Physiology 156 (1) (2011) 202-212.

33. M. Pre, M. Atallah, A. Champion, M. De Vos, C. M. Pieterse, J. Memelink, The AP2/ERF domain transcription factor ORA59 inte-grates jasmonic acid and ethylene signals in plant defense, Plant Physi-ology 147 (3) (2008) 1347-1357.

34. T. Vogt, Phenylpropanoid biosynthesis, Molecular Plant 3 (1) (2010)

35. D. Adams, S. Yang, Ethylene biosynthesis: identi cation of 1- aminocyclopropane-1-carboxylic acid as an intermediate in the conver-sion of methionine to ethylene, Proceedings of the National Academy of Sciences 76 (1) (1979) 170-174.

36. A. B. Bleecker, H. Kende, Ethylene: a gaseous signal molecule in plants, Annual Review of Cell and Developmental Biology 16 (1) (2000) 1-18.

37. Z.-B. Liu, T. Ulmasov, X. Shi, G. Hagen, T. J. Guilfoyle, Soybean GH3 promoter contains multiple auxin-inducible elements., The Plant Cell Online 6 (5) (1994) 645-657.

38. L. Van Der Fits, J. Memelink, The jasmonate-inducible AP2/ERF-domain transcription factor ORCA3 activates gene expression via inter-action with a jasmonate-responsive promoter element, The Plant Jour-nal 25 (1) (2001) 43-53.

39. N. Gutterson, T. L. Reuber, Regulation of disease resistance pathways by AP2/ERF transcription factors, Current Opinion in Plant Biology 7 (4) (2004) 465-471.

40. Z.-S. Xu, M. Chen, L.-C. Li, Y.-Z. Ma, Functions and Application of the AP2/ERF Transcription Factor Family in Crop Improvement, Journal of Integrative Plant Biology 53 (7) (2011) 570-585.

41. J. Mizoi, K. Shinozaki, K. Yamaguchi-Shinozaki, AP2/ERF family tran-scription factors in plant abiotic stress responses, Biochimica et Biophys-ica Acta 1819 (2) (2012) 86-96.

42. K. H. Sohn, S. C. Lee, H. W. Jung, J. K. Hong, B. K. Hwang, Expres-sion and functional roles of the pepper pathogen-induced transcription factor RAV1 in bacterial disease resistance, and drought and salt stress tolerance, Plant Molecular Biology 61 (6) (2006) 897-915.

43. M. Jakoby, B. Weisshaar, W. Droge-Laser, J. Vicente-Carbajosa, J. Tiedemann, T. Kroj, F. Parcy, bzip transcription factors in Ara-bidopsis, Trends in Plant Science 7 (3) (2002) 106-111.

44. S. C. Lee, H. W. Choi, I. S. Hwang, B. K. Hwang, Functional roles of the pepper pathogen-induced bZIP transcription factor, CAbZIP1, in enhanced resistance to pathogen infection and environmental stresses, Planta 224 (5) (2006) 1209-1225.

45. N. Sato, N. Ohta, DNA-binding speci city and dimerization of the DNA-binding domain of the PEND protein in the chloroplast envelope membrane, Nucleic Acids Research 29 (11) (2001) 2244-2250.

46. T. Eulgem, P. J. Rushton, S. Robatzek, I. E. Somssich, The WRKY superfamily of plant transcription factors, Trends in Plant Science 5 (5) (2000) 199-206.

47. Wang L, Feng Z, Wang X, Wang X, and Zhang X (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics, 26(1):136–138

48. K. Hara, M. Yagi, T. Kusano, H. Sano, Rapid systemic accumulation of transcripts encoding a tobacco WRKY transcription factor upon wound-ing, Molecular and General Genetics 263 (1) (2000) 30-37.

49. Tremblay, A., Hosseini, P., Li, Sh., Alkharouf, NW. and Matthews, B. Analysis of Phakopsora pachyrhizi transcript abundance in critical pathways at four time-points during infection of a susceptible soybean cultivar using deep sequencing.BMC Genomics 2013.

50. Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van

Baren, M. J.,Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology, 28(5), 511-515.

51. Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B.,

Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome research, 18(1), 188-196.

52. Barakat, A., DiLoreto, D. S., Zhang, Y., Smith, C., Baier, K., Powell, W.

Carlson, J. E. (2009). Comparison of the transcriptomes of American chestnut (Castanea dentata) and Chinese chestnut (Castanea mollissima) in response to the chestnut blight infection. BMC Plant Biology, 9(1), 51.

53. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B.

(2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq

Nature methods, 5(7), 621-628.

Curriculum Vitae

Mona Tavakolan

[redacted]

Education:

- **Doctor of Science in Information Technology:**
  Towson University, Towson, MD, USA

  Date: January 2011- July 2014

  GPA: 4.0

  Thesis title: Bioinformatics analysis of data and development of tools for high throughput biological experiments

- **M.Sc. in Information Technology:**
  Iran University of Science and Technology, Tehran, Iran

  Date: January 2006-March 2009

  Thesis title: Designing a web service with capabilities to apply user preferences based on privacy

- **B. Sc in Computer Science:**
  Mazandaran University of Science and Technology, Babol, Iran

  Date: September 2001-September 2005

  Thesis title: Simulation tool for Color Petri Net

Teaching Experience:

Instructor:

- Telecommunications (CIS 350), Towson University, Towson, MD, USA
  - Spring 2014
- Computers & Creativity (COSC 109) Towson University, Towson, MD, USA
  - Spring 2013
  - Fall 2013
  - Spring 2014 (2 Sections)

Responsibility:

- Provided general support and one-to-one assistance for students
- Created and managed exams, assignments and labs
- Supported students individually outside class and during off-classroom hours
- Provided individualized care to students with special needs
- Created learning environments that stimulate thinking and motivate students

Teaching Assistant:

- Principles of Computer Organization (COSC 290), Towson University, Towson, MD, USA
  - Fall 2011
  - Spring 2012
  - Fall 2012
- Fundamentals of Computer Networks, IUST, Tehran, Iran
  - Spring 2006 to Fall 2008

Responsibility:

- Managed the labs
- Graded all the assignments, exams and labs
- Monitored student's learning
- Made records of student's progress
- Met with students to solve their problems

Research Experience:

Research Assistant at Towson University, Towson, MD, USA (January 2011 - Present)

- Bioinformatics analysis of data and development of tools for high throughput biological experiments

- Producing databases for the identification of soybean seed proteins
- Analysis of next generation sequence data related to soybean treated with different hormones
- Comparison of statistical tools used in RNA-Seq analysis
- Updating the soybean genomics database to include next gen sequencing data

Research Assistant at IUST, Tehran, Iran (January 2006 – March 2009)

- Designed and implemented virtual university using Rational Rose2005
- Designed and implemented electronic auction using Rational Rose2005
- Investigated and implemented windows fail over clustering
- Conducted a survey on personalization E-commerce system which resulted in a new idea for their recommendations system
- Investigated on distributed shared memory
- Data mining on hygienic data to establish associations rules, clusters and classifications using Clementine tool
- Designed a web service with capabilities to apply user preferences based on privacy

Work Experience:

Company name: Pars Asan Afzar Company

Job Title: Systems Analyst (September 2006- September 2007)

Responsibility:

- Analyzed the company's software projects
- Gathered system information and interviewed system users to identify system use cases and actors
- Utilized Rational Rose in order to define system use cases, actors, related scenarios and existing relationship
- Designed system classes and sequence diagrams
- Coordinated the installation of computer programs and systems
- Identified software defects and solved technical problems
- Tested and maintained company systems

Company name: Pars Asan Afzar Company

Job Title: Database Administrator (September 2007-September 2008)

Responsibility:

- Designed, implemented, normalized, documented and maintained databases
- Created required transactions

- Defined users and their access levels to different database objects
- Implemented different database views as required
- Created and used data bases utilizing SQL server for software

Company name: Pahneh Negar Tehran Company

Job Title: Data Mining and R&D (September 2008-July 2009)

Responsibility:

- Getting user opinion about program interfaces and web site design
- Categorizing data to train and test and run data mining algorithms on train data and find their association rules, classification and clustering. Then run the same algorithm on test data to evaluate

Skills and Knowledge:

- **Bioinformatics Packages:** DEGseq, DESeq, NOISeq, edgeR, cuffdiff, FASTX-Toolkit, Bowtie
- **Network Packages:** Ns-2 (network simulator), OPNET, Wireshark, Active Directory
- **Programming Languages:** C, C#, Java, MATLAB, Assembly, HTML, ASP.NET, MySQL, Microsoft Sql Server
- **Operating Systems:** Linux, Unix, Windows
- **Other Software:** Rational Rose, Clementine, Microsoft Office (Word, Excel, Access, PowerPoint, Visio), Photoshop, Flash, Microsoft Project Management.

Publications:

- "SoyProDB: A database for the identification of soybean seed proteins", **Tavakolan M**, Alkharouf NW, Khan FH, Natarajan S. Bioinformation 2013.

- "SCNProDB: A database for the identification of cyst nematode proteins using 2D Gels", **Tavakolana M**, Alkharouf NW, Mathews B and Natarajanb S. Bioinformation 2014.

- "Identification of low abundance Soybean proteins by 2D Gels", **Tavakolan M**, Alkharouf NW, and Natarajanb S. In preparation for Bioinformation 2014.

- "SoyProDB: A database for analysis of next generation sequence data related to soybean treated with different hormones", **Tavakolan M**, Alkharouf NW. In preparation for Bioinformation 2014.

- "An Extensible Model for Improving the Privacy of Web Services", **Tavakolan M**, Zarreh M and Azgomi M. Proc. of the Fifth International Conference on Innovations in

Information Technology (Innovations'08), Al Ain, Emirates, Dec. 16-18, IEEE CS Press 175-179

- "Web Service Discovery Based on Privacy Preferences", **Tavakolan M**, Zarreh M and Azgomi M. International Journal of Web Services Practices (IJWSP), Vol.4, ISSN 1738-6535

Thesis:

- M.S. Thesis: Designing a web service with capabilities to apply user preferences based on privacy at Iran University of Science and Technology, March 2009

- B.Sc. Thesis: Simulation tool for Color Petri Net at Mazandaran University of Science and Technology, September 2005