12-13-2014

Computational Methods on Study of Differentially Expressed Proteins in Maize Proteomes Associated with Resistance to Aflatoxin Accumulation

Alka Tiwari

Computational methods on study of differentially expressed proteins in maize proteomes associated with resistance to aflatoxin accumulation

By

Alka Tiwari

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Molecular Biology
in the Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology

Mississippi State, Mississippi

December 2014

Computational methods on study of differentially expressed proteins in maize proteomes associated with resistance to aflatoxin accumulation

By

Alka Tiwari

Approved:

_____

Xueyan Shan
(Major Professor)

_____

Din-Pow Ma
(Co-Major Professor/Graduate Coordinator)

_____

W. Paul Williams
(Committee Member)

_____

Ashli Brown
(Committee Member)

_____

Andy D. Perkins
(Committee Member)

_____

George M. Hopper
Dean
College of Agriculture and Life Sciences

Name: Alka Tiwari

Date of Degree: December 13, 2014

Institution: Mississippi State University

Major Field: Molecular Biology

Major Professor: Xueyan Shan

Title of Study: Computational methods on study of differentially expressed proteins in maize proteomes associated with resistance to aflatoxin accumulation

Pages in Study: 106

Candidate for Degree of Doctor of Philosophy

Plant breeders have focused on improving maize resistance to *Aspergillus flavus* infection and aflatoxin accumulation by breeding with genotypes having the desirable traits. Various maize inbred lines have been developed for the breeding of resistance. Identification of differentially expressed proteins among such maize inbred lines will facilitate the development of gene markers and expedite the breeding process. Computational biology and proteomics approaches on the investigation of differentially expressed proteins were explored in this research. The major research objectives included 1) application of computational methods in homology and comparative modeling to study 3D protein structures and identify single nucleotide polymorphisms (SNPs) involved in changes of protein structures and functions, which can in turn increase the efficiency of the development of DNA markers; 2) investigation of methods on total protein profiling including purification, separation, visualization, and computational analysis at the proteome level. Special research goals were set on the development of open source

computational methods using Matlab image processing tools to quantify and compare protein expression levels visualized by 2D protein electrophoresis gel techniques.

DEDICATION

I would like to dedicate this research work to my parents Dr. Ram Krishan Tiwari and Mrs. Kiran Tiwari, my in-laws Mr. Yogendra Mani Tripathi and Mrs. (Late) Girija Mani Tripathi, my sibs Dr. Alok Tiwari and Dr. Ila Tiwari, and my husband Dr. Markandey Mani Tripathi.

ACKNOWLEDGEMENTS

I would like to acknowledge many people who helped me during my Ph.D. study. First, I would like to show my deepest appreciation to my advisor Dr. Xueyan Shan for her immense support in my research work. Dr. Shan has been a wonderful mentor and a very nice person. During my Ph.D. study, she was always ready to listen to my problems. She motivated me to develop problem solving approaches, accepted my ideas, and gave me confidence to incorporate those ideas. She encouraged me to develop independent problem solving skills throughout my Ph. D. study. Without her guidance and persistent help this dissertation would not have been possible.

I would like to thank my committee members, Dr. Din-Pow Ma, Dr. W. Paul Williams, Dr. Ashli Brown, and Dr. Andy D. Perkins for their enormous support. My thanks to Dr. Williams for providing research materials and research funds. My thanks to Dr. Ma, Dr. Perkins and Dr. Brown for helping me to understand Biochemistry, Molecular Biology, and Basics of Computational Biology in depth through my course study, which ultimately helped me to use those fundamentals in my research work.

I would like to thank Dr. Tibor Pechan for providing me the imager to take DIGE gels, which were used in one of my projects. Dr. Tibor Pechan is a very nice and knowledgeable person, who taught me the details to take good gel pictures. Also, I would like to thank Shannon Kate Thompson and Taylor Clark for helping me in the field work as well as in the lab.

Lastly I would like to thank my family. My parents and my sibs constantly inspired me to pursue my interest in life. A special thanks to my Mother, for believing me and giving me confidence to fulfill all my endeavors in life. I am especially grateful to my dear husband, Dr. Markandey Mani Tripathi, for his support, patience and encouragement to fulfill all my dreams.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

LITERATURE REVIEW

Maize: a real triumph of plant breeding

Plant domestication is one of the most wonderful events that happened in the past 13000 years of human civilization which has provided food and feed for humans and animals (Diamond, 2002). World's major crops such as maize (*Zea mays*), rice (*Oryza sativa*) and wheat (*Triticum aestivum*) are the greatest discoveries of human history. Agricultural development not only has provided staple food to individuals but also united the world together. Maize, one of the world's leading crop for food and fuel, was first domesticated at Mexico's Central Balsas River valley (Timms and Cramer, 2008; Tenaillon and Charcosset, 2011). Cultivation of maize supported millions of population worldwide and also helped in cultural development.

Maize is a monoecious plant that belongs to family Gramineae and tribe Maydeae. Gramineae is also known as Poaceae. Maize has 10 chromosomes and its genetic configuration is $2n = 2x = 20$. It is an annual plant that grows rapidly and reaches an average height of 2.5 meters at maturity. The optimum temperature for its growth is between 10ºC - 30ºC. Maize has the broadest geographical cultivation range of all crops and has been cultivated from Chile to Canada (Tenaillon and Charcosset, 2011). Maize also has the greatest morphological and genetic diversity compared to other crops, which has provided the fundamental basis for the breeding and the improvement of

agronomically important traits (Gauci et al., 2011). Maize is one of the important food staples worldwide, most prominent in sub-Saharan-Africa, India, China, and United States of America (Blackie et al., 1987; Henry, 2013). Maize is used in the production of a variety of food products (such as corn oil and corn syrup) as well as industrial products (such as ethanol). It has enormous food and industrial utilizations.

Many approaches have been employed by maize breeders and researchers to improve maize production by the enhancement of the agronomic traits for better yield and disease resistance. Plant breeders have focused on improving maize by breeding with genotypes having the desirable traits. Various maize resistant lines have been developed in recent decades for the breeding purposes to enhance maize yield. Molecular mapping techniques have also applied to identify regions that are responsible for maize disease resistance. Genetic engineering techniques have also been employed to transform susceptible maize inbred lines to make them resistant against pathogenic attacks.

Aspergillus flavus and Aflatoxins

Aflatoxins are toxic metabolites produced by fungi *Aspergillus flavus* and *Aspergillus parasiticus* (Efrat et al., 2002). While occurring worldwide on a number of agricultural commodities, *A. flavus* appears to be mostly associated with maize, peanut and cotton (Rodriguez-del-Bosque, 1996). *A. flavus* is a saprophytic and an opportunistic fungus which can proliferate in a variety of environments. Life cycle of this saprophytic fungus begins in soil where it can reside as mycelia or sclerotia stages. In the favorable environmental conditions, especially in dry and high heat weather, conidiophores produce conidia that spread through air or water. Plants that are infected by these conidia become good reservoirs for further dispersal of *A. flavus* to other plants (Wicklow and Donahue,

1984). Conidia that land on plant surface can serve as inoculums for secondary infections and consequently complete several life cycles in a single season, causing rot diseases and producing aflatoxins in infected plants (Dorner, 2004; Dorner and Abbas, 2005; Yin et al., 2008).

Aflatoxins are polyketide-derived toxins which are a class of secondary metabolites and carcinogenic to humans and animals. Major types of aflatoxins include AFB1, AFB2, AFG1, AFG2, AFM1, and AFM2 (Dhanasekaran et al., 2011; Tajkarimi et al., 2011). Aflatoxin B1 is classified as class I carcinogen and can severely affect the health of humans and animals. The carcinogenic effect of aflatoxin B1 is enhanced through the cytochrome p450 enzymatic systems in liver. Cytochrome p450 enzymes oxidize aflatoxin B1 to produce aflatoxin B1-8,9 epoxide (Kensler et al., 2011). The toxicity of aflatoxin B1-8, 9 epoxide is more severe than aflatoxin B1. It can conjugate with DNA, RNA, and proteins and cause carcinogenic effects. Aflatoxin B1 can also be converted to aflatoxin M1 which presents in milk and is a potent contaminant in dairy products.

Food highly contaminated with aflatoxins can cause acute health problems such as liver cirrhosis. Chronic consumption of sub-lethal concentration of aflatoxins can cause immunological problems, hemorrhage, acute liver damage, and liver cancer. Growth impairment has been seen among children due to the aflatoxin exposure at their early age and these children became vulnerable to infectious diseases (Ricci et al., 2006). Aflatoxin contamination not only affects humans but it also poses serious health threats to animals. Common health issues reported in animals include reduced feed intake and

weight loss (mule ducklings, Japanese quail and turkeys) (Giambrone et al., 1985; Sadana et al., 1992; Cheng et al., 2001).

Diversity of *A. flavus*

A. flavus exhibits great phenotypic diversity within species which can be divided into L and S strains, respectively. Although significantly differing in the aflatoxin synthesis ability, their natural coexistence is found in various habitats. S strain produces more aflatoxin compared with L strain. S strain produces smaller sclerotia and less conidia while L strain produces larger sclerotia and more conidia (Cotty, 1989). L strain is also known as the typical isolate of *A. flavus*. Aflatoxin producing fungi can also be categorized into the toxigenic category which produces more than 10^6 ppb in susceptible plants and the atoxigenic category (Probst and Cotty, 2012). S strain does not contain any atoxigenic isolates and is more stable than L strain. The genetic basis of atoxigenic isolates may involve single nucleotide polymorphisms (SNPs) or large deletions in aflatoxin biosynthesis genes (Donner et al., 2010). It has been reported that a SNP in a polyketide pathway gene resulted in an atoxigenic isolate AF36 (Ehrlich et al., 2007). The exact mechanism of the loss of aflatoxin production in atoxigenic isolates is still unknown.

Maize host resistance to *A. flavus* infection and aflatoxin accumulation

Reducing aflatoxin contamination in maize grains can be achieved to some extent by improving cultural practices as well as storage facilities of harvested maize grains (Rodriguez-del-Bosque, 1996; Chulze, 2010). Cultural practices include tillage, fertilization, irrigation, insect control, and planting date management (Wagacha and

Muthomi, 2008). These methods are very helpful in reducing aflatoxin accumulation in maize grains. However, the most effective strategy is to develop resistant maize lines for reduced aflatoxin accumulations (Williams, 2006). Several resistant maize inbred lines, such as Mp313E, Mp420, Mp715, Mp718 and Mp719, have been developed over the years and exhibit reduced aflatoxin accumulation in Mississippi (Windham and Williams, 2002; Boykov and Funka-Lea, 2006; Williams and Windham, 2012). These resistant lines were developing through multiple generations of self-pollination and aflatoxin screening. Two maize lines Mp718 and Mp719 were recently released having improved agronomic traits too. These two lines were generated by cross between parental line Mp715 and Va35 (Williams and Windham, 2012).

Homology Modeling

Protein 3D structural determination is a very crucial step to understand protein functions and interactions with other molecules. Enormous amount of protein structures have been resolved by the structural biologist in recent years. Experimental as well as computational methods have been developed to determine unique protein structures. Recently, protein sequence data have grown at a much greater speed in comparison to their structures due to the extensive genome sequencing projects. For this reason, it has become difficult to experimentally determine the structures of all available proteins. The gap between protein sequences and structures can be filled by comparative modeling methods based on homology among protein sequences (Larsson et al., 2008).

Basic principles behind comparative modeling or homology based methods are the alignment of target with template sequences. It is believed that proteins related evolutionarily often show structural similarities which can help to build an unknown

(target) structure of a protein using sequence alignment of a known (template) protein structure. Comparative modeling is one of the accurate methods available for protein modeling studies and plays important role in bridging the gap between protein sequences and their structures. The success of homology modeling is dependent upon the similarity between target and template sequence and their optimal alignment. Modeller is the software program most commonly used for this purpose in the area of homology modeling.

Modeller: A tool in three dimensional structure prediction of proteins

Modeller is a very important tool in the area of homology modeling. It is an easily accessible program which is primarily based on the alignment of target and template protein sequences (Sali and Blundell, 1993). Sequence similarity between target and template plays very important role in model design. Sequence identity, if it is less than 40%, can pose a great problem in alignment and ultimately generates an inaccurate protein structure (Fiser, 2004). Model inaccuracies can also be generated by structural divergence between target and template sequences, errors at the time of modeling, and inaccurate modeling of loops and side chain amino acids (Haas et al., 2013). In spite of these problems, high quality models can be treated in a similar way as an experimental structure and low quality models can still be used to get lower resolution information. For example, low resolution models can help to understand the role of a protein in mutagenesis experiment and understanding yeast ribosomal assembly (Baker and Sali, 2001; Haas et al., 2013).

Like comparative modeling approach, the de novo method was also employed for protein structure prediction if any experimentally determined 3D structure is not

available. De novo methods are based on identifying two key components: native state of a protein and their tertiary structure having minimum free energy state. Several advancements have been performed to develop de novo methods in recent years, e.g. Robetta (Kim et al., 2004). The accuracy of these methods is lower in compare to comparative modeling approaches. In recent studies, it has been observed that 58% protein structures are modeled with Modeller and Modpipe and deposited in protein data bases (Baker and Sali, 2001; Eswar et al., 2003). The role of Modeller program is increasing enormously in the area of three dimensional protein structure prediction and accurate design of protein models. This can enhance the scope of protein structure prediction and better understanding of their functions.

Proteomics a general view

The term proteomics refers to the study of all the proteins expressed in a cell at a particular time under a certain condition (Anderson et al., 2000). The investigation methods for total protein profiling include their purification, separation, visualization, and computational analysis at proteome level. Initially due to limitations of available analytical techniques, only a limited number of proteins could be analyzed in a single experimental run. More advanced methods have been developed at the urge of emergence of proteomics as an area of research. DNA sequence only provides basic information of all the possible genes a cell might use. In actual sense cells live in dynamic environments which are constantly changing due to the changes of environmental factors. Proteomics become an important qualitative and quantitative analysis tool in studying the total protein expression in various cell and tissue levels. Proteomic studies can answer questions beyond the gene level. It has been noticed that many of the genetic studies

could not match with proteomic studies because of complex regulatory mechanisms controlling the expression of proteins from genes. Often it has been seen that even mRNA expression cannot be matched with protein expression. Hence, the need for an efficient proteomic tool becomes necessary to decipher the functions of genes and proteins in various biological processes.

Proteomics methods include gel-based methods such as two dimensional gel electrophoresis (2DE) and non-gel-based techniques such as mass spectrometry (MS). 2DE protein gel technique is the most popular and traditional gel-based method of studying proteins. This technique was discovered by O'Farrell in 1975 and, it was able to identify 1100 proteins in *E. coli* (Cleveland et al., 2003). Continuous development and optimization have been achieved in the 2DE technique and thousands of proteins can be resolved in combination with the use of advanced chromatographic and centrifugation techniques (Nielsen et al., 2005). 2DE protein gel technique has been improved by the involvement of 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. A modified version of 2DE protein gel technique is the two dimensional difference in gel electrophoresis (2D-DIGE) technique. 2D-DIGE technique allows two samples per gel, and hence provides direct comparisons between two samples for quantitative analysis in protein expression. Nowadays this technique become a popular tool in clinical proteomics to study mechanism of disease and ultimately identify the biomarkers. MS-based techniques such as matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) or LC/electrospray ionization (ESI) ion trap are routinely used techniques nowadays for protein sequence identification (Nägele et al., 2004).

Two dimensional gel electrophoresis: An epitome of proteomics

2DE protein gel has always been a powerful tool in the area of proteomics with the capacity of revealing thousands of protein spots in a single gel. 2DE gel technique has been employed for decades and has been advanced constantly to overcome its limitations. It became a preferred choice in many areas of biology such as molecular biology, clinical pathology, and proteomics studies. The great success achieved in 2DE gel was the incorporation of immobilized pH gradient (IPG) strips and carrier ampholytes to form different pH ranges. Initial attempts with tube gel mode (Hochstrasser et al., 1986) and gel slab (Gianazza et al., 1985) were not very successful and the problems were later solved by the invention of IPG strips (Görg et al., 1987). Resolution of proteins was another problem, especially the resolution of hydrophobic and membrane proteins. This was partially solved by adding chaotropes in isoelectrofocusing (IEF) as a step of 2DE (Rabilloud et al., 1997; Santoni et al., 2000b; Rabilloud, 2009; Rabilloud et al., 2009). Certain chemicals such as 0.5% boiling SDS and 90% formic acid with cyanogen bromide can also help to solubilize membrane proteins (Mirza et al., 2007). 2DE gel visualization has also gone through several advancements from silver stain to fluorescence dyes and from nanogram to picogram range of protein detection precision. These advancements allow 2DE gel technique to detect low abundant proteins in a sample. In 2D-DIGE technique, cyanine minimal dyes (CyDyes; Amersham Biosciences) differing in their excitation and emission wavelengths are used (Ünlü et al., 1997). An internal standard is also incorporated in this method to reduce gel to gel variations.

Non-gel-based techniques: proteomics with mass spectrometry

Major advances in proteomics also happened with the invention of mass spectrophotometry-based methods. Fast atom bombardment and plasma desorption techniques were used for direct analysis of peptide and proteins. These techniques suffered several limitations. Later inventions of MALDI (Karas and Hillenkamp, 1988) and ESI MS (Fenn et al., 1989) techniques have overcome some of the problems and popularized MS based techniques in protein identification. The above mentioned techniques were used together with the enzymatic (trypsin) digestion of proteins to generate protein sequence database (Henzel et al., 1993; James et al., 1993; Mann et al., 1993; Pappin et al., 1993).

The basic steps used in the MS-based technologies include generation of ions from protein samples, separation based on mass and charge, and detection of ions. The most common analyzer used in MALDI is TOF. TOF analyzer accelerates a packet of ions with electric potentials, differentiates them by the time they take to traverse a flight tube, and calculates the m/z value, which is the time required to move from the ion source to detector (Van Breemen et al., 1983). MS based techniques usually employ two approaches “top down” and “bottom up” for protein identification. Top down method is generally used for the analysis of intact proteins and bottom up approach is used for the analysis of peptide mixtures. The basic difference between top down and bottom up approaches is that the former one uses intact protein and cleaves them in the gas phase and the latter one uses tandem mass spectrometry for the digestion of proteins. ESI-FTMS which is based on top down approach is able to analyze intact proteins with high resolution power. It uses off-line sample introduction with direct fragmentation of protein

ions in the device by using electron capture dissociation (ECD) (McLafferty and Senko, 1994). FTMS based techniques can detect protein sequence error and post translational modifications. Bottom up approach is quite different because it uses proteolytic enzymes to generate peptide fragments to obtain their amino acid sequences.

Analysis of two dimensional gel images

Commercial software packages for 2DE gel image analysis

A broad range of commercial software packages are available for 2DE gel image analysis for the detection and quantification of protein spots. In spite of being fully automated or semi-automated these techniques still need extensive human intervention for spot detection which is time consuming and error prone. Another drawback related to these software packages is that some of them are expensive and not in open sources. The need of fully automatic, open-source and fast analysis software or program is highly required in 2DE gel image processing. The overall success of image analysis is totally dependent upon the accuracy and reliability of these software packages which demand less user intervention during analysis. It is difficult to choose the software that meets the specific needs.

Several groups have compared different software packages to identify best suitable ones to meet the needs. Rosengren et al. (Rosengren et al., 2003) compared PDOquest and Progenesis software packages and Raman et al. (Raman et al., 2002) provided a list of ten commercially available software packages along with their price and platform. Most of the available software packages follow three basic steps of image analysis including spot detection, spot matching, and spot quantification. Melanie 3.0 was the first software invented in 1991 for 2DE gel image analysis. It has undergone several

advancements and became a most widely used software which was followed by PDQuest software with same fundamental steps in the image alignment. Another software Z3 was invented in 2000 which employs automatic alignment and dual color visualization system to see differences in protein expression (Appel et al., 1997). Delta 2D software also arrived in 2000, using a standard gel for spot matching (Millioni et al., 2010). Progenesis Same Spot is a more advanced software that provides fully automatic image analysis and has sophisticated data visualization and analysis tools (Rosengren et al., 2003). Pinnacle is a software that computes 'mean gel', which involves averaging the intensities pixel by pixel across all of the gels and detects spots by their pinnacles (Morris et al., 2009). It has also used fixed windows for spot detection and quantification. RegStaGel also uses mean gel image and provides advanced statistical tools for data analysis (Li and Seillier-Moiseiwitsch, 2011).

Further developments are needed to provide the best solution in 2DE gel image analysis. The Matlab image processing tool box can be a preferred method (Figure 1.1) for many users in compare to currently available software packages due to its easy accessibility, less user intervention, quick automation, and well written documentation. User can take advantage of their built-in compiler to generate their own applications. Matlab is a high-level language and an interactive software package used by scientists and engineers and it is developed by MathWorks. It allows development of algorithms, simulation, data analysis, communication, signal and image processing by the Image Processing Toolbox (Lyra and Georgantzoglou, 2011). Matlab can support different file formats and images can be saved and executed in extensions such as 'JPEG' ((Joint Photographic Experts Group), 'BMP' (Microsoft Windows Bitmap), 'TIFF' (Tagged

Image File Format), and 'PNG' (Portable Network Graphics). Challenges present in 2D gel image analysis by commercial software packages can potentially be substituted by "Matlab image processing" tool (IPT) box. The overall interpretation of 2D gel images requires extensive preprocessing.

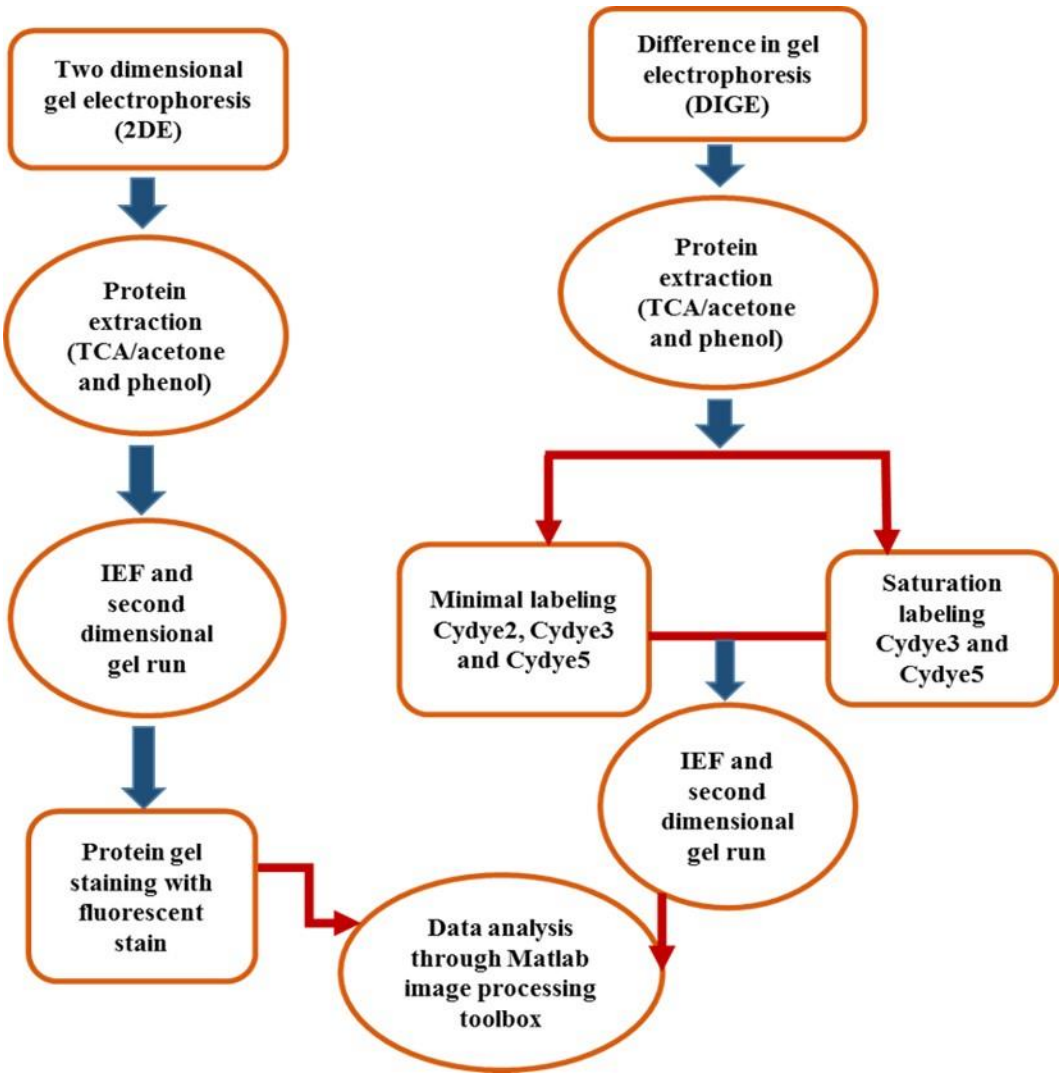

Figure 1.1 Schematic presentation of 2DE and 2D-DIGE with Matlab Image Processing toolbox

Preprocessing of 2DE gel images

Preprocessing of gel images is usually a beginning step in the image analysis pipeline. A summary of image processing using different method is shown in Fig. 1.2. Denoising of 2DE gel images can enhance the visual interpretation of gel images and allow the proper detection of spot boundary and spot shape which can ultimately affect protein quantification. Several methods are currently available for denoising 2DE gel images such as linear filtering, median filtering and filtering in the wavelet domain. Kaczmarek et al. (Kaczmarek et al., 2004) has divided the filtering in linear and non-linear. Two 2DE gel images can have different noise levels and even their backgrounds can vary significantly. Linear filtering methods such as Gaussian filter (Canny, 1986) and Wiener filter (Sonka et al., 2008) can be used during the time of noise removal. Non-linear filtering includes median filtering which can easily reduce spike like components (salt-and-pepper noise) (Meyer-Baese, 2004) that are often presents in 2DE gel images (Ahad, 2011). Donoho et al. has proposed another approach to remove the noise known as wavelet shrinkage which is mostly used in signal processing and image analysis area (Donoho and Johnstone, 1994; Donoho, 1995; Donoho and Johnstone, 1995; Donoho et al., 1995).

2D gel image alignment

Daszykowski et al. in 2009 used “Matlab image processing toolbox” (IPT) for matching 2DE gel images (Mehl et al., 2012). They employed IPT for 2DE gel image alignment on the basis of a subroutine present for manual marker selection. They described the need of a standard gel image which is highly required in 2DE gel image alignment step. The images which have highest mean correlation should be chosen as a

standard image. Nhek et al. also described that 2DE gel image alignment is an important step before spot detection (Nhek et al., 2012). The alignment of 2DE gel images falls in many categories such as Landmark based, spot-based, and intensity based alignment. Landmark based alignment which is commonly used by researchers requires matching of corresponding spots located by their spot centers and further align them (Rohr et al., 1999; Efrat et al., 2002). Another approach is intensity based alignment which does not require spot detection or matching at earlier stage of 2DE gel image processing (Smilansky, 2001; Yang, 2001; Woodward et al., 2004). This approach can be better after employing landmark based alignment. In other alignment methods combination of landmark and spot based alignments were used to improve protein quantification (Rogers and Graham, 2004; Rohr et al., 2004). These combinations are already used in the area of medical imaging (brain and heart imaging) (Collins et al., 1998; Wang and Staib, 1998; Cachier et al., 2001; Johnson and Christensen, 2002).

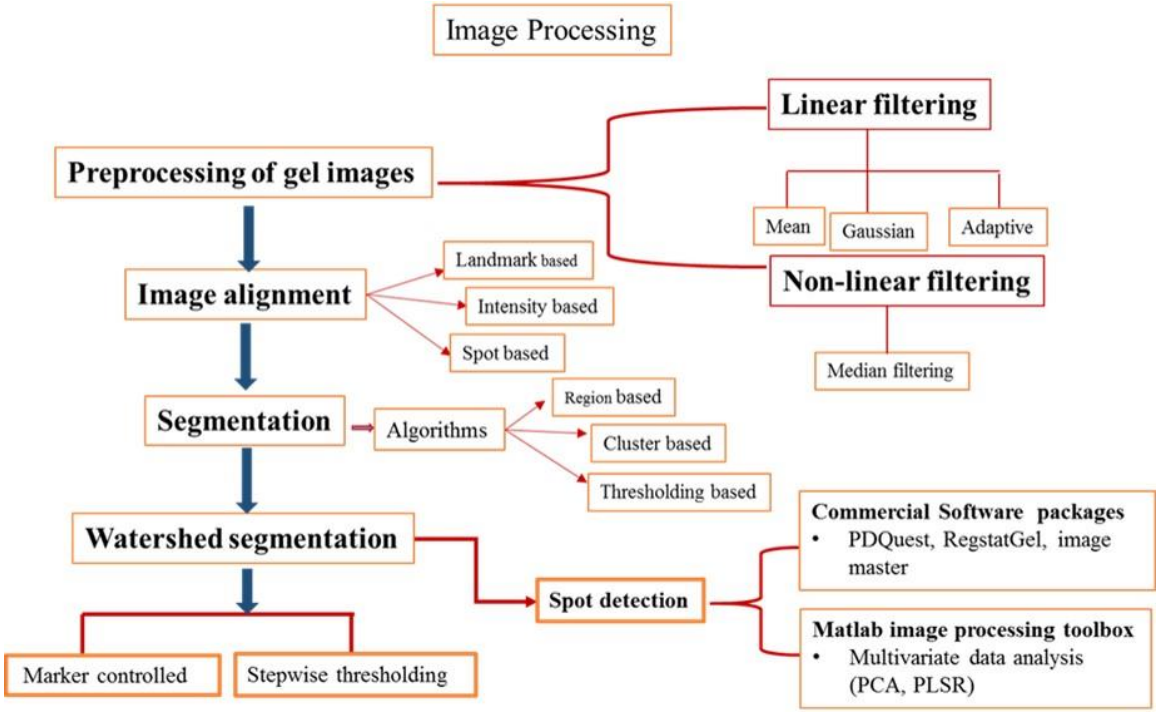

Figure 1.2 Overall methods of image processing.

2DE gel image segmentation

Segmentation is another major area in 2DE gel image analysis which generally referred to as the process of separating areas related to protein spots and eliminating the background. The final result of segmentation produces individual spots and spot boundaries. The main aim of segmentation is the partition of the image into the set of regions which are distinct from low level intensity regions such as color, gray level and texture (Phung et al., 2005). Region in an image can be defined as pixels or group of pixels having a boarder and shape such as circular in case of protein spots, ellipse and polygon. The main advantage of segmentation is that it makes an image more meaningful and easy to analyze. Several algorithms has been proposed to define image segmentation which is broadly classified into Edge detection methods, thresholding based methods,

region based methods, cluster based methods and graph based methods (Agrawal et al., 2010). Selection of algorithm which best suited for segmentation is purely dependent upon selecting different algorithm for different images. Evaluation of these algorithms can be checked by giving same parameter for the segmentation of multiple images or giving different values to the algorithm's parameters for segmenting some comparable images. However, there is no universally accepted algorithm present for the segmentation. Zhang et al. has proposed a classification of evaluation methods in several groups as "analytical", "empirical goodness", or "empirical discrepancy" (Zhang, 1996). Agrawal et al. (Agrawal et al., 2010) have employed several algorithms by using Matlab. They have applied canny algorithm (Canny, 1986), split and merge algorithm(Strasters and Gerbrands, 1991) and cluster based methods.

Mathematical morphology is a vast area in image analysis, and Beucher et al. (Beucher and Lantuejoul, 1979; Beucher, 1983) have used this method first time for analysis of 2DE images. This is basically including watershed based segmentation or marker controlled watershed segmentation to deal with the problem of over segmentation in 2DE images (Skolnick, 1986; Vincent, 1993). In watershed approach, viewing grayscale image as a landscape, identifying local minima in the landscape and finding of catchment basin associated with the local minima. Watershed lines are the outline of a catchment basin and each separated catchment basin represents the individual protein spot. Vincent and Soille (Vincent and Soille, 1991) has explained gray scale images as landscape with tops and valleys and used immersion principle to describe it. Wu et al. has described an open source software RegStatGel (Matlab-based) (Li and Seillier-Moiseiwitsch, 2010) and RegStatGel employs mean gel image for watershed

segmentation. Like RegStatGel, Pinnacle software also employs mean gel image for water shed segmentation to avoid over segmentation problem present in 2DE images (Wu and Zhang, 2011). Some of the methods have combined different algorithms to obtain better results. In conclusion, combination of segmentation methods can be more effective compared to single method.

Multivariate data analysis

Study of any phenomenon usually depends upon several factors. Data analytical methods using only single variable are known as univariate methods and the data analysis methods which use multiple variables are known as multivariate data analysis methods. Multivariate data analysis methods are more useful than univariate data analysis methods because univariate data analysis is often not sufficient for the complete and accurate data analysis.

Principle component analysis

Principle component analysis is one of the oldest and most popular methods for multivariate data analysis, which decompose a high dimensional data matrix into variables of lower dimensions. These variables are also known as principle components, eigenvectors, or loadings. PCA (Martens, 1989) is mostly represented as a graph in which the direction shows the data displaying high variability. PLSR (partial least square regression) (Wold et al., 1983; Rencher and Christensen, 2012) is also closely related to PCA. The major difference between PCA and PLSR is that PLSR deals with variations in covariance matrix. Another version of PLSR known as Discriminant PLSR (DPLSR) uses discrete type value instead of continues values (Martens and Martens, 2001).

Proteomics approach in plants in recent years

Advancements in proteomics include the refinement in 2DE techniques and the involvement of MS-based techniques for identification of proteins and acquisition of genome sequence information in plants (Griffin et al., 2001; Mann et al., 2001). As a consequence of these developments, comparative proteomic studies and proteome maps have emerged in plant systems to identify proteins and develop biomarkers. Porubleva et al. (Porubleva et al., 2001) have done global mapping of maize (*Zea mays*) leaves. In another study, proteome maps were also generated for wheat (*Triticum aestivum*) (Skylas et al., 2005), poppy (*Papaver somniferum*) and latex (*Hevea brasiliensis*) (Decker et al., 2000). In Arabidopsis subcellular proteomes including cell wall, plasma membrane and endoplasmic reticulum have also been mapped (Robertson et al., 1997; Mahon et al., 2000; Santoni et al., 2000a). Arabidopsis luminal proteins present in chloroplast and envelop proteins and thylakoid membrane proteins (Hippler et al., 2001) and plastid ribosomal subunit proteins from *C. reinhardtii* (Yamaguchi and Subramanian, 2000) were also mapped.

Plants lack an immune system and thus they are vulnerable to pathogen attack. To fight with these pathogen attacks, plants have evolved many defense related systems within themselves. Pathogenesis related proteins are the part of the evolutionary mechanism to protect plants from pathogen attacks. Several pathogenesis related proteins have been identified in maize plants in recent decades. Pathogenesis related proteins (PR) plays important role in plant defense system and they are grouped in 17 independent families (PR-1 to PR-17). PR-2, PR-3, PR-4, PR-8 and PR-11 proteins display β-1, 3-glucanase (glucan endo-1, 3-β-glucosidase) and endochitinase activity, respectively.

Some of the PR proteins are chitinase, endoproteinases, peroxidases, proteinase inhibitors, thaumatin-related proteins and other small protein molecules including thionins, defensins, lectins and heveins. The role of these enzymes are to protect plants from fungal infection (Datta and Muthukrishnan, 1999; Kitajima and Sato, 1999; Van Loon and Van Strien, 1999; Midoro-Horiuti et al., 2001). Chitinase (EC 3.2.1.14) plays important role against fungal pathogens because of its ability to digest chitin, a crucial cell wall material of fungi (Huynh et al., 1992). Purified plant chitinases display strong antifungal activity against two very important non-pathogenic plant fungi such as *Trichoderma reesei* and *Trichoderma hamatum*. Chitinases are classified in three forms, class I enzymes which contain a cysteine rich domain and a highly conserved domain and class II enzymes which lack a cysteine rich domain but have a catalytic domain similar to class I enzymes. Class III chitinases are a group of chitinases not showing homology with class I and class II chitinases. Class III chitinases are homologous to the acidic chitinases of cucumber and Arabidopsis (Metraux et al., 1989).

Antifungal properties of chitinase A (Chit A) and chitinase B (Chit B) which belong to class I chitinase have been reported in maize seeds. Several other types of chitinases have been isolated from maize kernels, embryo, pericarp and germinating seeds (Metraux et al., 1989). Enzyme β- 1, 3-glucanase also play important role in plant defense system, and it is present in maize endosperm and kernels and its level increases at the time of fungal infection. These antifungal proteins in maize has different mechanism to kill fungal pathogen such as zeamatin kill the fungus by permeabilizing fungal hyphal membrane and causes leakage of cytoplasmic contents. Another family of defense proteins, ribosome inactivating proteins (RIPs, EC 3.2.2.22), defends the plant by binding

to the pathogen's ribosomes and inhibiting protein synthesis (Virgilio et al., 2010). They basically depurinate a conserved adenine and make ribosomes impaired in transcriptional elongation process (Bass et al., 2004). RIPs are classified mainly in three types; RIPI, RIPII and RIP III. Maize RIP is different from others because it is synthesized as an inactive precursor known as proRIP or b-32.

Other novel proteins identified in maize were the basel-layer type antifungal proteins (BAPs) which attacks fungal plasma membrane. BAP genes are located in chromosomes 4 and 10 in maize and found in three different forms, *bap*1, *bap*2 and *bap*3 (Serna et al., 2001). The stress related proteins aldose-ketose reductase, antioxidants peroxidase, glyoxylase 1, globulins, late embryogenesis related proteins and several heat shock proteins were recognized in maize embryos and kernels that may be contributing to pathogen resistance.

CHAPTER II

COMPARATIVE STRUCTURAL ANALYSIS OF MAIZE PROTEINS TENTATIVELY ASSOCIATED WITH RESISTANCE TO *ASPERGILLUS FLAVUS* INFECTION AND AFLATOXIN CONTAMINATION

Abstract

A number of candidate resistance genes have been identified through previous gene expression and quantitative trail loci (QTL) studies. Sequencing of these genes revealed abundant DNA polymorphisms including SNPs. Despite the bulky amount of SNPs present in gene sequences, only a few of them may be involved in changes of protein structures. The research goals in this chapter were to investigate methods on identification of SNPs related to differences in protein structures. Homology modeling and comparative modeling methods were used to study structural changes caused by SNPs in proteins present in resistant maize (Mp313E and Mp715) and susceptible maize inbred line (Va35). Selected candidate genes were amplified by RT-PCR method from cDNAs and then the sequences of genes were determined by sequencing. SNPs in candidate gene sequences were identified through sequence analysis. The corresponding protein sequences were determined by using bioinformatics tools and by searching available maize protein database. The 3D structures of these proteins were determined by using the Modeller software in the following steps. In the first step, the protein sequence

to be modelled was used as ‘query sequence’. Four to six template sequences for each query sequence were then selected through searching protein database (PDB). Homology modeling was performed to build the 3D structural model from the query protein sequence. Protein structure comparisons between maize inbred lines were performed by using PyMOL software.

Keywords. *Aspergillus flavus*/Modeller/PDB/RT-PCR

Introduction

Computational modeling of protein 3D structures is an alternative approach for the understanding of protein structures and functions when experimentally determined protein models are not available. Protein structures play crucial roles in biological and developmental processes. The journey of protein synthesis starts with mRNA translation. In general, the product of mRNA translation is a polypeptide chain consisting of a linear sequence of amino acids residues. Through the folding of the polypeptide chain and the interactions between amino acid residues, a well-defined three dimensional structure is generated, which is known as the native state of the protein (Anfinsen, 1973). Under various conditions, proteins adopt primary structure (linear chain of amino acid residues), secondary structure (repetitive structure such as α-helix and β-sheet), tertiary structure (spatial arrangement of a polypeptide chain), and quaternary structure (spatial arrangement of multiple polypeptide chains) which ultimately defines the functions (Schulz and Schirmer, 1979; Tompa and Fersht, 2010). Proteins perform various functions in biological systems such as winding and unwinding of DNA, enzymatic reactions in metabolism, and repair of DNA sequences (Hsieh and Brutlag, 1980). Protein structure prediction, through either experimental or computational approaches, has

become an important area to understand the physical and chemical properties of proteins and their specific roles.

Through genome sequencing projects, enormous amounts of DNA sequence data have been generated in the recent decade. With the advent of PCR technique which amplifies a piece of DNA sequence exponentially (Bartlett and Stirling, 2003) and advancement of new sequencing techniques, additional large volume of sequence data are available for genome-wide sequence comparative studies. Analysis of DNA sequences have revealed abundant DNA polymorphisms including SNPs and indels, providing fundamental basis for the development of gene-based markers which are essential for quantitative trait loci studies in breeding programs. A number of candidate resistant genes have been identified through previous gene expression and quantitative trail loci (QTL) studies. Despite of the bulky amount of SNPs present in gene sequences, only a few of them may be involved in the changes of protein structures. SNPs associated with protein structure changes are considered as highly informative DNA markers. Screening of informative SNPs is crucial to the efficiency of gene-based DNA marker development.

Genome sequence data can be used to identify gene coding regions, DNA polymorphisms, and DNA markers. Genome sequence data also can be used to develop DNA microarrays for the study of gene expression. Genome sequencing projects have rapidly grown and yielded nearly 100,000 predicted protein sequences (Martí-Renom et al., 2000). However, only a few thousands of protein 3D structures have been determined experimentally by X-ray crystallography or NMR spectroscopy methods. Especially, the development of protein crystals (the precipitation of proteins from its solution) of membrane proteins are the limiting factors present in X ray crystallography and NMR

spectroscopy methods (Durbin and Feher, 1996). The disparity between protein sequences and their available 3D structures can be resolved by the employment of computational methods.

Computational based methods such as comparative modeling or homology modeling are among the most successful methods for the determination of 3D structure of protein. Homology modeling method predicts the 3D structure of the query protein sequence based on its similarity with protein sequences of known structures in the Protein Data Bank (PDB) (Roy and Zhang, 2012). It is a well-known fact that if two different proteins share a high sequence similarity they are likely to have a very similar three-dimensional (3D) structure and similar functions too (Floudas, 2007b). The approaches include fold recognition, target-template alignment, and multiple template based alignments. Homology modeling provides the initial information about protein 3D structures which can be easily investigated further by experimental methods (McGuffin et al., 2000). Protein modeling also provides a powerful tool to identify changes (if any) in the highly similar sequences of homologous proteins that may cause changes in the 3D structures and functions.

Accurate modeling method requires high similarity between query and template sequences. Modeling accuracy decreases sharply if the query and template sequence identity is in the lower range of 30-50% or below 30% template based modeling is less effective (Daszykowski et al., 2009; Nhek et al., 2012). However, it is often difficult to identify the best template among the similar templates. Several approaches such as BLAST and PSI-BLAST have been developed to identify optimal template. Template identified by BLAST (Altschul et al., 1990) and PSI-BLAST (Altschul et al., 1997) are

based on E-value which reflects the probability of a given template to be homologous to the query sequence. Template sequence which has zero E-value is considered as a best template for the alignment purpose. Modeller programs satisfy spatial restrains that include stereo chemical restrains (bond angle and bond length) and homology derived restrains (Eswar et al., 2006). It also satisfies statistical restrains which is expressed as probability density factor (PDB) for dihedral angles, non-bonded inter-atomic distances (Hess et al., 2008) and restrains obtained from experimental procedures such as NMR spectroscopy, image reconstruction from electron microscopy and site-directed mutagenesis (Eyrich et al., 2001).

Modeller program has been applied on multiple alignment of protein sequences (Madhusudhan et al., 2005), calculation of phylogenetic tree (Fiser, 2004), identification of disease resistant related protein (Dehury et al., 2013), and the design of inhibitors (Xiong et al., 2003). Homology modeling becomes a preferred approach to design 3D structure of proteins and to understand their functions in the vast area of proteomics research. Homology modeling study was performed to get the structural information of resistant protein xa5 in rice when no crystal structure was present. This protein is a bacterial blight disease resistant protein which is caused by pathogen Xanthomonas oryzae (Dehury et al., 2013). In another study, structure of CYP51 protein of A. fumigatus was determined to characterize mutations present in 3D models (Snelders et al., 2010). Maize (Zea mays L.) is also not an exceptional case where homology modeling approach was incorporated to identify structural difference between homologous proteins such as Hm1 and Hm2 resistant to fungus Cochliobolus carbonum. This is a most destructive biotic fungus which can infect susceptible maize plants at any

stage of their development (Dehury et al., 2014). The examples mentioned above demonstrated the role of homology modeling in the protein 3D structure design to get an insight of their function.

In the current study, homology modeling approach was used due to lack of crystal structures of proteins in resistant and susceptible maize (*Zea mays* L.) inbred lines related to aflatoxin research. This approach can provide the information on predicted structures of susceptible and resistant related proteins of maize inbred lines. It also can provide methods on identification of highly informative SNPs associated with differences in protein structures. Characterization of informative SNPs is crucial to gene-based DNA marker development.

Materials and Methods

Collection of the plant material and RNA extraction

Two resistant (Mp313E, Mp715) and one susceptible maize inbred line (Va35) were planted in the green house and harvested for leaf sample collection. All leaf tissues were ground with liquid nitrogen and total RNA was extracted by using an Aurum™ Total RNA kit. Around 100 mg of ground leaf tissue was transferred to a 2 ml micro centrifuge tube and 1 ml of PureZol was added immediately. Once the sample has been disrupted in PureZol, the lysate was incubated at room temperature for 5 min. The sample was then centrifuged at 12,000g for 10 minute at 40C and followed by 5 min incubation at room temperature. Then 0.2 ml of chloroform was added and the lysate was shaken vigorously for 15 sec to 30 sec and mixed periodically for 5 minutes. Sample was then centrifuged at 12,000 g for 15 min at 40C. In this step the sample was separated into two phases and the upper aqueous phase containing RNA (200-400 µl) was transferred to a

new 2 ml microfuge tube and an equal volume of 70% ethanol was added and mix by pipetting. A RNA binding mini column was placed in a cap less 2 ml tube and RNA sample was poured into RNA binding mini column and was centrifuged for 1 min at 12,000g. In the next step, 400 µl of low stringency wash solution was added in the RNA binding column and centrifuged for 30 sec at 12,000 g. The contaminating genomic DNA was removed by adding 80 µl of DNase I enzyme into the tube and incubating for 15 min at room temperature. High stringency wash solution (400 µl) was then added two times into the RNA binding column and followed by centrifugation for 30 sec at 12,000g. Centrifugation step was performed for additional 2 min to remove residual wash solution. RNA binding column was then transferred to a 1.5 ml capped micro centrifuged tube. Finally 30 µl of elution buffer was added in the RNA binding column, incubated for 1 min, centrifuged for 2 min at 12,000 g to elute total RNA. RNA samples were stored at -800C until use.

cDNA synthesis and RT-PCR

cDNA synthesis was performed from total RNA samples. cDNA was synthesized by mixing 5 µl of total RNA (approximately 2.5 µg) with a 7 µl of Master mix I (1 µl of 50 µM oligo dT primer, 2 µl 10 mM dNTP and 4 µl DEPC water) and 8 µl of Master mix II (4µl 5X cDNA synthesis buffer, 1 µl 0.1M DTT, 1 µl 40 U/ µl RNaseOut, 1 µl DEPC water and 1 µl of 15 U\µl of Thermoscript Transcriptase). RNA and Master mix I were denatured by incubating at 65 ^{0}C for 5 min and then placed on ice. Master mix II was then added and the reaction mixture was incubated at 50 ^{0}C for 2 h. Reaction was terminated by heating at 80^{0}C for 5 min. cDNA was stored at -80 ^{0}C until use.

RT-PCR method was performed using cDNA. Primer sequences were developed by Primer 3 program. A list of primers and their sequences used in this study is shown in Table 2.1. Thermo cycler programmed to incubate at 95 ^{0}C for 3 min and then to proceed with 40 cycles of denaturation (94°C, 30 s), primer annealing (58^{0}C for 50 s), and primer extension (72 ^{0}C for 1 m 20 s). Resulting PCR products were examined with 1% agarose gels (0.5g agarose in 50 ml of 1X TAE buffer). Ten µl of 1Kb ladder and 20 µl of samples were loaded into the wells. The gels were run in 1X TAE (diluted from 50X TAE that contains 121g Tris base, 28.55ml glacial acetic acid and 50 ml of 0.5 M EDTA pH 8.0) buffer for 30 min at 75 V and stained with 10 µl of ethidium bromide for 15 minutes. RT-PCR products were visualized with UV light.

Cloning and sequence analysis

RT-PCR products were purified and cloned into pGEM^R-T Easy plasmid vector (Promega). Cloning was performed in three steps. In the first step, A-tailing reaction start with 5µl of purified PCR product generated by a proofreading polymerase (*Pfu* DNA Polymerase) was added into 0.2ml PCR tube. Then 1µl of 5 units of *Taq* DNA Polymerase, 1µl of 0.2mM dATP and ddH_2O was added to make the final reaction volume of 10 µl. Reaction mixtures was incubated at 70^{0}C for 30 minutes and in then hold for 4^{0}C. Cloning was performed with pGEM^R-T Easy plasmid vector (Promega, Madison, WI). The ligation reaction for a sample was prepared having 5 µl of 2x ligation buffer, 1 µl plasmid vector, 3 µl PCR product, and 1 µl T4 DNA ligase buffer. A control reaction was prepared having 5 µl of 2x ligation buffer, 1 µl plasmid vector, 3µl ddH_2O and 1 µl T4 DNA ligase. The ligation mixture was incubated for 1h at RT. In the final step, 2 µl from the ligation mixture was used for the transformation using heat-shock

method. Tubes containing ligation mixture and the JM109 (Promega, Madison, WI) competent cell were put on ice for 20 min and then on water bath at 42^0C for 50 sec. Tubes were removed from water bath and again put on ice for 2 min and then were kept on shaker at 37^0C for an hour. A 500 µl of the transformation culture was plated per; LB plate containing 100 µl of 40 mM IPTG and 100 µl of 2% X-gal. Plates were incubated overnight at 37^0C. Plasmid DNA was extracted by using Zyppy™ Plasmid miniprep kit according to the manufactures protocol and the sequencing were performed by sending samples to SeqWright program.

3-D structure prediction of proteins

The sequencing data of cloned maize genes were analyzed employing Biology Workbench program (Phylip rooted tree-Phenogram). The protein structural modeling was performed by using Modeller 9.13 program (http://salilab.org/modeller/download_installation.html) which uses Python as its control language. In protein structure modeling experiments, when the sequences obtained for the maize inbred lines were not all complete gene sequences. The closest related sequences were selected by employing NCBI Blast program. The nucleotide sequences were translated into amino acid sequences by using program SDSC Biology Workbench 3.2. The amino acid sequences of the maize lines (susceptible and resistant) were used as "query" sequences. Five promising "template" sequences were selected for each query sequence. The query and template sequences were saved in a basic file folder obtained from Modeller program.

Building of profiles

First, the query sequence was put in PIR format and saved with an .ali extension. These ali sequences were input files for profile.build () command. The output of the "build_profile.py" script was written to the "build_profile.log" file produced by Modeller program. Modeller also generated a text file "build_profile.prf" file. It displayed PDB IDs for the templates having more than >35% sequence similarity with the query sequence. In total twelve columns were generated by this command. Second column basically reported the IDs of PDB sequence which was used to compare with the query sequence. These PDB sequences have generally more than 95% sequence identity and less than 35% sequence length difference with each other. Sixth column represented E-values generated by Modeller. The eleventh column represented the sequence identity between a query sequence and a PDB sequence normalized by the length of the alignment.

Template selection

General method for best template selection was by considering the e-value generated with the build.profile () command. In this study, five templates were first selected on the basis of high sequence identity and zero E-value for each query sequence. The most appropriate template among these five templates were then selected using alignment.compare_structures () command. This command was used to assess the structural and sequence similarity between the templates. Finally a compare.log file was generated after running this command which showed a dendrogram having templates used in this study. Difference between these templates can be identified through this dendrogram. It calculated a clustering tree from the input sequences. The most

appropriate template was then selected on the basis of high sequence similarity with target sequence and best crystallographic resolution.

Alignment of query and template sequences

Query sequence was then aligned with template sequence using align2d () command. This command took structural information from the template and place gaps in solvent exposed regions or curved regions. This command used dynamic programming algorithm instead of standard alignment methods such as alignment of one sequence to other sequence. Once query-template alignment was constructed, Modeller 9.13 automatically calculated 3D model from the query sequence using model-single.py command. It generated five models with .pdb extension. Finally model-single.log file was generated which contains DOPE and GA341 scores for all the five models.

Model building

Best model for each query sequence was selected by observing lowest DOPE (Discrete optimized protein energy) assessment score and highest GA341 score. Once the modeling was performed, 3D structure of the model was visualized by the software PyMOL v3.1.0 (www.pymol.org).

Results

Nucleotide sequence alignment and template selection

RT-PCR cloning was performed on fifteen candidate genes or gene families using primers designed and listed in Table 2.1. All the PCR products were sequenced. The modeling of glycine rich RNA binding proteins (GRBPs) was given here as an example showing the methodology of protein modeling approaches. Forward primer AGP2F1 (5'

ATGGCGGCGTCGGATGTTGA 3') and reverse primer AGP2R1 (5' TCAGTTCCTCCAGTTCCCGT 3') were used for RT-PCR cloning of GRBP genes. Multiple sequence alignment was performed using ClustlW program of SDSC Biology Workbench 3.2 for GRBP genes in three maize inbred lines (Mp313E, Mp715 and Va35). As shown in Figure 2.1, GRBPs in two resistant maize inbred lines Mp313E (279 bp) and Mp715 (276 bp) were closely matched with genbank sequence BT085135.1 (794 bp). BT085135.1 is a Zea mays cDNA full length clone, complete cds. As shown in Figure 2.1, genebank sequences (EU961238.1, BT035160.1, EU955998.1 and BT060694.1) were very well matched with the GRBP gene in susceptible maize inbred line Va35 (438bp) nucleotide sequences. Through multiple sequences alignment, it is observed that SNP's are present in GRBPs between Va35 and Mp715 or Mp313E inbred lines.

Original nucleotide sequence of the GRBP gene in Va35 and the BT085135.1 gene (794 bp) which has sequence identity 98% with Mp715 were selected for 3D protein structure prediction. Amino acid sequence of BT085135.1 was used as Mp715 in the result section. Template (PDB ID: 4C7Q) was selected for both Mp715 and Va35 for alignment purposes. It has 63% sequence identity, 38% query coverage with Mp715 and crystallographic resolution was (1.0 Å) and 77% sequence identity, 56% query coverage with Va35 and crystallographic resolution was (1.0 Å). This was one of the best templates selected for Va35 and Mp715.

Selection of the best model and visualization through PyMOL

A total of five models for each gene were generated through modeler program. The best model was selected on the basis of lowest DOPE and highest GA341 scores as shown in Table2.2 and Table 2.3.

Five models were constructed for GRBP gene in resistant maize inbred MP715. As shown in Table 2.3. These are listed as (Mp715.B99990001.pdb, Mp715.B99990002.pdb, Mp715.B99990003.pdb, Mp715.B99990004.pdb and Mp715.B99990005.pdb). On the basis of lowest DOPE score (-6594.04102) and highest GA341 scores (0.22658) Mp715.B99990002.pdb was selected as a best model among five models.

Five models were also constructed for susceptible maize inbred Va35. As shown in Table 2.2. These are listed as (Va35.B99990001.pdb, Va35.B99990002.pdb, Va35.B99990003.pdb, Va35.B99990004.pdb, and Va35.B99990005.pdb). On the basis of lowest DOPE score (-8072.1902) and highest GA341 scores (1.00000) Va35.B99990002.pdb was selected as a best model among five models.

These five best models were visualized through PyMOL software. The 3D models of Va35 and Mp715 are shown in Figure 2.4 and 2.5, respectively. 3D model of Va35 contains two α helices and six parallel β sheets and Mp715 contains two α helices and four parallel β sheets. These helices and sheets are conserved in the 3D models but the loop region contains significant difference. To compare the structural differences in GRBPs between resistant and susceptible maize inbred lines, models were aligned with each other. Alignment of models are shown in the Figure 2.6. Best models, Mp715.B99990002.pdb and Va35.B99990002.pdb were aligned to find out the differences present in GRBPs between these two maize inbred lines.

Table 2.1 A list of primers and their sequences used in this study

Primers	**Forward and Reverse sequences**	**Genes**
AGP2F1	ATGGCGGCGTCGGATGTTGA	Glycine Rich RNA Binding Protein Family
AGP2R1	TCAGTTCCTCCAGTTCCCGT	Glycine Rich RNA Binding Protein Family
AGP2F2	ATGGCGTCACTTCTCCGCCC	Glycine Rich RNA Binding Protein Family
AGP2R2	CTTAGGATAAAGCCTGACGG	Glycine Rich RNA Binding Protein Family
AGP2F3	ATGCCACTACCCTACTCAAC	Glycine Rich RNA Binding Protein Family
AGP2R3	AGGATGAACACAGCACTAAC	Glycine Rich RNA Binding Protein Family
BM379345SF	CCCCCACCACAAGCTAGATA	Metallothionein-Like Protein (MTLP)
BM379345SR	CCAAGGTCGAGGATGTTTTG	Metallothionein-Like Protein (MTLP)
BQ538849SF	CACAGAGGAAGAATGTTTGAGC	C2H2-type Protein
BQ538849SR	ACATGGCAACGATACACGAA	C2H2-type Protein
EU954539 F1	GGTCCTCGAGTCCAAGATCA	Glycine Rich RNA Binding Protein Family
EU954539 R1	ACACAGATGGGCAACAACAA	Glycine Rich RNA Binding Protein Family
EU954539 F2	GCGAGGTCCTCGAGTCCAAG	Glycine Rich RNA Binding Protein Family
EU954539 R2	AAACACAGATGGGCAACAAC	Glycine Rich RNA Binding Protein Family
EU957887.1F1	TGAGTACCGTTGCTTCGTTG	Glycine Rich RNA Binding Protein Family
EU957887.1R1	CTTGCGGTCAAAAACACAGA	Glycine Rich RNA Binding Protein Family
EU963153.1F1	ACGACCACTCCCTCAACAAC	Glycine Rich RNA Binding Protein Family
EU963153.1R1	ACACGGTAGCAGAAGCGAAC	Glycine Rich RNA Binding Protein Family
RBP2F1	ACAACGCCTTCAGCACCTAC	Glycine Rich RNA Binding Protein Family
RBP2R1	ACACAGATGGGCAACAACAA	Glycine Rich RNA Binding Protein Family

Table 2.1 Continued

RBP2F2	CGTCGGATGGGATCAGGA	Glycine Rich RNA Binding Protein Family
RBP2R2	CTGTTTGATGCATGGGTCTG	Glycine Rich RNA Binding Protein Family
RBP2F3	AACGAGGTCAGGTCAGTCAC	Glycine Rich RNA Binding Protein Family
RBP2R3	TTGCCTTGATTCGCACTCAA	Glycine Rich RNA Binding Protein Family
RBP2F4	AGCAAGTGTCCATATCCACC	Glycine Rich RNA Binding Protein Family
RBP2R4	TCCAGCTCCTTGCCGTTCAT	Glycine Rich RNA Binding Protein Family
TC207503SF	CTTCCTTTGGAAGCCAGTTG	Prenylated Rab Acceptor (PRA1) Family
TC207503SR	AGCACAGCTCACAAACAACG	Prenylated Rab Acceptor (PRA1) Family
TC218605SF	CTGCACTATCCAGCCACTGA	Phytochrome A (PHYA)
TC218605SR	TCTGAAAGCCAGAAGGCAAT	Phytochrome A (PHYA)
TC223372SF	CGCTGGTTGGTGAACTTGTA	Cinnamoyl CoA Reductase (CNCR 2)
TC223372SR	GTCACCGATGACCCTGAGAT	Cinnamoyl CoA Reductase (CNCR 2)
TC223736SF	CGGAGAACGAGTAGAGATGGA	Heat Shock Protein 90 (HSP90)
TC223736SR	GCCTAGCACTGCGAACTCTT	Heat Shock Protein 90 (HSP90)
TC226528SF	ATGAACCGTGTTCTCGCTCT	Uracil Permease (UPS)
TC226528SR	TCCCGAGTTTGTAGGCATTC	Uracil Permease (UPS)
TC232785SF	TCTTCCACCTCCTGGATCAC	F-Box Protein SKIP31-Like
TC232785SR	CAGCGTGGAATATTGTGCTG	F-Box Protein SKIP31-Like
TC234808SF	AAAACCCTAGCAGGCTTCG	Ribosomal Protein L30 (RPL30)
TC234808SR	CACTGGTGGAGCATTCGTT	Ribosomal Protein L30 (RPL30)
TC237311SF	GATCTGGTCATGCAGGAGGT	Heat Shock Protein 101 (HSP101)

Table 2.1 Continued

TC237311R	CCACATTCACGGGCTTATCT	Heat Shock Protein 101 (HSP101)
TC238832SF	TCTTGGGATCGGCTAGAGAA	Lecithin Cholesterol Acyltransferase (LCAT)
TC238832SR	AGCTCCATCAGCTCCTTGAA	Lecithin Cholesterol Acyltransferase (LCAT)
TC241201SF	CATGTTGCAGGATTTCATGG	Uncharacterized Protein
TC241201SR	AAGCTTCTCCGTCCAGTTCA	Uncharacterized Protein
TC245683SF	CCGGTCGGTCAGATATAAGC	Uncharacterized Protein
TC245683SR	AAGAACAATGGCACACGACA	Uncharacterized Protein
TC247683SF	TCTACGTAGTGTGCGGCTTG	Uncharacterized Protein
TC247683SR	CGCATTGAAATGATCCTCCT	Uncharacterized Protein

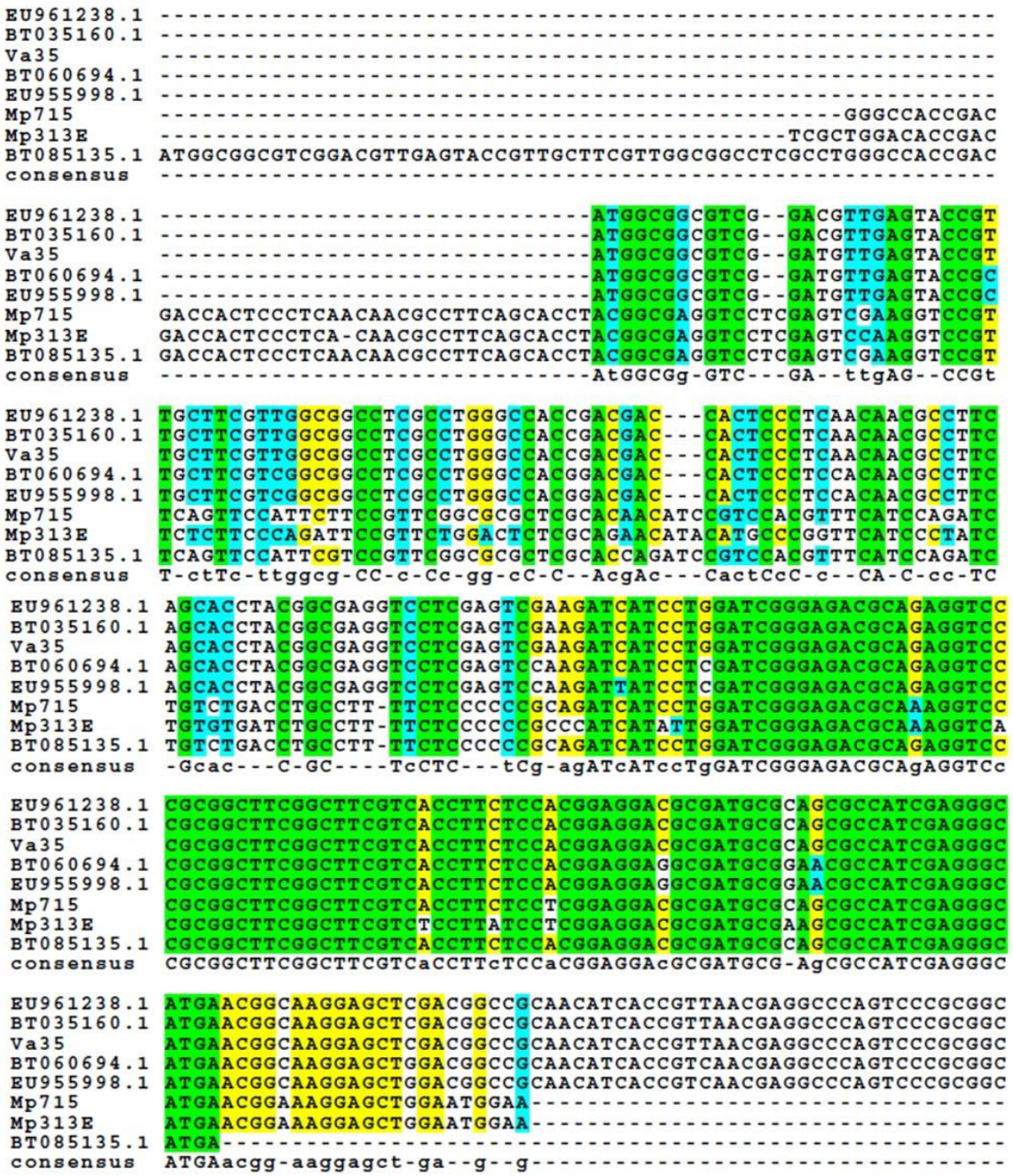

Figure 2.1 Alignment of nucleotide sequence of the GRBP gene in Va35 with its closely related sequences and the GRBP genes in resistant (Mp313E and Mp715) maize inbred lines.

Sequences closely related to Va35 and Mp313E and Mp715 were obtained from Genbank. Alignment was performed using ClustlW program of SDSC Biology Workbench 3.2. GRBPs in Mp313E and Mp715 were closely matched with genbank sequence BT085135.1. Conserved sequences were marked in green. Sequence polymorphisms were marked in blue, white, and yellow.

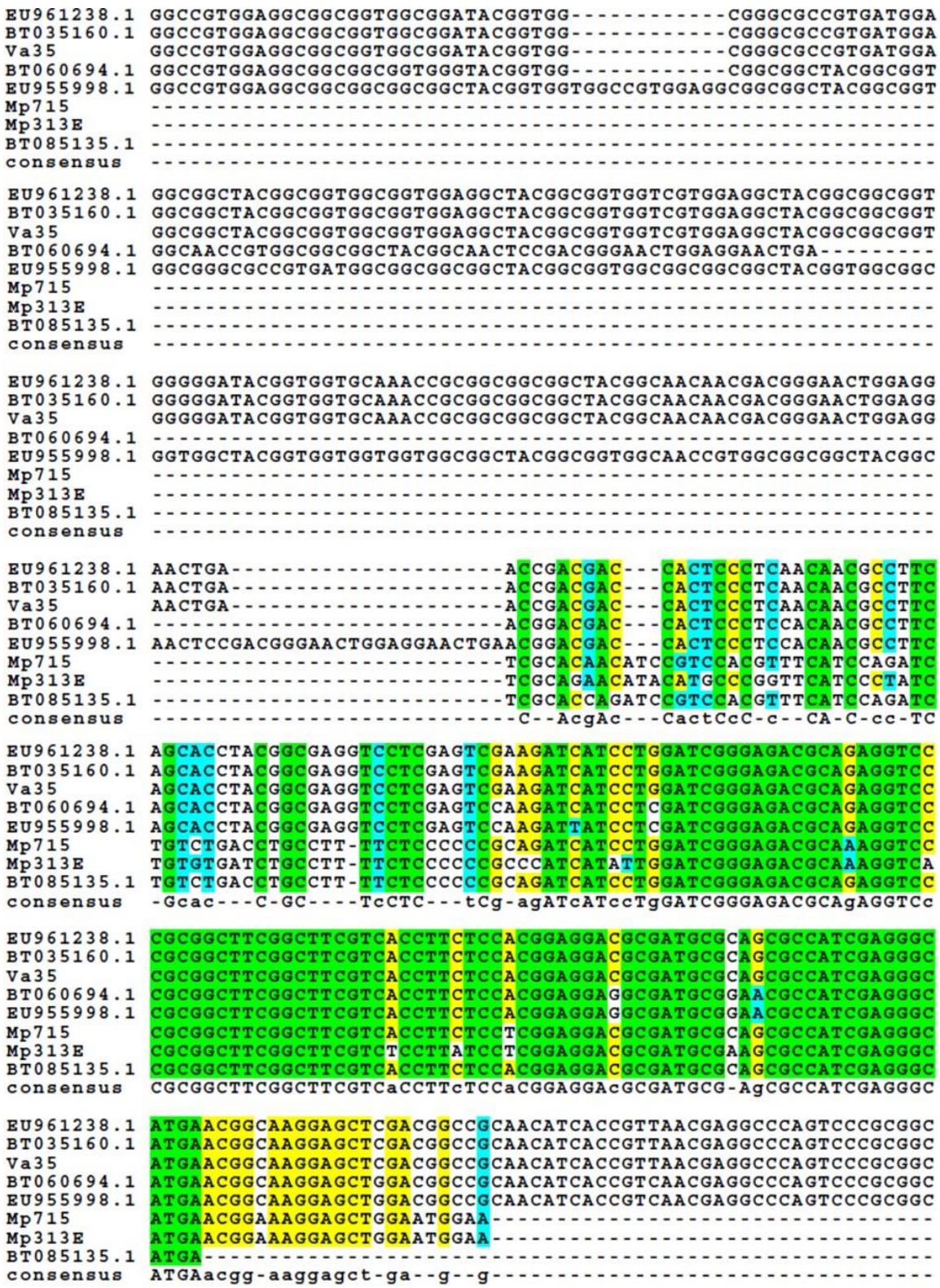

Figure 2.1 (continued)

```
EU961238.1 GGCCGTGGAGGCGGCGGTGGCGGATACGGTGG------------CGGGCGCCGTGATGGA
BT035160.1 GGCCGTGGAGGCGGCGGTGGCGGATACGGTGG------------CGGGCGCCGTGATGGA
Va35       GGCCGTGGAGGCGGCGGTGGCGGATACGGTGG------------CGGGCGCCGTGATGGA
BT060694.1 GGCCGTGGAGGCGGCGGCGGTGGGTACGGTGG------------CGGCGGCTACGGCGGT
EU955998.1 GGCCGTGGAGGCGGCGGCGGCGGCTACGGTGGTGGCCGTGGAGGCGGCGGCTACGGCGGT
Mp715      ------------------------------------------------------------
Mp313E     ------------------------------------------------------------
BT085135.1 ------------------------------------------------------------
consensus  ------------------------------------------------------------
```

Figure 2.1 (continued)

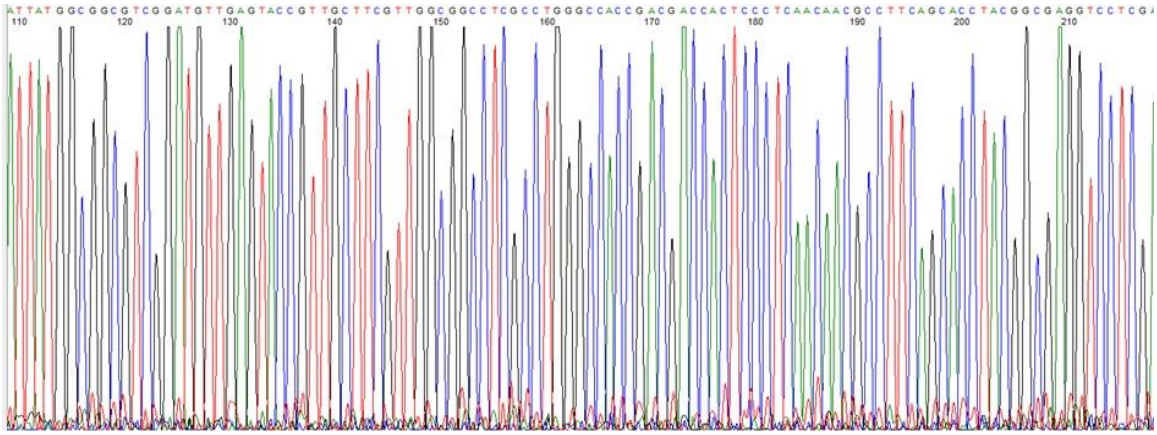

Figure 2.2 Nucleotide sequence chromatogram of gene GRBP in maize inbred line Va35.

Figure 2.3 A phylogenetic tree constructed using Biology Work Bench (Phylip rooted tree- Phenogram) for GRBP genes.

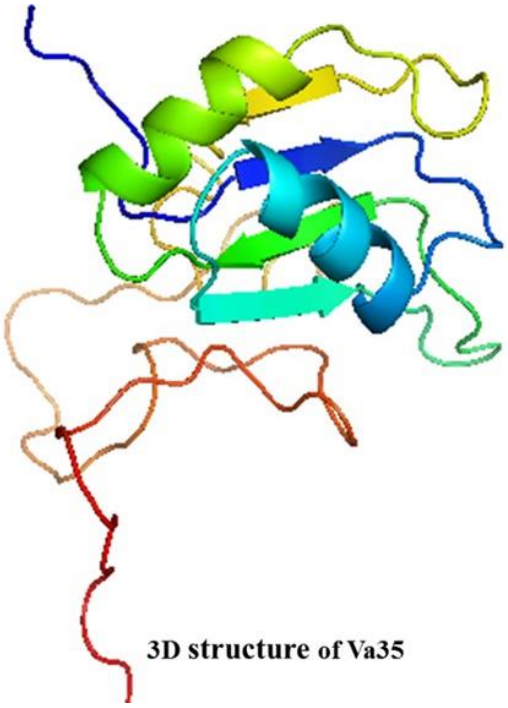

Figure 2.4 3D model of the GRBP protein in susceptible maize inbred line Va35.

The 3D protein structure was constructed by Modeller v 9.13 and visualized by PyMOL v3.1.0.

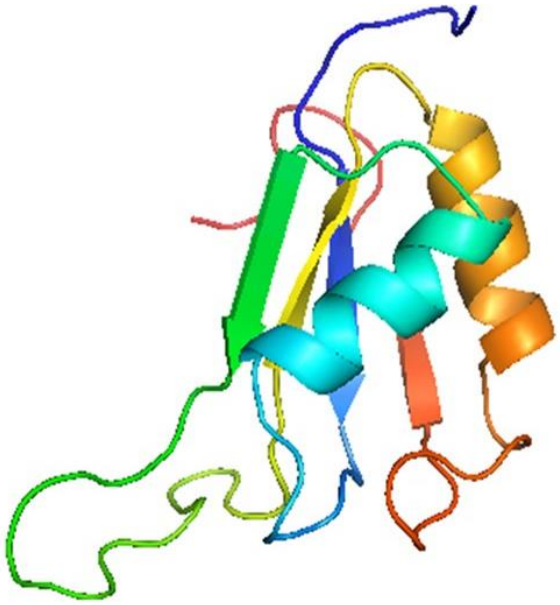

Figure 2.5 3D model of the GRBP protein in resistant maize inbred line Mp715.

The 3D structure was constructed by Modeller v9.13 and visualized by PyMOL v3.1.0

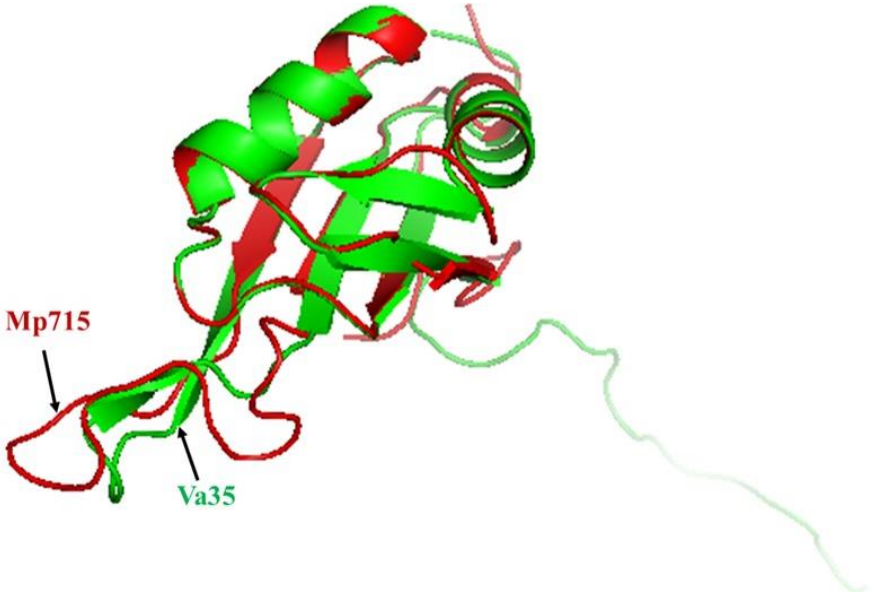

Figure 2.6 Alignment of 3D models of the GRBP protein of the susceptible maize inbred line Va35 with that of the resistant maize inbred line Mp715.

Green color represents Va35 and red color represents Mp715. Three dimensional structure alignment was constructed by PyMOL v3.1.0.

Table 2.2 Representation of DOPE and GA341 scores of models for the GRBP protein in susceptible maize inbred line Va35

Va35	Molpdf	DOPE score	GA341 score
Va35.B99990001.pdb	565.79291	-7907.72021	1.00000
Va35.B99990002.pdb	559.11224	-8072.19092	1.00000
Va35.B99990003.pdb	582.92566	-8035.28613	1.00000
Va35.B99990004.pdb	577.73541	-7784.59961	1.00000
Va35.B99990005.pdb	579.21600	-7941.20605	1.00000

On the basis of lowest DOPE score (-8072.19092) and highest GA341 scores (1.00000) Va35.B99990002.pdb was selected as the best model among five models.

Table 2.3 Representation of DOPE and GA341 scores of models for the GRBP protein in resistant maize inbred line Mp715

MP715	Molpdf	DOPE score	GA341 score
MP715.B99990001.pdb	745.29187	-6286.53125	0.26712
MP715.B99990002.pdb	778.20740	-6594.04102	0.22658
MP715.B99990003.pdb	743.55212	-6559.12305	0.20268
MP715.B99990004.pdb	858.77148	-6579.29102	0.16105
MP715.B99990005.pdb	717.34637	-6508.47070	0.35049

On the basis of lowest DOPE score (-6594.04102), Mp715.B99990002.pdb was selected as the best model among five models.

Discussion

Amino acids are the building blocks of proteins and the sequence of a polypeptide uniquely determines protein 3D structure. During evolution protein structures remain conserved while protein sequences changes in a much faster rate. Proteins with similar

sequences likely have similar structures. However, distantly related sequences may also fold into the similar 3D structures like closely related sequences, which ultimately proves the evolutionary theory (Krieger et al., 2003). For these reasons, protein 3D structure determination becomes an important approach to identify protein functions and protein-protein interactions, because structure and function are more conserved than sequences themselves. Protein 3D structures are very useful in the area of structure based drug design (to develop a novel drug), rational design of more stable proteins and in molecular modeling studies. Experimental methods, such as X-ray crystallography and NMR spectroscopy, are the very effective techniques to obtain the structure of proteins (Martí-Renom et al., 2000). However, size of proteins and difficulty in protein crystallization can pose great problem to employ these techniques in the protein structure determination.

Comparative modeling or homology modeling methods substitute experimental techniques and solve the problems persistent in these techniques. The basic principles behind comparative modeling techniques are the alignment of query (structure unknown) protein with template (structure known) protein sequences. Modeller relies on protein data bank (PDB) or other databases such as Swissprot for the selection of proteins with known structures.

Mainly three major steps are used by comparative modeling methods such as alignment of query with template, model building and evaluation of models (Madhusudhan et al., 2005).

First step used by Modeller program was the alignment of query with template sequence. This is one of the crucial steps in 3D structure prediction. The sequence identity between query and template must be more than 30% to perform best alignment.

Sequence identity less than 30% can cause errors and gaps in the alignment step (Madhusudhan et al., 2005). We have incorporated template sequences with identity greater than 30% to the target sequence. Templates having zero E-value were considered as best templates for the alignment purpose. Sometimes these methods can lead to generate errors and gaps in model building because ranking based on E-value does not necessarily ensure that the selected template is the best one. So, consideration of structural and functional features are also important in the selection of best template.

Selection of multiple templates are also recommended in comparison to selection of single best template to increase model accuracy (Sanchez and Sali, 1997; Floudas, 2007a). However, this is not always the true case because incorporation of multiple templates can also lead to production of inaccurate models (Larsson et al., 2008). The selection of single best template vs multiple templates intuitively depends upon users. Other approaches such as *ab initio* do not require any template sequence. Hence these methods are not as reliable as the template based method for protein 3D structure prediction (Liu et al., 2011).

Query and template alignment uses dynamic programing algorithm and a variable gap penalty function. In this command, gap penalty does not affect helices and sheets, buried regions and distance residues (Sanchez and Sali, 1997). Errors can occur at the time of modeling when query sequence does not match with template sequence. This can lead to misalignment. The advantage of this technique is that it needs only minimum level of sequence similarity between query and template rather than 100%. Programs such as CASP, CAFASP, Live Bench (Bujnicki et al., 2001), and EVA (Eyrich et al., 2001) have been developed by researchers to overcome the misalignment problem. EVA

is the most widely used web server for evaluating blind predictions from prediction servers and can access large set of data to check misalignment.

Various approaches have been developed for model building such as modeling with rigid-body assembly, modeling by segment matching and modeling by satisfying spatial restrains (Eswar et al., 2006). The Modeller program calculates distance and dihedral angle restrains at the time of model building to satisfy spatial restrains. It is one of the most promising and preferred method used in comparative modeling studies. Any method can provide best models if used optimally.

In this study, 3D models of Mp715and Va35 were selected on the basis of lowest DOPE (Discrete Optimized Protein Energy) and highest GA341 scores (Pieper et al., 2011) as shown in the Table 2.2 and Table 2.3. GA341 scores mostly are found in the range of 0 to 1. DOPE scores basically highlight the correct and incorrect aligned regions. Mp715.B99990002.pdb was selected as the best model having lowest DOPE score (-6594.04102) among five models. Va35.B99990002.pdb was selected as the best model having lowest DOPE score (-8072.19092) and highest GA341 scores (1.00000) among five models. The purpose of this last step was to identify the best model among all the alternative conformations generated by Modeller program that are closest to the native 3D structure of protein. The accuracy and validity of those models can also be checked by some of the structure quality scores or energy functions which can better estimate the quality of those models. These scores can easily distinguish correct 3D structure from rest of the protein structures. The quality of models can also be checked by model quality assessment program (MQAP) and SPICKER (Roy and Zhang, 2012) which choose the structures in a hierarchical way. Another programs such as PROCHECK, AQUA and

SFCHECK check bond length, bond angles and peptide bonds in the models to evaluate model quality (Liu et al., 2011).

Application of 3D models and their biological usefulness is entirely dependent upon accuracy in 3D model building and in homology approach. These 3D protein structures can open the door for the study of mutations, identifying SNPs, in drug designing and identifying unknown similar protein structures. Homology modeling approach can provide a clue of protein structure without involving any experimental methods. Not only high resolution models but also low resolution models roughly in the range of (2-5 A^0) can also provide useful information such as identifying active binding sites of protein and disease associated mutations.

In this study, high resolution models (1.0 A^0) of selected proteins in maize inbred lines were generated through Modeller program which provides greater understanding of 3D structure of proteins and the difference in their function. It was found that loop region of 3D structure placed in result section has great difference between two maize inbred lines. Comparative modeling or homology modeling studies are also effective in identifying resistant related proteins to aflatoxins and can also predict 3D structures of novel candidate proteins in maize inbred lines. This approach is easy to follow and provides accurate results.

CHAPTER III

COMPUTATIONAL PROGRAM FOR QUANTITATIVE ANALYSIS OF 2DE PROTEIN GEL IMAGES ON IDENTIFICATION OF DIFFERENTIALLY EXPRESSED PROTEINS ASSOCIATED WITH RESISTNACE TO AFLATOXIN ACCUMULATION IN MAIZE

Abstract

Two-dimensional gel electrophoresis (2DE) has been widely used in proteomic studies for revealing differences in proteomic profiles expressed in tissues of interest under different treatments. However, the limitations such as gel to gel variation and lack of an effective quantification method to digitally analyze all the protein spots across multiple gels have limited the adaptation of this technique as a quantitative tool. In this research, we established a working protocol for detection and quantification of protein spots from multiple maize 2DE protein gel images with Matlab Image Processing Toolbox. Developing kernels from two resistant (Mp715 and Mp719) and two susceptible (Va35 and Mp04:87) maize inbred lines were used in this research to identify differentially expressed proteins associated with resistance to *Aspergillus flavus* infection and aflatoxin accumulation. Protein extraction was performed using TCA/acetone precipitation in combination of a phenol extraction step. The protein gel electrophoresis was performed using PROTEAN IEF Cell (Bio-Rad) and PROTEAN II XL cell (Bio-Rad). Gel images were obtained with an Alpha imager. Matlab image processing tools

were used to analyze all the 2DE protein gel images in the experiment simultaneously. The 2DE protein gel images were first aligned and cropped. A total 12 protein gel images were used to compute mean gel image. Mean gel image was used for image segmentation and detection of protein spots. Differentially expressed maize proteins associated with resistance to *Aspergillus flavus* infection and aflatoxin accumulation were identified through principal component analysis of the mean intensities of selected protein spots. This approach will provide a means of high throughput quantitative proteomic studies with large amount of 2DE protein gel images.

Keyword *Aspergillus flavus*/Two dimensional gel electrophoresis/ Matlab/PCA

Introduction

Genomics and proteomics are the two very important approaches complimenting each other. Genomics deals with the analysis of genome sequences and proteomics deals with the analysis of large scale of proteins present in different physiological conditions. Major area of proteomics deals with the separation of mixtures of proteins and their further identification. For this reason, a broad range of techniques have been developed such as two dimensional gel electrophoresis (2DE) for protein separation and mass spectroscopy (MS) based techniques for protein sequence identification. 2DE is a traditional technique in proteomics which identifies thousands of proteins in a polyacrylamide gel (Jungblut and Thiede, 1997; Hames, 1998) based on their isoelectric point (pI) and molecular mass (mw). Each protein spot separated uniquely on 2D gels which is visualized by staining with Coomasssie Brilliant Blue or florescent stains (Gauci et al., 2011). Proteomic differences present in different samples can be easily identified across the gels through this method.

A major bottleneck that persists in 2DE is the gel image analysis. Several commercially available software packages such as PDQuest, Dymension, Image Master, Protein Mine, and Delta 2D (Clark and Gutstein, 2008; Mehl et al., 2012) are available for spot detection and spot matching. However, to properly use these software packages user need considerable efforts. The user has to perform substantial manipulation at many stages such as proper input of parameters, selection of threshold values and types of normalization according to their need (Daszykowski et al., 2009). The major difficulties occur in image processing are the spot detection (a step of separating actual protein spots from the background), spot matching, and quantification of protein spots. Spot matching generally involves matching of a spot in a gel to the corresponding spot present in the other gels. This way spots corresponding to the same protein can be identified. The problem arises when large number of gel images are involved in a study, only few spots match very well in the spot matching step. Further statistical analysis becomes very challenging with this scattered data. These software packages can intensify the variability created by the experimental procedure of 2DE (Wheelock and Buckpitt, 2005). In reality, software induced variance can be much higher than the variance present in the data itself in classical statistical analysis tests such as t-test and F-test (Li and Seillier-Moiseiwitsch, 2011). This variances increase as the number of the gels increases in the study. The analysis of multiple gel images became a challenging task with these software packages. Advancement of these software packages and their automated analysis still could not provide the best solution in the 2DE gel image analysis. Further developments are needed to provide the best solution in this area.

Matlab image processing tool box can be a preferred method for many users when comparing it to currently available software packages: This is due to its user friendly syntaxes and advanced statistical tools make it a complete package in one platform. Matlab is a high-level language and an interactive software package developed by MathWorks and released in 1984. It supports high level languages such as Fortran, C and C++ (Varjo, 2014). Image Processing Toolbox™ (IPT) provides various algorithms for multiple functions such as image visualization, denoising of images, image enhancement, image segmentation, image transformation and statistical analysis (Gonzalez and Woods, 2002). Matlab image processing toolbox can analyze grey scale as well as colored (RGB) images.

The workflow of image analysis contains mainly five broad steps: alignment of the gel images based on land marks, mean gel image construction, denoising through filters, segmentation, and finally statistical analysis. Detection of each spot, matching of one spot to the same spot in other gels and quantification of protein spots are extremely important steps in image analysis. Proper detection of protein spots is highly recommended because improper detection can miss differentially expressed proteins or not detect significant difference. For this reason, Matlab image processing toolbox™ was used to acquire a mean gel image, average of pixel intensities across all the gels used in the experiment. Protein spot alignment and incorporation of mean gel image is an effective way to reduce variations in data analysis (Li and Seillier-Moiseiwitsch, 2010). Mean gel image can be selected for segmentation purpose and for further analysis.

The purpose of segmentation is to remove artifacts generated in 2DE experiments, such as cracks, fingerprints, distorted gel shapes and low-grey-level regions (Rye and

Alsberg, 2008). Several watershed algorithms has been developed in recent years for image segmentation. However, classical watershed can create over-segmentation of gel images having tiny and shallow watershed instead of deep ones. The only solution available for this problem till now is the use of marker controlled watershed segmentation (Bleau and Leon, 2000). Markers are mainly used to identify low-grey-level regions and utilize homotopy image modifications for proper segmentation. Segmentation procedure alone is not sufficient to remove these artifacts. Combination of segmentation and spot filtering method can make it better. Spot filtering method usually assigns a threshold limit (Rye et al., 2008) for protein spots. Selection of threshold value is totally dependent upon users, but too low and too high values can cause addition or removal of protein spots, respectively (Rye and Alsberg, 2008). Gaussian filter is a general spot filtering method applied for spot filtering because protein spots mostly contain a Gaussian shape and deviation of this can leads to detection of false protein spot (Kimori et al., 2010).

In this study, an optimized Matlab based 2DE gel image processing methodology was developed to identify differentially expressed proteins in two resistant (Mp715, Mp719) and susceptible (Va35 and Mp04:87) maize inbred lines. Through this method, we have compared spot intensities across the gel images which ultimately were used to compare the expression levels of same proteins in susceptible and resistant maize inbred lines.

Material and Methods

Sample collection

In order to perform 2DE gel electrophoresis, resistant maize inbred lines (Mp715, Mp719) and susceptible maize inbred lines (Va35, Mp04:87) were obtained from Corn

Host Plant Resistance Research Unit (USDA-ARS-CHPRRU) at Mississippi State University. All maize lines were planted at the R. R. Foil Plant Science Farm at Mississippi State. For proteomics study, maize kernels were collected 14 days after inoculation with *A. flavus*. Kernels were collected and immediately frozen in liquid nitrogen and stored at -80°C for further analysis.

Protein extraction, solubilization and measurement of concentration

TCA-acetone extraction

Kernels were ground into a fine powder in liquid nitrogen using pre-chilled mortar and pestle. We have tried three different protein extraction methods in our study and finally adopted the TCA-Acetone with phenol extraction method. We have modified the method of Damerval, et al. and used in this study. In this method, approximately 500 mg of kernel powder was transferred in a 2 ml tube and 1.8 ml of cold TCA-2ME was added. Sample was vigorously mixed and incubated for 1h and then centrifuged for 15 min at 12,000g (below 4^0C). Supernatant was discarded and pellet was resuspended in 1.8 ml of cold rinsing solution (2-ME-acetone). Sample was stored -20^0C for 1h and then centrifuged for 15 min at 12,000g (below 4^0C). Supernatant was discarded and pellet was dried under air for 1 h. Solubilization buffer of 100 µl containing 9.5 M urea, 5 mM K_2CO_3, 1.25 % SDS, 5% DTT, 6% Triton X-100, 2% ampholines (pH 3-10) was used for dissolving pellets. Protein pellet was mixed vigorously and centrifuged for 15 min at 10,000 g at 25^0C. Sample was stored at -20^0C until use.

Phenol extraction

Protein was extracted by grinding tissues into a fine powder in liquid nitrogen using pre-chilled mortar and pestle. We have modified the method of Pelpor, et al. (Faurobert et al., 2007) according to our need. In this method, approximately 200mg-400 mg of powder was transferred in a 2 ml tube and 1.8 ml of extraction buffer (500 mM Tris-HCL, 50 mM EDTA, 700 mM sucrose, 100 mM KCl (pH 8.0)) was added. Tube was vortexed, and incubated on ice for 10 min. An equal volume of Tris buffered phenol was added and incubated in shaker for 10 min at RT. It was centrifuged for 10 min at 12,000 g at 4^0C. After centrifugation phenol layer was separated carefully and back extracted with 1.8 ml of extraction buffer. Sample was mixed again, vortexed, and centrifuged for 10 min at 12,000 g at 4^0C. Phenol phase was recovered again carefully. Precipitation solution (0.1 M ammonium acetate in cold methanol) of 1.8 ml was added in recovered phenol solution. Tube was mixed well and incubated overnight at -20^0C. Proteins were pelleted by centrifugation for 10 min at 12,000 g at 4^0C. Protein pellet was washed with cooled acetone solution. Five minute centrifugation step was performed at each step of wash. Pellet was dried under air and solubilized in IEF (9 M urea, 4% CHAPS, 0.5% Triton X-100, 20 mM DTT, 1.2% Pharmalytes pH 3-10) buffer. The sample was incubated for 1 hour at RT under agitation and stored at -20^0C until use.

Protein extraction in combination of phenol and TCA-acetone

Proteins were extracted via TCA/acetone precipitation in combination of a phenol extraction step (Wang et al., 2006). The method was modified according to our need. The fine powder of 500 mg sample was put in a 2 ml tube and it was filled with 1.8 ml of 10% TCA/acetone solution for overnight incubation at -20^0C. Sample was vortexed vigorously

and mixed well and centrifuged (16,000g, 10min) at 4^0C. After the centrifugation the supernatant was discarded. Tube was again filled with 80% acetone, vortexed vigorously and mixed well. Sample was centrifuged for 10 min (16,000g) at 4^0C. Temperature was kept constant at 4^0C for these steps. Supernatant was discarded and sample was kept in room temperature (RT) and air dried briefly to remove excessive acetone. Once the sample was air dried 1:1 phenol (pH 8.0, Sigma)/ SDS buffer was added (approximately 500 µl phenol and 500 µl SDS). Sometimes more phenol was added in Phenol/SDS buffer for proper separation of two phases. Sample was mixed and incubated for 1 hour and then centrifuged for 15 min (16,000g) at 4^0C. The upper phenol layer (approximately 200-400 µl) was transferred in to a new 2 ml tube and filled with 1.8 ml of methanol containing 0.1M ammonium acetate solution. The sample was stored at this step for 10 min or overnight at -20^0C. Again the sample was centrifuged for 15 min (16,000g) at 4^0C. The supernatant was discarded and protein pellet was washed once with 100% methanol and once with 80% cold acetone. During each wash step, protein pellet was mixed by vortexing and then centrifuged as above. The protein pellet was allowed briefly to air dry. Protein pellets were stored at -80°C until use. The protein pellets were solubilized with 200 µl of solubilization buffer containing (7.0 M urea, 2.0 M thiourea, 4% CHAPS, 1M DTT, 2% ampholites pH 3.0-10.0). Protein concentration was determined by Bio-Rad protein assay dye reagent concentrate at 595 nm.

Isoelectric focusing and second dimensional electrophoresis

Samples were thawed and applied to an immobilized pH gradient strip (Bio-Rad ReadyStrip™ IPG Strips, pH 3-10, 17cm) by overnight active rehydration at 50V. Isoelectric focusing was performed using Bio-Rad IEF Protean cell. Strips were

dehydrated for 16 h before isoelectric focusing for 20 min at 250V, 10000V for 2.5 h with linear gradient and then at 10,000V until it reached at 40000 V-h with rapid gradient. Proteins were separated according to charge in the electro-focusing system. After electrofocusing, the strips were either stored at -80°C or immediately put in equilibration buffer containing equilibration buffer I and II. IPG strip equilibration was performed by employing 4 ml of equilibration buffer I (6 M urea, 2% SDS, 0.375 M Tris-HCL (pH 8.8), 20% glycerol, and 2% DTT) and 4 ml of equilibration buffer II (6 M urea, 2% SDS, 0.375 M Tris-HCL (pH 8.8), 20% glycerol) for 1 hour with gel side up. After equilibration, IPG strips were dipped in a 100 ml of 1X Tris-glycin-SDS running buffer to remove excessive mineral oil.

The equilibrated IPG strips were sealed at the top of 1 mm thick (20 x 22.3 cm) SDS-PAGE gel and sealed with 0.5 % of low melting agarose with trace of bromophenol blue. By the use of forceps any air bubble beneath the strips were removed. The separation in the second dimension was performed by Bio-Rad Protean R Xi cell and PROTEAN II XL cell (Bio-Rad) on a 12 % SDS PAGE gel (Acrylamide/Bis-acrylamide (30%), 10% SDS, 1.5 M Tris-HCL, pH 8.8, 10% (w/v) ammonium persulfate, TEMED and ddH2O). A 5X SDS-electrophoresis running buffer (25 mM Tris, 192 mM glycine and 0.1% SDS, pH 8.3) was prepared for the gel run. Electrophoresis was carried out at 16mA for 30 minutes and followed by 24 mA for 5 hour for each gel until the bromophenol dye front reached at the bottom of the gel. All the gels for one comparative analysis were run under same experimental conditions to reduce variations caused by experimental errors.

Protein gel staining and imaging

Protein gel was stained overnight with Oriole^{TM} fluorescent gel stain (Bio-Rad) and then stored in water before imaging. Protein gels can be stored in the water or wrapped in a plastic wrap and were stored at -20°C until use. Protein gel images were taken by an Alpha Innotech MultiImager (Alpha Innotech, San Leandro, CA) system. All the gels were imaged at the same resolution and exposure to reduce any gel to gel variation. Gel images were stored in standard "tif" format and used for further analysis.

2DE gel image analysis

Image alignment

Gel image which contains maximum protein spots were selected as a standard gel image for the alignment of other gel images present in this study. Two landmarks were selected for the alignment purpose. Landmarks can be defined as the protein spots present across all the gels used for the alignment purpose. After alignment images were cropped to make them of equal size. Aligned and cropped images were saved in Matlab folder in standard 'tif' format.

Mean gel image construction

In this experiment, 16-bit intensity and 75 dpi resolution protein gel images were used. Twelve aligned and cropped gel images were used for mean gel image construction. Mean gel image was computed by averaging the intensities pixel-by-pixel across all the 12 protein gels in this work flow. Mean gel image was used for spot detection and for segmentation. We have performed denoising after construction of mean gel image. We have incorporated single filter and then multiple filters for denoising. Finally, two filters

were employed for denoising purpose. Log (Laplacian of Gaussian) transform with default values h size 5 and 0.5 for sigma, 6 pixel averaging and Top Hat Transform ('disk', 12) was applied to denoise the mean gel image.

Watershed Segmentation and PCA analysis

Marker-controlled watershed segmentation was chosen to reduce the over segmentation in the mean gel image. Watershed segmentation basically choose watershed ridge lines in an image. In the third step of this experiment, mean gel image (300×300 pixel width) was divided to five (150×150) pixel region and watershed algorithm (Vincent and Soille, 1991; Koyuncu et al., 2012) was applied for spot detection. The whole mean gel image was divided in to upper-left, upper-right, down –left, down- right and central area of the mean gel image. To identify more differentially expressed protein spots, the upper left region was further divided to nine (50×50) pixel regions and piece wise segmentation was performed. External minima and internal minima markers were chosen to do proper segmentation. Internal markers were chosen to decide the protein spot and to obtain watershed lines. These watershed lines were chosen as external markers for the background. The purpose of overall process was the partitioning of protein spots from the background. Regionprops function was used for the calculation of centroid, weighted centroid and mean intensity for the 12 protein gels.

In the final step of mean gel image analysis, principal component analysis (PCA) was performed in the upper-left region of mean gel image. Two-way analysis of variance (ANOVA) test was applied on the upper left region of the mean gel image which gives significant p-value. This test can provide the information of spots which are differentially

expressed in resistant and susceptible maize inbred lines. All program codes for analysis of gels were written in Matlab version 8.1.0.

Results

Two dimensional gel electrophoresis was performed to get 12 protein gels. All the gel images were aligned and cropped to make a mean gel image. Mean gel image covers all the protein spots present across twelve protein gels. Before segmentation, the Log and Tophat filtering methods were employed to denoise the mean gel image. Segmentation was then performed to identify boundaries of each single protein spot. Finally, PCA analysis was performed to identify differentially expressed proteins among susceptible and resistant maize inbred lines.

Mean gel image construction and spot detection

Twelve protein gel images were used to compute mean gel image and protein spots detected across the mean gel image. Total 31 protein spots were detected in the mean gel image in a 300 ×300 pixel intensity region without using any filter. To identify more spots we have used Tophat filter for better denoising of mean gel image. Total 80 protein spots were identified by using Tophat filter alone. Tophat filter provides better spot detection in comparison to no filter but still many spots were unidentified in the mean gel image. To identify those spots, Log (Laplacian of Gaussian) filter was employed. It identified total 573 protein spots in mean gel image. This filter identified more spots but over segmentation was also observed. To overcome the over-segmentation problem Tophat and log filters were applied together. In these steps no manual editing

was performed. With the use of image processing toolbox, the whole procedure was automated to avoid any user intervention.

A gray scale protein gel image of susceptible maize inbred line Va35 and resistant inbred line Mp715 are shown in Figure 2.7. The black color in gel image represents protein spots and white color represents the background. The gel images were inverted for visualization purpose only. These images were not incorporated for image analysis. We have employed gel images with black background and white protein spots.

In Figure 2.9. A mean gel image is shown which was denoised by using Tophat transform filters. A default threshold was chosen for this purpose. In the Figure 2.10, internal minima markers (2.10A) and ROI (region of interest) function (2.10B) were chosen to calculate centroid locations to count the number of protein spots and total pixel intensity. Selection of markers depends upon the size of object. The effectiveness of this method basically depends upon exact selection of marker size. Figure 2.10C shows watershed segmentation applied in the mean gel image to get the binary image of segmentation.

Watershed segmentation

Watershed segmentation applied in 150×150 pixel region of the mean gel image is shown in Figure 2.10 A) The 300×300 pixel region of mean gel image was divided to five (150 x150) pixel regions; UL (upper left), UR (upper right), DL (down left), DR (down right) and CEN (central) regions to do better segmentation for these areas and for the identification of more protein spots. Total 64 protein spots were identified in the UL region and total 100 protein spots were identified in the UR region. Total 43 protein spots were identified in DL region, and total 80 protein spots were identified in DR region. In

the CEN region total 109 protein spots were identified. We further employed PCA analysis to identify differentially expressed proteins in these areas. We have successfully identified differentially expressed proteins in UL region. To further improve our segmentation step, upper left region was further piece wise segmented to nine 50× 50 pixel regions to get better segmentation Figure 2.10 B) and to identify more protein spots. In these nine regions, only UL regions (UL-1, UL-3, UL-5, UL-7 and UL-8) has very good segmentation, and total number of protein spots also increases in comparison to the segmentation of 150× 150 pixel regions. Total 20 protein spots were identified in the UL-1 region, 25 protein spots were identified in UL-3, 16 spots identified in UL-5, 9 spots identified in UL-7 and total 17 protein spots were identified in UL-8 region. For the better presentation of segmented areas in these five regions, binary images include upper left regions (UL-1, UL-3, UL-5, UL-7 and UL-8) were generated in (Figure 2.10C).

Multivariate data analysis

In the upper left region PCA analysis was performed to identify protein spots in the all four maize inbred lines. In this study, two resistant maize inbred (Mp715, Mp719) lines and two susceptible maize inbred lines (Va35 and Mp04:87) were clustered together in the upper left region of the mean gel image. PCA analysis was performed in UR, DL, DR and CEN, 150×150 pixel intensity region but clustering of resistant and susceptible maize inbred lines were not observed.

We further analyzed the UL region which has a good separation. Upper left region having (50×50) pixel shows separation of two resistant and two susceptible maize inbred lines. This separation confirms that the proteins belonging to this area are differentially expressed aflatoxin and resistance related proteins. Differentially expressed protein

regions are shown by an arrow in Figure 2.11. In this PCA plot scores of protein spots, 3, 13 and 24 corresponds to Va35, score 1 corresponds to Mp04:87, score 23 corresponds to Mp715 and scores 7, 11, 15, and 21 corresponds to Mp719. These differentially expressed protein spots are related to aflatoxin resistance which was confirmed by a two way anova (ANOVA) test to calculate the p-value >0.05 for each.

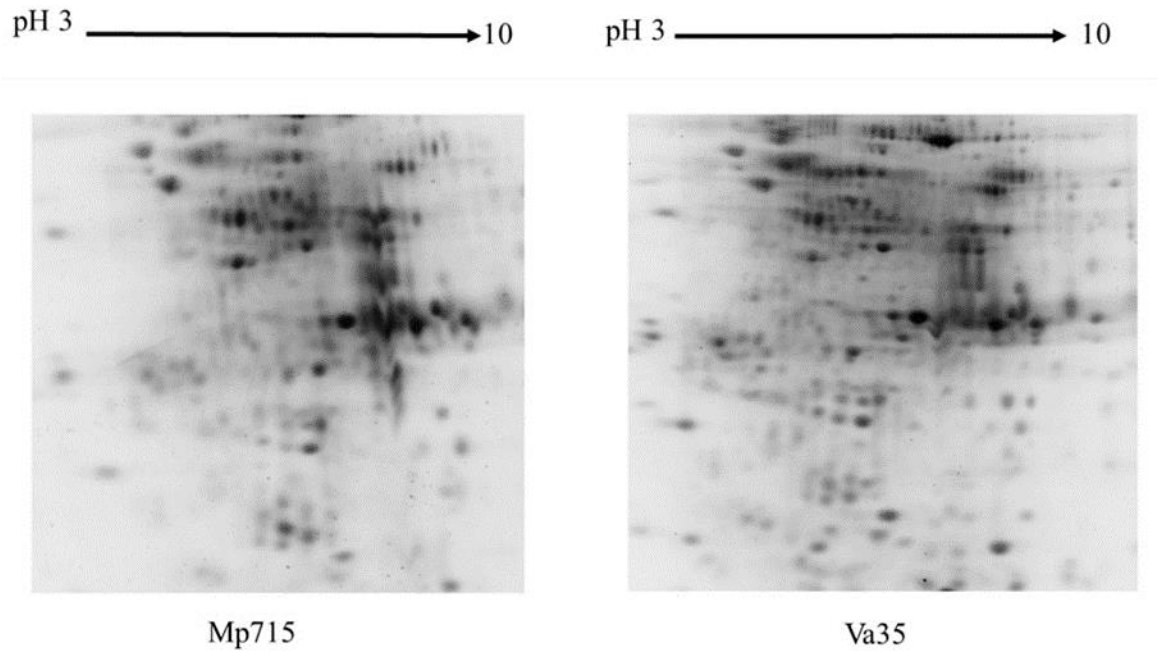

Figure 3.1 Gray scale 2D gel image of resistant inbred line Mp715 and susceptible inbred line Va35

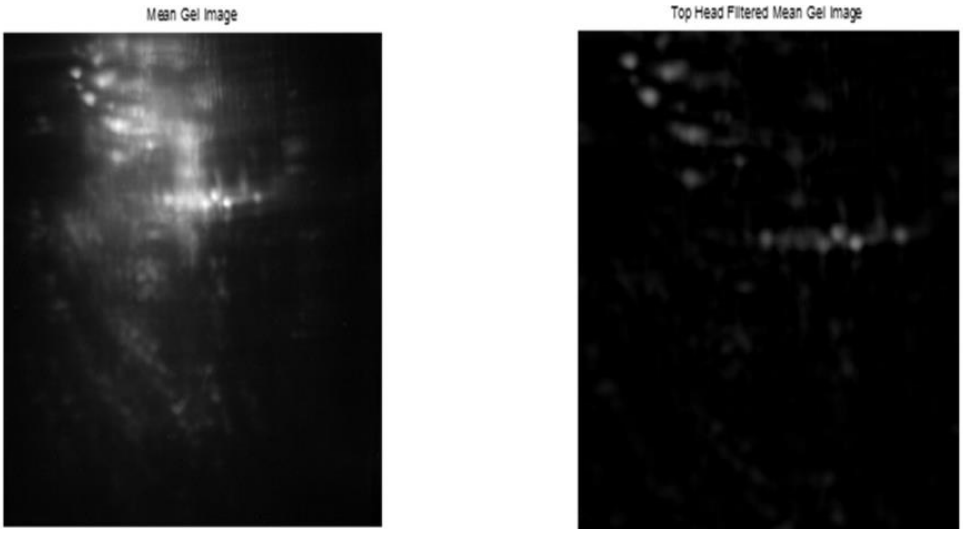

Figure 3.2 A) Mean gel Image. B) Top Hat Transform Image

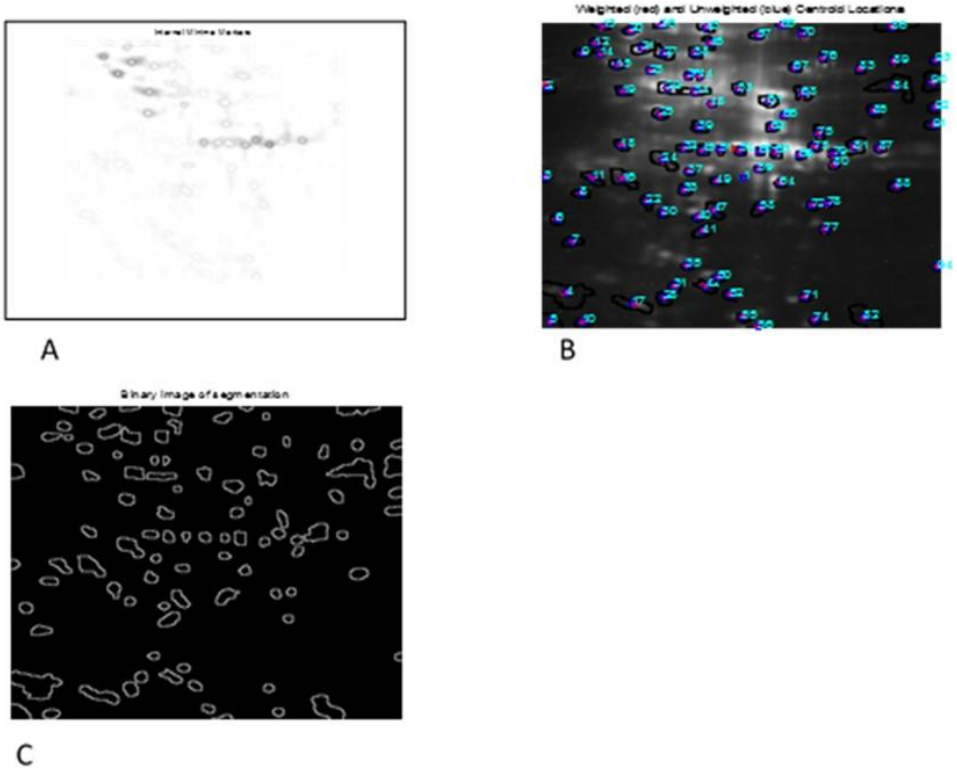

Figure 3.3 Internal minima markers and centroid locations.

A) Internal minima markers and B) ROI (region of interest) function was chosen to calculate weighted and unweighted centroid locations and C) watershed segmentation was applied in the mean gel image to get the binary image of segmentation.

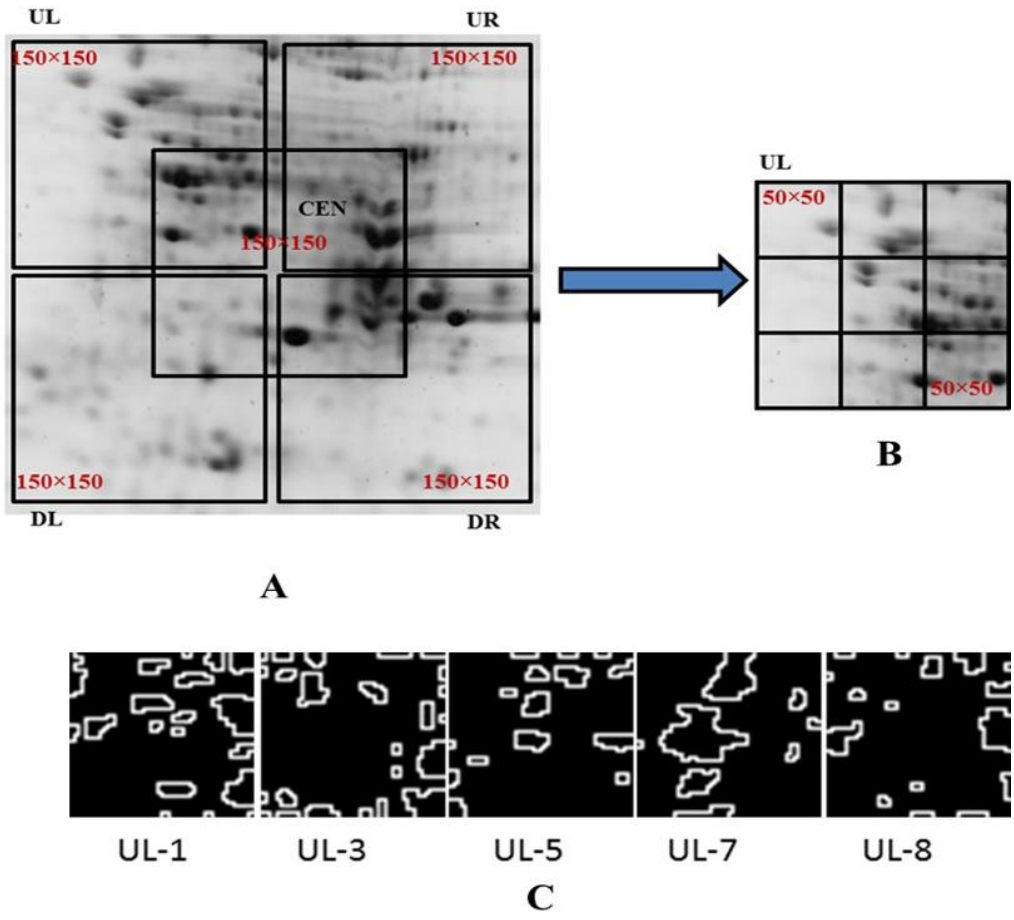

Figure 3.4 Images of watershed segmentation.

A) A presentation of watershed segmentation applied in 150×150 pixel region of mean gel image. The image is divided in to five regions UL (upper left), UR (upper right), DL (down left), DR (down right) and CEN (central) and in B) Upper left region was further piece wise segmented in to nine 50× 50 pixel regions C) Binary images are shown includes upper left region, UL-1, UL-3, UL-5, UL-7 and UL-8.

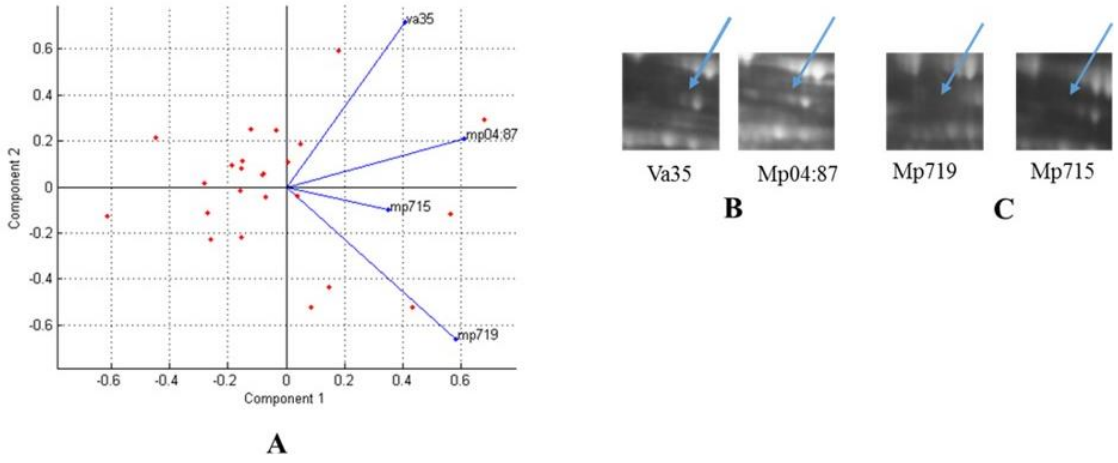

Figure 3.5 Principle component analysis of 50x50 pixel upper left region.

A) Principle Component Analysis (PCA) of upper left region 50×50 pixel of mean gel image of four maize inbreds (Mp04:87, Va35, Mp715 and Mp719). B) Shows two susceptible lines Va35 and Mp04:87 C) Shows two resistant maize inbred lines (Mp719 and Mp715). Area containing differentially expressed proteins are shown by arrows which was also proved by PCA analysis.

Improved segmentation method and results

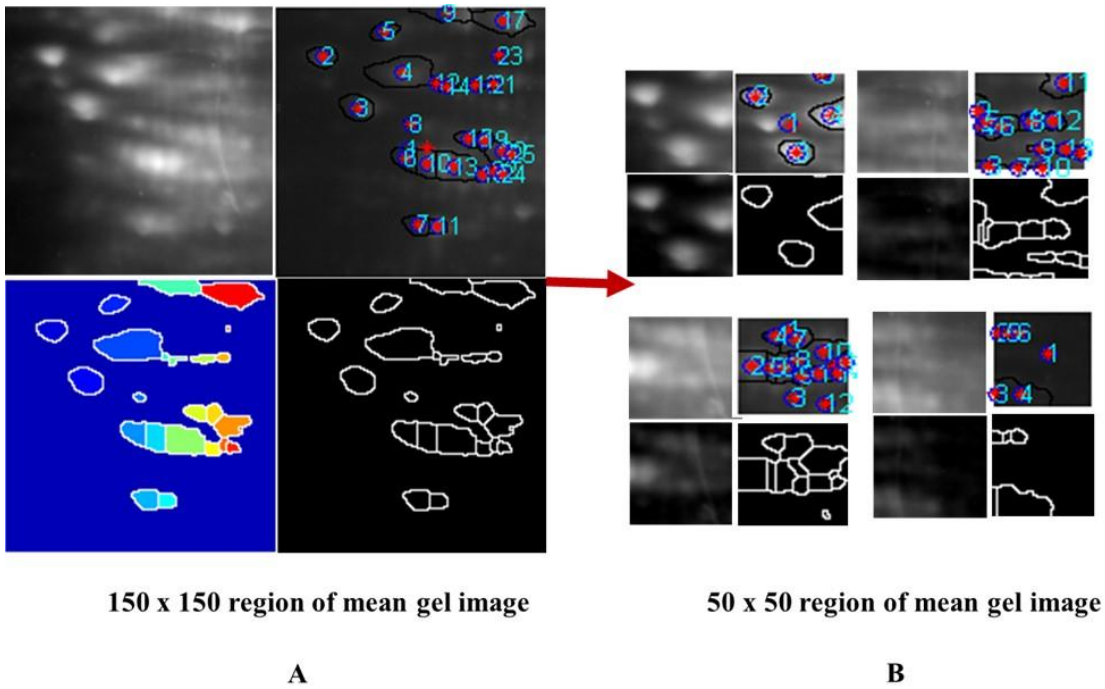

Figure 3.6 Improved segmented mean gel image.

Segmentation was first performed in the A) upper left (150x150) area of the mean gel image and then divided to four B) (50x50) pixel region to identify more protein spots.

The matlab code for image analysis further improved the identification of more protein spots. In the Figure 2.12A, total 25 protein spots were identified and in these four subdivisions, total 39 protein spots were identified as shown in Figure 2.12B. Ultimately piece wise segmentation can improve the number of protein spots and can also identify low intensity spots in a gel image

Multivariate data analysis

In the upper left region PCA analysis was performed to identify protein spots in the all four maize inbred lines. In this study, two resistant maize inbred (Mp715, Mp719) lines and two susceptible maize inbred lines (Va35 and Mp04:87) were clustered together in the upper left region of the mean gel image as shown in Figure 2.13A). This separation confirms that the proteins belonging to this area are differentially expressed aflatoxin and resistance related proteins. Figure (2.13B) labeled protein spots and their centroid locations are shown, (2.13C) differentially expressed protein spots are shown by arrow in each inbred line.

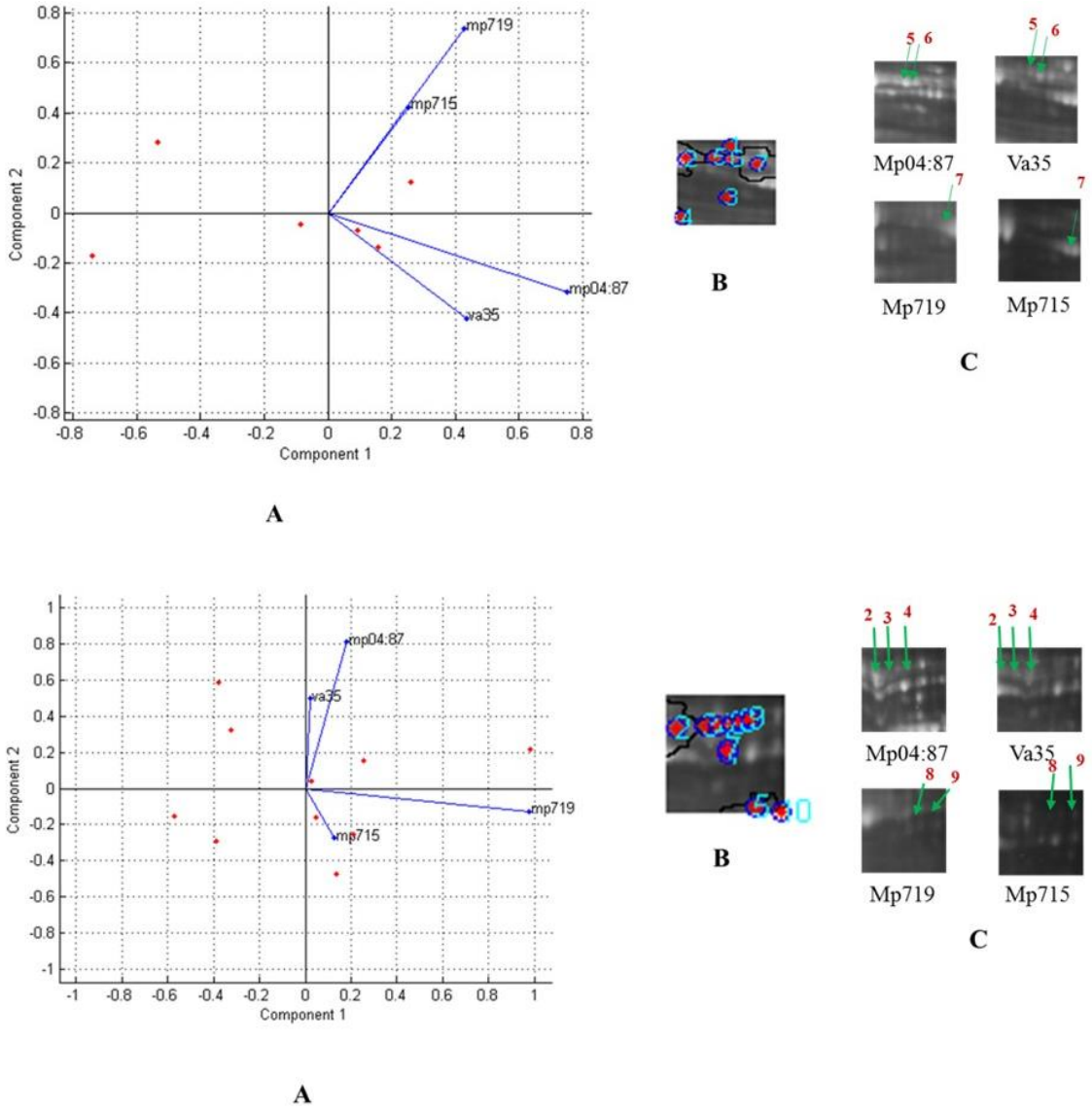

Figure 3.7 Principle component analysis of 50x50 pixel upper left region.

A) Principle Component Analysis (PCA) of upper left region 50×50 pixel of mean gel image of four maize inbreds (Mp04:87, Va35, Mp715 and Mp719). B) Labeled protein spots and their centroid locations C) Shows two susceptible (Mp04:87, and Va35) and two resistant maize inbred lines (Mp719 and Mp715). Area containing differentially expressed proteins are shown by arrows which was also proved by PCA analysis.

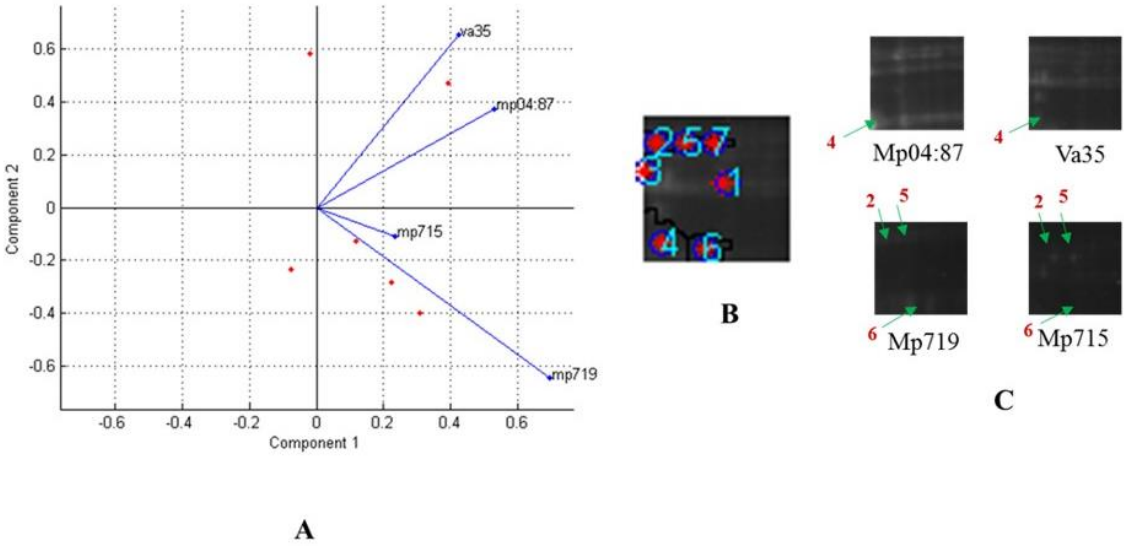

Figure 3.7 (continued)

Discussion

The goal of 2DE gel electrophoresis is to separate, identify and quantify the major differences between protein samples. Two DE gel can also provide information about protein modifications such as phosphorylation, glycosylation and changes in their expression level (Rabilloud et al., 2010; Magdeldin et al., 2012). Identification of changes in protein spot through 2D gel image analysis is a crucial step in proteomic studies. Manual analysis of protein spots through visual comparison is a difficult and time consuming process. Though several automatic commercial software packages are available, they could not provide the best results because of being expensive and including only very basic statistical tools (Li et al., 2011).

We have developed a new approach for 2D gel image analysis by using Matlab image processing toolbox. It is an efficient, accurate, less user intervention required, and automatic analysis method in comparison to other software packages. First step in image

processing is to select “standard gel image”. It is recommended that the standard gel image which has highest mean correlation compared to other gel images used in the experiment be chosen (Mehl et al., 2012). Maize inbred line Va35 replication 3 has been chosen as a “standard gel image” because of being a parental variety and having highest mean correlation compared with other gels used in this experiment. All the four maize inbred lines were aligned and cropped by a 300×300 coordinate frame. This is a very quick step taking approximately minutes to align and crop all 12 gel images. Two DE gel images contains many artifacts such as dust, distorted spot shapes, cracks and finger prints (Dowsey et al., 2010) which needs denoising steps to remove these artifacts. Log and Tophat filters were used to denoise these images. Both of them come under 2-D filter category that uses fspecial function which creates a two- dimensional fitter having specified values and parameters. Default values 5 and 0.5 were selected for these filters. Mean gel image was constructed after denoising step, employing all the 12 gel images. Construction of mean gel image is an important step to avoid spot matching. Methods which are based on pixel intensities provide huge background noise and sensitivity (Conradsen and Pedersen, 1992) and can create problems in spot matching step. Several other software packages use mean gel image to overcome with these limitations such as pinnacle (Morris et al., 2009) and RegStatGel (Li and Seillier-Moiseiwitsch, 2010). These software packages use mean gel image as a template to compare protein spots. Pinnacle uses fixed window and RegStatGel uses advanced statistical tools to analyze differences between protein spots. The commercial software packages use their own methods and algorithms to analyze protein gel images hidden from the users (Raman et al., 2002). The other drawback using these software packages are that the output results

may vary according to the selected package. It has been observed that software packages can generate more than 25% variance in comparison to data variance itself (Wheelock and Buckpitt, 2005). Comparison of many protein gel images still remains a challenge but Matlab Image Processing toolbox can provide better solution to this problem.

The amount of data produced by 2D gel electrophoresis is quite large and to define individual spots is a daunting task. Marker-controlled (Bleau and Leon, 2000) watershed algorithm was applied to overcome with over-segmentation problem which generally persists in classical watershed approach. Over-segmentation can increase higher numbers of tiny and shallow protein spots in 2DE gels. It can be effectively solved by deleting regions not corresponding to actual protein spots and creating marker regions. Each marker region corresponds to a particular watershed ridge and finally considered as a watershed ridge (Parvati et al., 2009).

This strategy provides the overall coverage of protein spots present in different gels. Sometimes, mean gel image contains a slightly larger and blurred protein spot centering about the locations of the same protein from all the gels due to slight misalignment of protein gels which can be removed by the master watershed map. Watershed segmentation was performed in the five 150 × 150 pixel UL, UR, DL, DR and CEN regions in this study. In these five regions, the upper left region shows significant differences in protein expression level as shown in Figure 2.11. The upper left region was again divided to nine 50 × 50 pixel region and piece wise watershed segmentation was applied. Piece wise segmentation was very effective in identifying protein spots which were not properly segmented in the 150 × 150 pixel region.

Multivariate data analysis was then performed in the upper left regions to identify differentially expressed protein spots. 2DE protein data analysis is a challenge step because multivariate data analysis has certain requirements for the data. The results may be misleading if requirement is not fulfilled (Gustafsson et al., 2004; Almeida et al., 2005). The possible requirements are distribution of measured values and their variance which is sometimes a big challenge in the case of 2DE data analysis (Meunier et al., 2005). Principle component analysis (PCA) is a mathematical procedure and a well-known multivariate analysis technique (Karthikeyan et al., 2012) which decompose matrix y into set of latent variables known as principle components. In this study, principle component analysis (PCA) has proved that proteins present in the upper left region of the gel image have significant differences in protein expression level and those spots contribute in resistance against *A. flavus* infection and aflatoxin accumulation. Differentially expressed protein spots are shown by an arrow in the Figure 2.12. These differentially expressed protein spots are related to aflatoxin resistance which was evaluated by a two way anova (ANOVA) test to calculate the p-value >0.05 for each. In summary, Matlab image processing toolbox is a highly advanced and quick analysis method of 2DE gel images which provides a platform to users for manipulations without much effort.

CHAPTER IV

TWO DIMENSIONAL DIFFERENCE IN GEL ELECTROPHORESIS (2D-DIGE) ANALYSIS ON IDENTIFICATION OF DIFFERENTIALLY EXPRESSED PROTEINS

Abstract

Traditional 2D protein electrophoresis technique can reveal thousands of protein spots in a single gel which contains one sample. A modified version of the 2D protein gel technique is the two dimensional difference gel electrophoresis (2D-DIGE) technique which allows running two samples in one gel, with each sample being labeled by different fluorescent dyes. The aim of this study is to apply 2D-DIGE technique on the identification of differentially expressed proteins associated with resistance to *Aspergillus flavus* infection and aflatoxin accumulation. Developing kernels from two resistant (Mp718and Mp719) and two susceptible (Va35 and Mp04:85) maize inbred lines were collected 14 days after inoculation. Total protein extraction was performed using TCA/acetone precipitation in combination with a phenol extraction step. Protein samples were labeled with fluorescent Cyanine dyes (Cy2, 3 and 5). Gel images were obtained using the Typhoon imager. Image analysis was performed by using Matlab image processing toolbox.

Keywords *Aspergillus flavus,* Difference in gel electrophoresis, Matlab, Cyanine dyes

Introduction

Traditional two dimensional gel electrophoresis (2DE) technique can resolve thousands of proteins in a single gel and it is very effective to identify differentially expressed proteins in diseased vs. non-diseased samples. Proteomics is a branch which deals with the study of proteins and their physiological state at a particular time in the cell. Cellular functions are mainly dependent upon the interplay between its proteins. Protein study has been always challenging due to its complexity and diversity in comparison to DNA and RNA. Protein diversity can be created by mRNA splicing, post translational modifications such as phosphorylation, alkylation, acetylation and sub-cellular localization (Edmond et al., 2011). These events could produce different number of proteins from a single mRNA. The beginning step of protein study is the separation and visualization of individual proteins.

One of the common and well-established methods of protein separation is two dimensional gel electrophoresis (2DE). It involves isoelectrofocusing (IEF) and sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) steps for the proper separation of proteins based on their isoelectric (pI) point and molecular weight (mw) (Sidman, 1981). 2DE is a very reliable technique to compare difference between replications of the same samples and relative abundance of the proteins present in different samples. However, traditional Two DE gel technique have limitations such as gel to gel variations observed due to inhomogeneity in polyacrylamide gels, thermal fluctuation during the run, pH differences in first dimension and post processing of gels (Wang et al., 2011).

Difference in gel electrophoresis (DIGE) developed by Amhersam Biosciences (Ünlü et al., 1997) aims to circumvent some of these problems. DIGE is a sensitive technique which can detect low amount (<0.5fmol) of protein in a sample and its detection range is also higher in comparison to traditional 2DE (Granlund et al., 2009). DIGE technique involves fluorescent cyanine dyes (Cy2, 3 and 5) for labeling of proteins. These dyes covalently attached to proteins and display different spectral peaks which provides the facility to run two samples in a single gel. Cydye minimal labeling and saturation labeling are the two labeling methods mostly used in DIGE experiments. Minimal labeling is the most stable method which primarily reacts with lysine amino groups and are also able to reduce multiple protein spots. The dynamic range of these dyes are high and they are more sensitive in comparison to other staining (silver) methods (Lilley and Friedman, 2004).

Minimal labeling often required an internal standard which consisting of pooled samples comprised of equal amounts of all samples present in the experiment. Internal standard is used for avoiding system generated variations, and it can be compared with other gels for matching protein patterns (Lilley and Friedman, 2004). It provides better resolution and high dynamic range in proteomic studies (Kimori et al., 2010). Unlike conventional 2DE technique, 2D-DIGE allows running two samples on a single gel. Multiplexing of diseased vs. non-diseased samples in a single gel offers simultaneous analysis of different samples. Another advantage of 2D-DIGE technique is that it skips the post staining procedure of the gels. Traditional 2DE technique always rely on staining of gels and uses coomassie blue, SYPRO ruby, colloidal CBB, or silver stain for the visualization purpose of proteins (Miller et al., 2006). Some of the stains employed in the

2DE have low dynamic range for the detection of smallest proteins or have poor sensitivity. 2D-DIGE technique potentially can detect small amount of proteins due to its high sensitivity and dynamic range (Hu et al., 2003). Combination of 2D-DIGE with MS based techniques can enhance detection limit and provide high throughput yield, which makes this technique a powerful tool for the differential analysis of the proteins.

Difference in gel electrophoresis approach has been used to identify proteins related to developmental defects, proteins related to metabolic pathways, and stress related proteins in plants. In maize the unstable factor for orange 1(Ufo1) in pericarp which is responsible for pleotropic growth and developmental defects was studied to check Ufo1 induced changes in proteome (Robbins et al., 2013). In another study, 2D-DIGE was used in proteomic analysis of *Saccharomyces cerevisiae* under metal stress (Hu et al., 2003). Information about major protein changes during drought, high temperature conditions, and aflatoxin contamination can be identified with this approach. Sufficient data related to protein changes during stress conditions and aflatoxin accumulations are lacking in maize plants.

The objectives of the current study was to identify differentially expressed proteins among resistant and susceptible maize inbred lines employing the technique 2D-DIGE. Two resistant (Mp718 and Mp719) and two susceptible maize inbred lines (Va35and Mp04:85) were used for this study. Two dimensional-DIGE technique was performed to obtain protein gel images using Typhoon TRIO Variable Mode Imager (GE Healthcare). Gel images were analyzed with ‘Matlab Image processing toolbox’ which provides diverse and efficient tools for image analysis.

Materials and method

Sample Collection

Resistant maize inbred lines (Mp715) and susceptible maize inbred lines (Va35) were obtained from Corn Host Plant Resistance Research Unit (USDA-ARS-CHPRRU) at Mississippi State University (Williams and Windham, 2001). All maize lines were planted at the R. R. Foil Plant Science Farm at Mississippi State. For proteomics study, maize kernels were collected 14 days after inoculation with *A. flavus* and three replications of each maize inbred line were collected. Kernels were collected and immediately frozen in liquid nitrogen and stored at -80°C for further analysis (Kelley et al., 2012).

Protein extraction in combination of phenol and TCA-acetone

Proteins were extracted using TCA/acetone precipitation in combination of a phenol extraction step (Wilkins et al., 1996). The fine powder of 1 g sample was placed in a 2 ml tube and it was filled with 10% TCA/acetone for overnight incubation. Sample was vortexed vigorously and mixed well and centrifuged (16,000g, 15min) at 4^0C. After centrifugation, supernatant was discarded. Tube was again filled with 80% acetone, vortexed vigorously and mixed well. Again centrifuged for 15 min, (16,000g) at 4^0C. Temperature was kept constant at 4^0C for these steps. Supernatant was discarded and sample was kept in room temperature (RT) and air dried. This process was performed briefly to remove residual acetone. Once the sample was air dried 1:1 phenol (pH 8.0, Sigma)/ SDS buffer was added (approximately 500 µl phenol and 500 µl SDS). Sometimes more phenol was added Phenol/SDS buffer for proper separation of two phases. Sample was mixed and incubated for 2 hour and then centrifuged for 15 min

(16,000g) at 4^0C. The upper phenol layer (approximately 200-400 µl) was transferred to a 2ml tube and filled with 1.8 ml of methanol containing 0.1M ammonium acetate solution. The sample was stored at this step for overnight at -20^0C. Again the sample was centrifuged for 15 min (16,000g) at 4^0C. The supernatant was discarded and protein pellet was washed once with 100% methanol and once with 80% acetone. During each wash step, the protein pellet was mixed by vortexing and centrifuged as above. The protein pellet was allowed briefly to air dry. Protein pellets were stored frozen at -80°C until use. Protein pellets were solubilized in 200 µl of buffer containing 7.0 M urea, 2.0 M Thiourea, 4% CHAPS, 30mM Tris pH 8.5. Protein concentration was determined by Bio-Rad protein assay dye reagent concentrate at 595 nm.

CyDye labeling for 2D-DIGE

CyDye DIGE fluors minimal dye was used for labeling of protein samples obtained from resistant and susceptible maize inbred lines. The NHS ester (N-hydroxy succinimidyl) group of CyDyes reacts with amino group of lysine to form amide bond. Three CyDyes; Cy3, Cy 5 and Cy 2 were chosen for minimal labeling (DeCyder™, Amersham Biosciences) reaction. Protein samples (50 µg) from Va35 and Mp719 were covalently labeled with 400 pmol (1 µL) of Cy3 dye and Cy5 dye, respectively. The reactions were allowed to proceed in the dark for 30 min on ice and were terminated by the addition of 1 µL of 10 mM lysine and samples were again incubated in the dark for 10 min at RT. Labeled samples were stored at -20^0C or used immediately for IPG strip rehydration.

Isoelectrofocusing and 2D gel elctrophoresis

To perform isoelectric focusing step, an equal volume of sample buffer was added into the labeled protein samples. A volume of 300 µl volume was maintained for 17 cm IPG strips. Samples were applied to an immobilized pH gradient strip (Bio-Rad ReadyStrip^{TM} IPG Strips, pH3-10,17cm) by overnight active rehydration at 50V and 23^0C. Isoelectric focusing was performed by Bio-Rad IEF protean cell. After active rehydration, isoelectric focusing was performed for 2 h at 250 V with linear gradient, 10,000 V for 4 h with linear gradient and then until it reached 80,000 V-hr. Proteins were separated according to charge in the electro-focusing system.

Second dimension

After electrofocusing experiment, the IPG strips were either stored at -80°C or immediately put in equilibration buffers. IPG strip equilibration was performed by employing 4 ml of equilibration buffer I (6 M urea, 2% SDS, 0.375 M Tris-HCL (pH 8.8), 20% glycerol, and 2% DTT) and 4 ml of equilibration buffer II (6 M urea, 2% SDS, 0.375 M Tris-HCl (pH 8.8), 20% glycerol, 1.5g idoacetamide) for 1hour. Separation in the second dimension was performed by Bio-Rad Protean R Xi cell and PROTEAN II XL cell (Bio-Rad) on a 12 % SDS-PAGE gel and run at 50 mA until the dye reached at the bottom of the gel. All gels for one comparative analysis were run under the same experimental conditions to reduce variation caused by experimental error. All the 2D gel images were stored in standard “tiff” format and used for further analysis.

Image Acquisition with Typhoon TRIO Variable Mode Imager

The 2 protein gels were scanned for Cy3, Cy5 labeled proteins using Typhoon Trio Variable Mode Imager (GE Healthcare). Cy3 labeled images were scanned using 532 nm excitation and 580BP30 emission filter; Cy5 images were scanned using 633 nm excitation and 670BP30 emission filter. All the gels were scanned with a PMT of 600.

Image analysis using Matlab Image Processing Toolbox

In this study, one resistant (Mp719) and one susceptible (Va35) gel images were selected for the analysis purpose. These two gel images were cropped before analyses to cover all valuable protein spots and to remove un-wanted. Mp719 gel image was cropped to make the dimension (379 x 288) which was supported by Matlab. Va35 was also cropped to make the dimension (392 x 288) to cover all the protein spots. Resolution of these gel images was 150 dpi.

Gel images were saved in standard 'jpg' format and image analysis was performed using Matlab image processing toolbox. Images were first read by using imread command of Matlab. Two images were changed to greyscale-level by employing rgb2gray command. Further analysis was performed in these images.

Images were denoised employing Log and Top Hat filters together. The default values selected for Log filter was (h 5 and sigma 1) and disk size for Tophat filter was 10. This parameter was applied on both images. Marker controlled watershed segmentation was then performed on the Mp719 and Va35. Finally a binary image was generated containing weighted and un-weighted centroid locations. All the selected protein spots were assigned with numbers for visual interpretation.

Results

In this study, we have employed DIGE technique for protein analysis. We have also compared DIGE with traditional 2DE and proposed its pros and cons. Only two images Va35 and Mp719 were used for the gel image analysis. The images of the two gels with fluorescence dyes Cy3-and Cy5 labeled samples are shown in Fig. 2. 14.

The difference between Mp719 (green) and Va35 (red) through this DIGE image was observed in Figure 2.18. Four Cy3 labeled protein spots were highly expressed in Va35. These protein spots are not present in Mp719. The clear visual interpretation of these protein spots can be seen in a gray scale image of Va35 and Mp719 as shown in Figure 2.15. The gray protein spots present in Va35 and Mp719 gel images corresponds to the volume of the proteins. We can identify the difference between these two images by comparing these protein spots.

Most of the protein spots are found in the acidic region of the gel roughly from pH 4 to pH 7. A total of 47 protein spots in Va35 and 43 protein spots in Mp719 were detected by using Matlab Image processing toolbox. These spots are assigned with numbers on Binary image as shown in Figures 2.16 and 2.17. The sliced images from upper left, upper right and middle regions having spots differently expressed in these two gel images are shown in the Figure 2.18. Spot numbers 21, 24, 29 and 35 are highly expressed in Va35 and spot number 12 and 13 are highly expressed in Mp719.

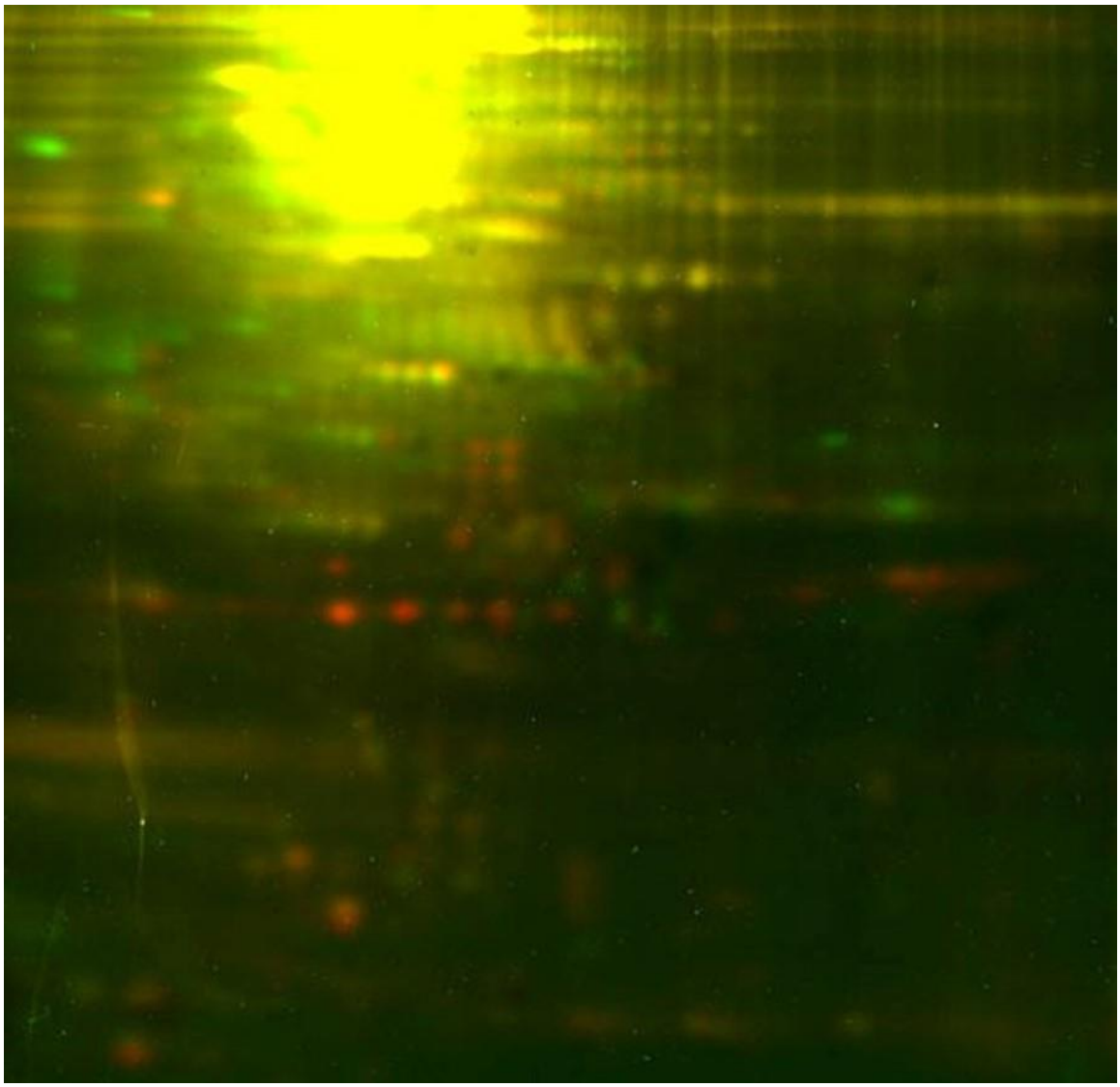

Figure 4.1 A DIGE gel image.

Va35 was labeled with Cydye 3 and Mp719 was labeled with Cydye 5. Red color represents Va35 and green color represents Mp719. The yellow color in this figure represent common protein spots present in both maize inbred lines.

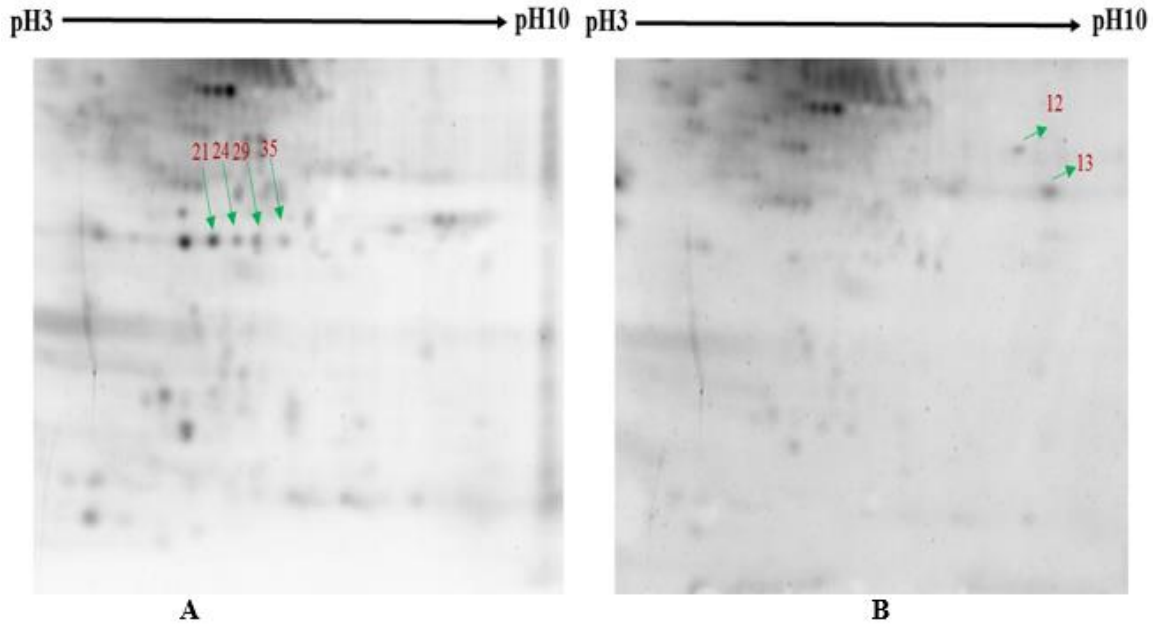

Figure 4.2 Gray scale image of **A**) Va35 and **B**) Mp719.

Gel images are shown from pH 3 to pH 10. Differentially expressed protein spots are pointed with arrow and labeled in red color.

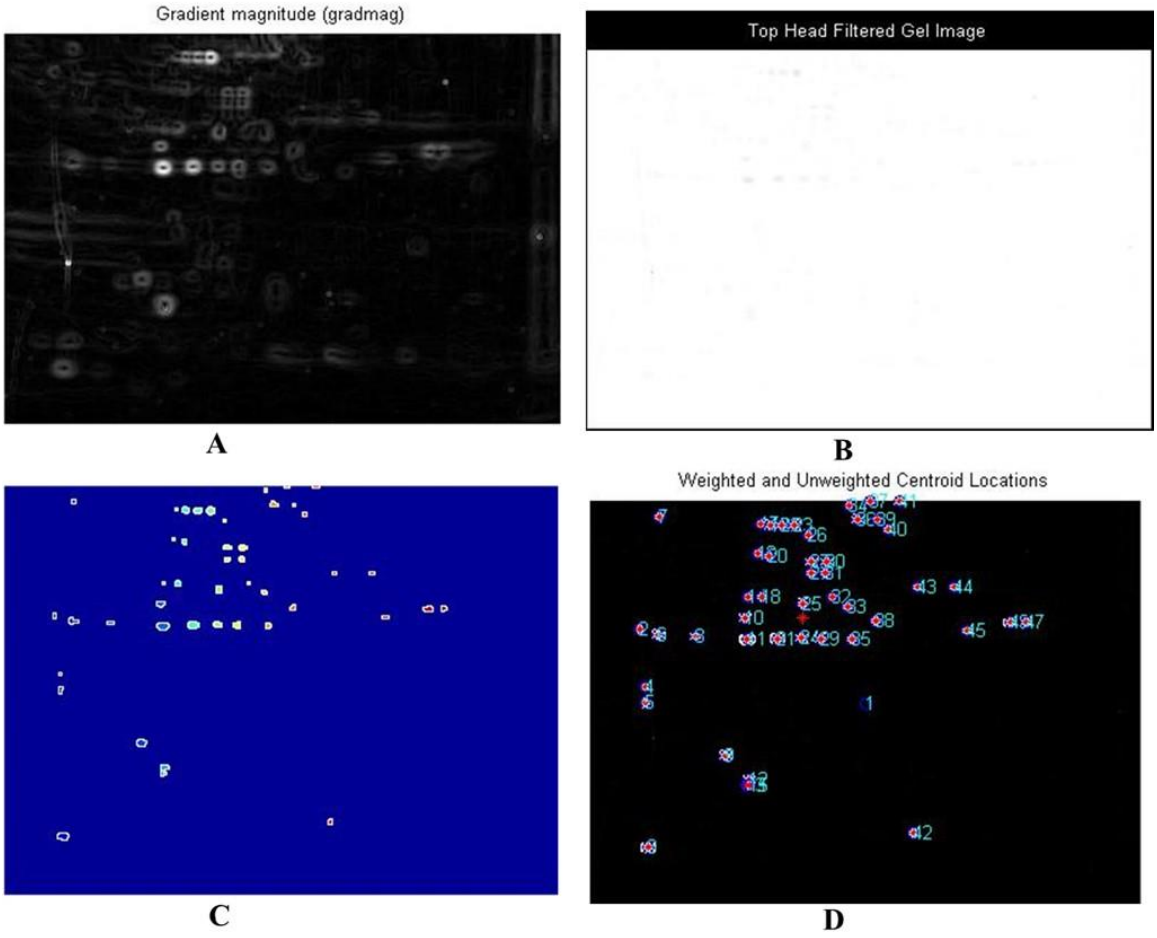

Figure 4.3 Overall presentation of segmentation.

A) Gradient magnitude image of Va35. B) Top Hat filtered image C) Segmented image D) Weighted and unweighted centroid location image.

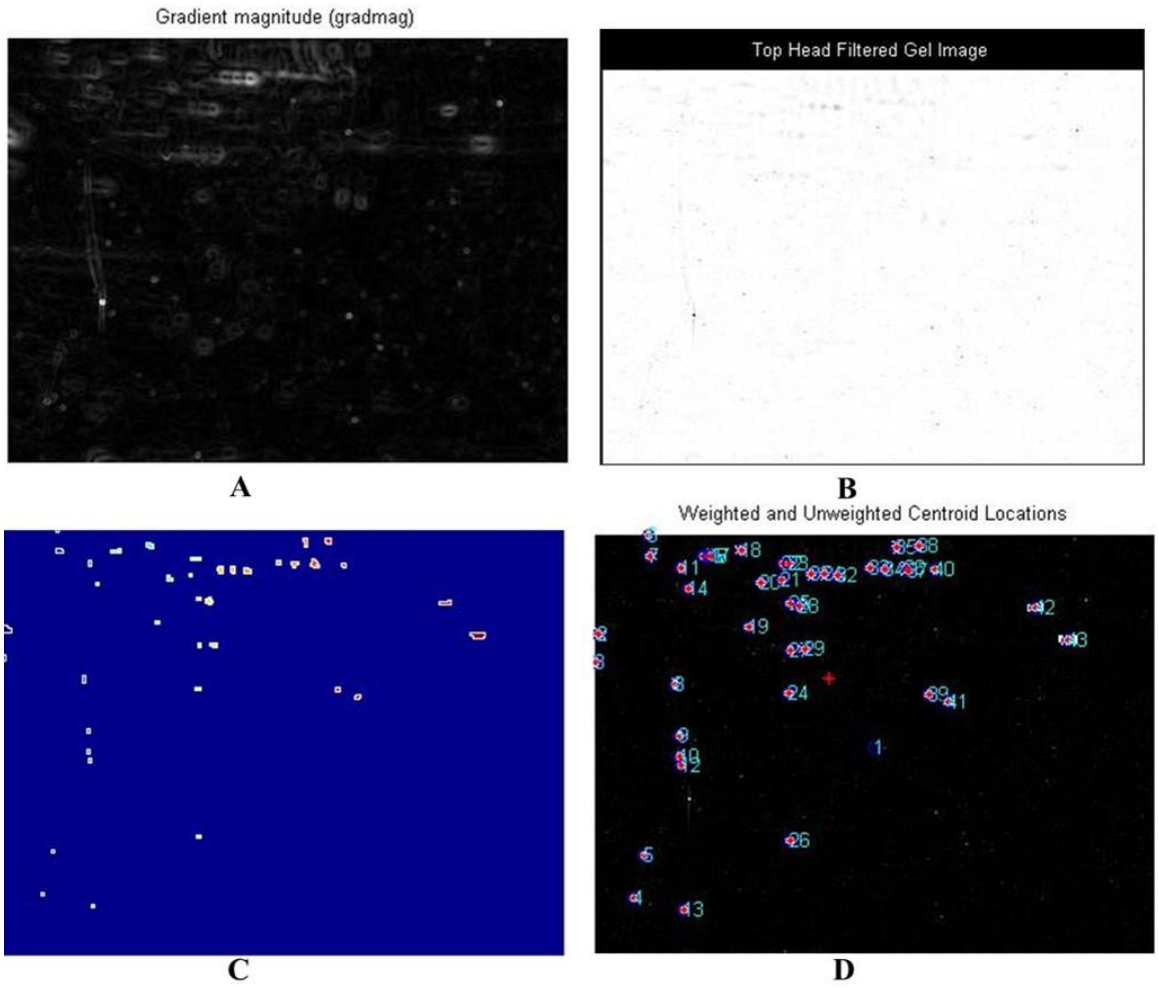

Figure 4.4 Overall presentation of segmentation.

A) Gradient magnitude image of Mp719. B) Top Hat filtered image C) Segmented image D) Weighted and unweighted centroid location image.

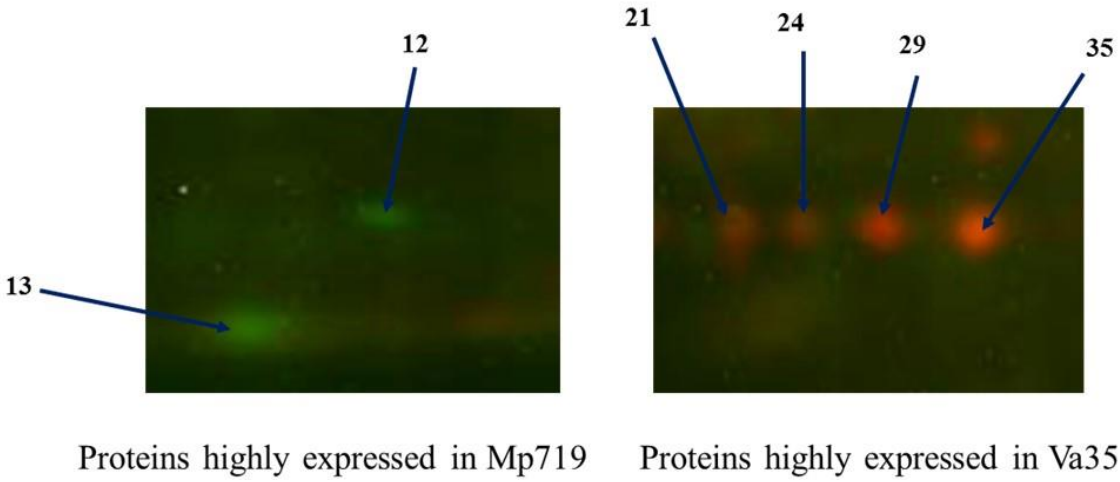

Figure 4.5 A presentation of differentially expressed proteins in Mp719 and V35. Left piece of the gel image shows proteins highly expressed in Mp719 and right piece of the gel image shows proteins highly expressed in Va35. A zoomed version of gel images are presented here.

Proteins highly expressed in Mp719 are located mostly in the upper right and upper left portion of the gel image and proteins highly expressed in Va35 are mainly located in the middle region of the gel image

Discussion

Maize (Zea mays L.) is an economically important crop which provides food and feed to millions of people and is greatly affected by aflatoxin contamination worldwide (Brown et al., 2013). Nowadays it becomes a great concern for farmers and scientists to develop methods which can prevent aflatoxin contamination in maize in pre and post-harvest conditions. Various pre and post-harvest strategies were applied by scientists to circumvent this problem. The development of resistant maize inbred lines is among one of the best strategies (Brown et al., 2004). Identification of biomarkers and differentially expressed proteins through Molecular Biology and Proteomic studies can be valuable

resources in developing new maize inbred lines with resistance to aflatoxin contamination.

A tremendous advancement has been observed in the area of proteomics in recent decades. From traditional 2DE to DIGE, proteomics has achieved several milestones. Being a very convenient, cheap and reproducible method, 2DE has faced several drawbacks such as gel to gel variation which was later circumvented by DIGE.

DIGE technique involves three basic steps for protein quantification. First one is the labeling of proteins with different fluorescent dyes (Cy2, Cy3 and Cy5) and multiplexing of samples in a single gel. Second step of DIGE is the loading of standard (pool of all the samples) in the gel and finally scanning of gels at different wavelengths (Timms and Cramer, 2008). These steps makes this technique highly advanced and more efficient for the identification of differentially expressed proteins in diseased/non-diseased plant samples in comparison to traditional 2DE. However, the use of Cydye2 in the labeling of standard samples are still controversial due to its weak signal production in comparison to Cy3 and Cy5 (Minden et al., 2009). In this study, only Cy3 and Cy5 dyes were used to label Va35 and Mp719 to identify the difference. DIGE overpowers traditional 2DE technique because of its sensitivity, highly dynamic range for protein detection and multiplexing of samples. These advancements allowed DIGE technique to analyze different samples in a single gel and also reduces statistical variation. Like other technique, DIGE has also faced some challenges such as Cydye minimal and saturation labelling is costly and it can only label lysine and cysteine amino acids present in the protein sample. These different steps makes this technique time consuming and labor intensive which requires highly skilled workers.

DIGE technique also required low fluorescent glass plates for gel casting which should be made of quartz or fused silica, because normal glass plates can interfere in imaging process. They can create high background fluorescence which can ultimately affect gel analysis.

The purchasing cost of these low fluorescent glass plates is also higher than normal glass plates used in 2DE. DIGE gel visualization is also a costly approach because it requires special equipment for the visualization purpose (Chandramouli and Qian, 2009). Typhoon trio variable mode imager which produces images of fluorescent, radioactive and chemiluminescent labeled proteins is very useful but the initial setup cost is high. Detection of fluorescently labeled gel images are also very challenging due to detection of signals from non-protein spots and overlapping signals from two different spots. These problems can be reduced by avoiding any dust or residue from glass plates and filtering out non-protein related signals. This scanner also contains a software for imaging and image analysis purpose. DIGE gel analysis generally requires several software packages such as DeCyder^{TM} (GE healthcare), Progenesis same spot, Melanie and Delta 2D. Different software package provides different results in data analysis. Therefore, selection of software package for DIGE gel analysis is also a crucial step (Granlund et al., 2009). The biggest problem persists with these software packages is matching of protein spots which sometimes need manual intervention of users. However, this problem can be solved by employing Matlab image processing toolbox which needs no user intervention in protein spot matching. Total 59 protein spots were identified using Matlab image processing toolbox in this workflow. The only requirement was to know the Matlab programing for employing Matlab image processing tool in image analysis.

Matlab program can convert colored images in to gray scale image for image processing. So, this is suitable for analysis of both 2DE and DIGE gel images. DIGE gel images were analyzed separately to extract the useful information's. For denoising of the gel images log and Tophat transform was applied. It is a crucial step which removes all the artifacts (dust particles, finger prints, distortion in the gel) present in the gel images and extract only true protein spots. For the segmentation of gel images, marker controlled watershed segmentation was performed to reduce the over segmentation problems. Differentially expressed protein spots were identified in this study.

Further analysis of protein spots are also a challenging step. It requires mostly MS based techniques for spot picking. Sometimes, minimal labeling can create problems because it only labels less than 3% of the lysine residue in a protein and MS based techniques cannot detect low resolution protein spots. MS based techniques also required large number of preparative gels for spot picking purpose and sometimes post staining of the gels as well.

Overall, DIGE has been a very advanced and highly sensitive technique in protein quantification, but the cost of initial setup of DIGE is very high. Other pitfalls with DIGE is that it cannot efficiently detect hydrophobic, highly acidic and basic proteins and proteins having molecular weights less than 10 kDa or higher than 150 kDa (Timms and Cramer, 2008). It is suggested that verification of DIGE results are highly required due to these limitations. In the present study, DIGE technique was not so helpful due to these limitations.

REFERENCES

Agrawal, P., Shriwastava, S., Limaye, S., 2010. MATLAB implementation of image segmentation algorithms, Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on, IEEE, pp. 427-431.

Ahad, M.A.R., 2011. Low-level Image Processing for Action Representations, Computer Vision and Action Recognition, Springer, pp. 9-37.

Almeida, J.S., Stanislaus, R., Krug, E., Arthur, J.M., 2005. Normalization and analysis of residual variation in two-dimensional gel electrophoresis for quantitative differential proteomics. Proteomics 5, 1242-1249.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of molecular biology 215, 403-410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25, 3389-3402.

Anderson, N.L., Matheson, A.D., Steiner, S., 2000. Proteomics: applications in basic and applied biology. Current Opinion in Biotechnology 11, 408-412.

Anfinsen, C.B., 1973. Principles that govern the folding of protein chains. Science 181, 223-230.

Appel, R.D., Vargas, J.R., Palagi, P.M., Walther, D., Hochstrasser, D.F., 1997. Melanie II–a third-generation software package for analysis of two-dimensional electrophoresis images: II. Algorithms. Electrophoresis 18, 2735-2748.

Baker, D., Sali, A., 2001. Protein structure prediction and structural genomics. Science 294, 93-96.

Bartlett, J.M., Stirling, D., 2003. A short history of the polymerase chain reaction, PCR protocols, Springer, pp. 3-6.

Bass, H.W., Krawetz, J.E., OBrian, G.R., Zinselmeier, C., Habben, J.E., Boston, R.S., 2004. Maize ribosome-inactivating proteins (RIPs) with distinct expression patterns have similar requirements for proenzyme activation. Journal of experimental botany 55, 2219-2233.

Beucher, S., 1983. Extrema of grey-tone functions and mathematical morphology. Rapport du CGMM, Ecole des Mines, Fontainebleau.

Beucher, S., Lantuejoul, C., 1979. Use of watersheds in contour detection.

Blackie, M.J., Mellor, J.W., Delgado, C.L., 1987. Accelerating food production in sub-Saharan Africa. Cambridge Univ Press.

Bleau, A., Leon, L.J., 2000. Watershed-based segmentation and region merging. Computer Vision and Image Understanding 77, 317-370.

Boykov, Y., Funka-Lea, G., 2006. Graph cuts and efficient ND image segmentation. International journal of computer vision 70, 109-131.

Brown, R.L., Bhatnagar, D., Cleveland, T.E., Chen, Z.-Y., Menkir, A., 2013. Development of maize host resistance to aflatoxigenic fungi.

Brown, R.L., Zhi-Yuan, C., Menkir, A., Cleveland, T.E., 2004. Using biotechnology to enhance host resistance to aflatoxin contamination of corn. African journal of biotechnology 2, 557-562.

Bujnicki, J.M., Elofsson, A., Fischer, D., Rychlewski, L., 2001. LiveBench-1: Continuous benchmarking of protein structure prediction servers. Protein Science 10, 352-361.

Cachier, P., Mangin, J.-F., Pennec, X., Rivière, D., Papadopoulos-Orfanos, D., Régis, J., Ayache, N., 2001. Multisubject non-rigid registration of brain MRI using intensity and geometric features, Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001, Springer, pp. 734-742.

Canny, J., 1986. A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 679-698.

Chandramouli, K., Qian, P.-Y., 2009. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. Human Genomics and Proteomics 1, 239204.

Cheng, Y.-H., Shen, T.-F., Fei Pang, V., Chen, B.-J., 2001. Effects of aflatoxin and carotenoids on growth performance and immune response in mule ducklings. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology 128, 19-26.

Chulze, S., 2010. Strategies to reduce mycotoxin levels in maize during storage: a review. Food Additives and Contaminants 27, 651-657.

Clark, B.N., Gutstein, H.B., 2008. The myth of automated, high-throughput two-dimensional gel analysis. Proteomics 8, 1197-1203.

Cleveland, T.E., Dowd, P.F., Desjardins, A.E., Bhatnagar, D., Cotty, P.J., 2003. United States Department of Agriculture—Agricultural Research Service research on pre-harvest prevention of mycotoxins and mycotoxigenic fungi in US crops. Pest management science 59, 629-642.

Collins, D.L., Le Goualher, G., Evans, A.C., 1998. Non-linear cerebral registration with sulcal constraints, Medical Image Computing and Computer-Assisted Interventation—MICCAI'98, Springer, pp. 974-984.

Conradsen, K., Pedersen, J., 1992. Analysis of two-dimensional electrophoretic gels. Biometrics, 1273-1287.

Cotty, P., 1989. Virulence and cultural characteristics of two Aspergillus flavus strains pathogenic on cotton. Phytopathology 79, 808-814.

Daszykowski, M., Færgestad, E.M., Grove, H., Martens, H., Walczak, B., 2009. Matching 2D gel electrophoresis images with Matlab 'image processing toolbox'. Chemometrics and Intelligent Laboratory Systems 96, 188-195.

Datta, S.K., Muthukrishnan, S., 1999. Pathogenesis-related proteins in plants. CRC press.

Decker, G., Wanner, G., Zenk, M.H., Lottspeich, F., 2000. Characterization of proteins in latex of the opium poppy (Papaver somniferum) using two-dimensional gel electrophoresis and microsequencing. Electrophoresis 21, 3500-3516.

Dehury, B., Patra, M.C., Maharana, J., Sahu, J., Sen, P., Modi, M.K., Choudhury, M.D., Barooah, M., 2014. Structure-Based Computational Study of Two Disease Resistance Gene Homologues (Hm1 and Hm2) in Maize (Zea mays L.) with Implications in Plant-Pathogen Interactions. PloS one 9, e97852.

Dehury, B., Sahu, M., Sarma, K., Sahu, J., Sen, P., Modi, M.K., Sharma, G.D., Choudhury, M.D., Barooah, M., 2013. Molecular Phylogeny, Homology Modeling, and Molecular Dynamics Simulation of Race-Specific Bacterial Blight Disease Resistance Protein (xa 5) of Rice: A Comparative Agriproteomics Approach. Omics: a journal of integrative biology 17, 423-438.

Dhanasekaran, D., Shanmugapriya, S., Thajuddin, N., Panneerselvam, A., 2011. Aflatoxins and Aflatoxicosis in Human and Animals. Aflatoxins-Biochemistry and Molecular Biology, 221-254.

Diamond, J., 2002. Evolution, consequences and future of plant and animal domestication. Nature 418, 700-707.

Donner, M., Atehnkeng, J., Sikora, R., Bandyopadhyay, R., Cotty, P., 2010. Molecular characterization of atoxigenic strains for biological control of aflatoxins in Nigeria. Food Additives and Contaminants 27, 576-590.

Donoho, D.L., 1995. De-noising by soft-thresholding. Information Theory, IEEE Transactions on 41, 613-627.

Donoho, D.L., Johnstone, I.M., 1995. Adapting to unknown smoothness via wavelet shrinkage. Journal of the american statistical association 90, 1200-1224.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., Picard, D., 1995. Wavelet shrinkage: asymptopia? Journal of the Royal Statistical Society. Series B (Methodological), 301-369.

Donoho, D.L., Johnstone, J.M., 1994. Ideal spatial adaptation by wavelet shrinkage. Biometrika 81, 425-455.

Dorner, J.W., 2004. Biological control of aflatoxin contamination of crops. Toxin Reviews 23, 425-450.

Dorner, J.W., Abbas, H., 2005. Biological control of aflatoxin crop contamination. Aflatoxin and food safety, 333-352.

Dowsey, A.W., English, J.A., Lisacek, F., Morris, J.S., Yang, G.Z., Dunn, M.J., 2010. Image analysis tools and emerging algorithms for expression proteomics. Proteomics 10, 4226-4257.

Durbin, S., Feher, G., 1996. Protein crystallization. Annual review of physical chemistry 47, 171-204.

Edmond, V., Moysan, E., Khochbin, S., Matthias, P., Brambilla, C., Brambilla, E., Gazzeri, S., Eymin, B., 2011. Acetylation and phosphorylation of SRSF2 control cell fate decision in response to cisplatin. The EMBO journal 30, 510-523.

Efrat, A., Hoffmann, F., Kriegel, K., Schultz, C., Wenk, C., 2002. Geometric algorithms for the analysis of 2D-electrophoresis gels. Journal of Computational Biology 9, 299-315.

Ehrlich, K.C., Montalbano, B.G., Cotty, P.J., 2007. Analysis of single nucleotide polymorphisms in three genes shows evidence for genetic isolation of certain Aspergillus flavus vegetative compatibility groups. FEMS microbiology letters 268, 231-236.

Eswar, N., John, B., Mirkovic, N., Fiser, A., Ilyin, V.A., Pieper, U., Stuart, A.C., Marti-Renom, M.A., Madhusudhan, M.S., Yerkovich, B., 2003. Tools for comparative protein structure modeling and analysis. Nucleic acids research 31, 3375-3380.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M., Eramian, D., Shen, M.y., Pieper, U., Sali, A., 2006. Comparative protein structure modeling using Modeller. Current protocols in bioinformatics, 5.6. 1-5.6. 30.

Eyrich, V.A., Martı-Renom, M.A., Przybylski, D., Madhusudhan, M.S., Fiser, A., Pazos, F., Valencia, A., Sali, A., Rost, B., 2001. EVA: continuous automatic evaluation of protein structure prediction servers. Bioinformatics 17, 1242-1243.

Faurobert, M., Pelpoir, E., Chaïb, J., 2007. Phenol extraction of proteins for proteomic studies of recalcitrant plant tissues, Plant Proteomics, Springer, pp. 9-14.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M., 1989. Electrospray ionization for mass spectrometry of large biomolecules. Science 246, 64-71.

Fiser, A., 2004. Protein structure modeling in the proteomics era. Expert Review of Proteomics 1, 97-110.

Floudas, C., 2007a. Computational methods in protein structure prediction. Biotechnology and bioengineering 97, 207-213.

Floudas, C.A., 2007b. Computational methods in protein structure prediction. Biotechnol. Bioeng. 97, 207-213.

Gauci, V.J., Wright, E.P., Coorssen, J.R., 2011. Quantitative proteomics: assessing the spectrum of in-gel protein detection methods. Journal of chemical biology 4, 3-29.

Giambrone, J., Diener, U., Davis, N., Panangala, V., Hoerr, F., 1985. Effects of aflatoxin on young turkeys and broiler chickens. Poultry Science 64, 1678-1684.

Gianazza, E., Astrua-Testori, S., Giacon, P., Righetti, P.G., 1985. An improved protocol for two-dimensional maps of serum proteins with immobilized pH gradients in the first dimension. Electrophoresis 6, 332-339.

Gonzalez, R.C., Woods, R.E., 2002. Digital image processing, Prentice hall Upper Saddle River, NJ.

Görg, A., Postel, W., Weser, J., Günther, S., Strahler, J.R., Hanash, S.M., Somerlot, L., 1987. Horizontal two-dimensional electrophoresis with immobilized pH gradients in the first dimension in the presence of nonionic detergent. Electrophoresis 8, 45-51.

Granlund, I., Hall, M., Schröder, W.P., 2009. Difference gel electrophoresis (DIGE). eLS.

Griffin, T.J., Goodlett, D.R., Aebersold, R., 2001. Advances in proteome analysis by mass spectrometry. Current Opinion in Biotechnology 12, 607-612.

Gustafsson, J.S., Ceasar, R., Glasbey, C.A., Blomberg, A., Rudemo, M., 2004. Statistical exploration of variation in quantitative two-dimensional gel electrophoresis data. Proteomics 4, 3791-3799.

Haas, J., Roth, S., Arnold, K., Kiefer, F., Schmidt, T., Bordoli, L., Schwede, T., 2013. The Protein Model Portal—a comprehensive resource for protein structure and model information. Database 2013, bat031.

Hames, B.D., 1998. Gel Electrophoresis of Proteins: A Practical Approach: A Practical Approach. Oxford University Press.

Henry, W.B., 2013. Maize Aflatoxin Accumulation Segregates with Early Maturing Selections from an S2 Breeding Cross Population. Toxins 5, 162-172.

Henzel, W.J., Billeci, T.M., Stults, J.T., Wong, S.C., Grimley, C., Watanabe, C., 1993. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proceedings of the National Academy of Sciences 90, 5011-5015.

Hess, B., Kutzner, C., Van Der Spoel, D., Lindahl, E., 2008. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of chemical theory and computation 4, 435-447.

Hippler, M., Klein, J., Fink, A., Allinger, T., Hoerth, P., 2001. Towards functional proteomics of membrane protein complexes: analysis of thylakoid membranes from Chlamydomonas reinhardtii. The Plant Journal 28, 595-606.

Hochstrasser, D., Augsburger, V., Funk, M., Appel, R., Pellegrini, C., Muller, A.F., 1986. Immobilized pH gradients in capillary tubes and two-dimensional gel electrophoresis. Electrophoresis 7, 505-511.

Hsieh, T.-s., Brutlag, D., 1980. ATP-dependent DNA topoisomerase from D. melanogaster reversibly catenates duplex DNA rings. Cell 21, 115-125.

Hu, Y., Wang, G., Chen, G.Y., Fu, X., Yao, S.Q., 2003. Proteome analysis of Saccharomyces cerevisiae under metal stress by two-dimensional differential gel electrophoresis. Electrophoresis 24, 1458-1470.

Huynh, Q.K., Hironaka, C.M., Levine, E.B., Smith, C., Borgmeyer, J., Shah, D., 1992. Antifungal proteins from plants. Purification, molecular cloning, and antifungal properties of chitinases from maize seed. Journal of Biological Chemistry 267, 6635-6640.

James, P., Quadroni, M., Carafoli, E., Gonnet, G., 1993. Protein identification by mass profile fingerprinting. Biochemical and biophysical research communications 195, 58-64.

Johnson, H.J., Christensen, G.E., 2002. Consistent landmark and intensity-based image registration. Medical Imaging, IEEE Transactions on 21, 450-461.

Jungblut, P., Thiede, B., 1997. Protein identification from 2-DE gels by MALDI mass spectrometry. Mass spectrometry reviews 16, 145-162.

Kaczmarek, K., Walczak, B., de Jong, S., Vandeginste, B.G., 2004. Preprocessing of two-dimensional gel electrophoresis images. Proteomics 4, 2377-2389.

Karas, M., Hillenkamp, F., 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Analytical chemistry 60, 2299-2301.

Karthikeyan, B., Vaithiyanathan, V., Venkatraman, B., Menaka, M., 2012. Analysis of Image Segmentation for Radiographic Images. Indian Journal of Science and Technology 5, 3660-3664.

Kelley, R.Y., Williams, W.P., Mylroie, J.E., Boykin, D.L., Harper, J.W., Windham, G.L., Ankala, A., Shan, X., 2012. Identification of maize genes associated with host plant resistance or susceptibility to Aspergillus flavus infection and aflatoxin accumulation. PloS one 7, e36892.

Kensler, T.W., Roebuck, B.D., Wogan, G.N., Groopman, J.D., 2011. Aflatoxin: a 50-year odyssey of mechanistic and translational toxicology. Toxicological Sciences 120, S28-S48.

Kim, D.E., Chivian, D., Baker, D., 2004. Protein structure prediction and analysis using the Robetta server. Nucleic acids research 32, W526-W531.

Kimori, Y., Baba, N., Morone, N., 2010. Extended morphological processing: a practical method for automatic spot detection of biological markers from microscopic images. BMC bioinformatics 11, 373.

Kitajima, S., Sato, F., 1999. Plant pathogenesis-related proteins: molecular mechanisms of gene expression and protein function. Journal of Biochemistry 125, 1-8.

Koyuncu, C.F., Arslan, S., Durmaz, I., Cetin-Atalay, R., Gunduz-Demir, C., 2012. Smart markers for watershed-based cell segmentation. PloS one 7, e48664.
Krieger, E., Nabuurs, S.B., Vriend, G., 2003. Homology modeling. Methods of biochemical analysis 44, 509-524.

Larsson, P., Wallner, B., Lindahl, E., Elofsson, A., 2008. Using multiple templates to improve quality of homology models in automated homology modeling. Protein Science 17, 990-1002.

Li, F., Seillier-Moiseiwitsch, F., 2010. Differential analysis of 2D gel images. Methods in enzymology 487, 595-609.

Li, F., Seillier-Moiseiwitsch, F., 2011. Analyzing 2D gel images using a two-component empirical bayes model. BMC bioinformatics 12, 433.

Li, F., Seillier-Moiseiwitsch, F., Korostyshevskiy, V.R., 2011. Region-based statistical analysis of 2D PAGE images. Computational statistics & data analysis 55, 3059-3072.

Lilley, K.S., Friedman, D.B., 2004. All about DIGE: quantification technology for differential-display 2D-gel proteomics.

Liu, T., W Tang, G., Capriotti, E., 2011. Comparative modeling: the state of the art and protein drug target structure prediction. Combinatorial chemistry & high throughput screening 14, 532-547.

Lyra, M., Georgantzoglou, A.P.A., 2011. MATLAB as a Tool in Nuclear Medicine Image Processing. Edited by Clara M. Ionescu, 477.

Madhusudhan, M., Marti-Renom, M.A., Eswar, N., John, B., Pieper, U., Karchin, R., Shen, M.-Y., Sali, A., 2005. Comparative protein structure modeling, The Proteomics Protocols Handbook, Springer, pp. 831-860.

Magdeldin, S., Zhang, Y., Xu, B., Yoshida, Y., Yamamoto, T., 2012. Two-dimensional polyacrylamide gel electrophoresis–a practical perspective. Gel Electrophoresis-Principles and Basics.

Mahon, P., Packman, L.C., Dupree, P., 2000. A proteomic analysis of organelles from Arabidopsis thaliana. Electrophoresis 21, 3488-3499.

Mann, M., Hendrickson, R.C., Pandey, A., 2001. Analysis of proteins and proteomes by mass spectrometry. Annual review of biochemistry 70, 437-473.

Mann, M., Højrup, P., Roepstorff, P., 1993. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biological mass spectrometry 22, 338-345.

Martens, H., 1989. Multivariate calibration. John Wiley & Sons.

Martens, H., Martens, M., 2001. Multivariate analysis of quality. An introduction. Measurement Science and Technology 12, 1746.

Martí-Renom, M.A., Stuart, A.C., Fiser, A., Sánchez, R., Melo, F., Šali, A., 2000. Comparative protein structure modeling of genes and genomes. Annual review of biophysics and biomolecular structure 29, 291-325.

McGuffin, L.J., Bryson, K., Jones, D.T., 2000. The PSIPRED protein structure prediction server. Bioinformatics 16, 404-405.

McLafferty, F.W., Senko, M.W., 1994. Mass spectrometry in the development of drugs from traditional medicines. Stem Cells 12, 68-73.

Mehl, H.L., Jaime, R., Callicott, K.A., Probst, C., Garber, N.P., Ortega-Beltran, A., Grubisha, L.C., Cotty, P.J., 2012. Aspergillus flavus diversity on crops and in the environment can be exploited to reduce aflatoxin exposure and improve health. Annals of the New York Academy of Sciences 1273, 7-17.

Metraux, J., Burkhart, W., Moyer, M., Dincher, S., Middlesteadt, W., Williams, S., Payne, G., Carnes, M., Ryals, J., 1989. Isolation of a complementary DNA encoding a chitinase with structural homology to a bifunctional lysozyme/chitinase. Proceedings of the National Academy of Sciences 86, 896-900.

Meunier, B., Bouley, J., Piec, I., Bernard, C., Picard, B., Hocquette, J.-F., 2005. Data analysis methods for detection of differential protein expression in two-dimensional gel electrophoresis. Analytical biochemistry 340, 226-230.

Meyer-Baese, A., 2004. Pattern recognition for medical imaging. Academic Press.

Midoro-Horiuti, T., Brooks, E.G., Goldblum, R.M., 2001. Pathogenesis-related proteins of plants as allergens. Annals of allergy, asthma & immunology 87, 261-271.

Miller, I., Crawford, J., Gianazza, E., 2006. Protein stains for proteomic applications: which, when, why? Proteomics 6, 5385-5408.

Millioni, R., Miuzzo, M., Sbrignadello, S., Murphy, E., Puricelli, L., Tura, A., Bertacco, E., Rattazzi, M., Iori, E., Tessari, P., 2010. Delta2D and Proteomweaver: Performance evaluation of two different approaches for 2-DE analysis. Electrophoresis 31, 1311-1317.

Minden, J.S., Dowd, S.R., Meyer, H.E., Stühler, K., 2009. Difference gel electrophoresis. Electrophoresis 30, S156-S161.

Mirza, S.P., Halligan, B.D., Greene, A.S., Olivier, M., 2007. Improved method for the analysis of membrane proteins by mass spectrometry. Physiological genomics 30, 89-94.

Morris, J.S., Clark, B.N., Wei, W., Gutstein, H.B., 2009. Evaluating the performance of new approaches to spot quantification and differential expression in 2-dimensional gel electrophoresis studies. Journal of proteome research 9, 595-604.

Nägele, E., Vollmer, M., Hörth, P., Vad, C., 2004. 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. Expert review of proteomics 1, 37-46.

Nhek, S., Mosleth, E., Høy, M., Griessl, M., Tessema, B., Indahl, U., Martens, H., 2012. Nonlinear visualisation and pixel-based alignment of 2D electrophoresis images. Chemometrics and Intelligent Laboratory Systems 118, 97-108.

Nielsen, P.A., Olsen, J.V., Podtelejnikov, A.V., Andersen, J.R., Mann, M., Wiśniewski, J.R., 2005. Proteomic mapping of brain plasma membrane proteins. Molecular & Cellular Proteomics 4, 402-408.

Pappin, D.J., Hojrup, P., Bleasby, A.J., 1993. Rapid identification of proteins by peptide-mass fingerprinting. Current biology 3, 327-332.

Parvati, K., Rao, P., Mariya Das, M., 2009. Image segmentation using gray-scale morphology and marker-controlled watershed transformation. Discrete Dynamics in Nature and Society 2008.

Phung, S.L., Bouzerdoum, A., Chai Sr, D., 2005. Skin segmentation using color pixel classification: analysis and comparison. Pattern Analysis and Machine Intelligence, IEEE Transactions on 27, 148-154.

Pieper, U., Webb, B.M., Barkan, D.T., Schneidman-Duhovny, D., Schlessinger, A., Braberg, H., Yang, Z., Meng, E.C., Pettersen, E.F., Huang, C.C., 2011. ModBase, a database of annotated comparative protein structure models, and associated resources. Nucleic acids research 39, D465-D474.

Porubleva, L., Velden, K.V., Kothari, S., Oliver, D.J., Chitnis, P.R., 2001. The proteome of maize leaves: use of gene sequences and expressed sequence tag data for identification of proteins with peptide mass fingerprints. Electrophoresis 22, 1724-1738.

Probst, C., Cotty, P.J., 2012. Relationships between in vivo and in vitro aflatoxin production: reliable prediction of fungal ability to contaminate maize with aflatoxins. Fungal biology 116, 503-510.

Rabilloud, T., 2009. Membrane proteins and proteomics: love is possible, but so difficult. Electrophoresis 30, S174-S180.

Rabilloud, T., Adessi, C., Giraudel, A., Lunardi, J., 1997. Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 18, 307-316.

Rabilloud, T., Chevallet, M., Luche, S., Lelong, C., 2010. Two-dimensional gel electrophoresis in proteomics: past, present and future. Journal of proteomics 73, 2064-2077.

Rabilloud, T., Vaezzadeh, A.R., Potier, N., Lelong, C., Leize-Wagner, E., Chevallet, M., 2009. Power and limitations of electrophoretic separations in proteomics strategies. Mass spectrometry reviews 28, 816-843.

Raman, B., Cheung, A., Marten, M.R., 2002. Quantitative comparison and evaluation of two commercially available, two-dimensional electrophoresis image analysis software packages, Z3 and Melanie. Electrophoresis 23, 2194-2202.

Rencher, A.C., Christensen, W.F., 2012. Methods of multivariate analysis. John Wiley & Sons.

Ricci, K.A., Girosi, F., Tarr, P.I., Lim, Y.-W., Mason, C., Miller, M., Hughes, J., Von Seidlein, L., Agosti, J.M., Guerrant, R.L., 2006. Reducing stunting among children: the potential contribution of diagnostics. Nature 444, 29-38.

Robbins, M.L., Roy, A., Wang, P.-H., Gaffoor, I., Sekhon, R.S., de O Buanafina, M.M., Rohila, J.S., Chopra, S., 2013. Comparative proteomics analysis by DIGE and iTRAQ provides insight into the regulation of phenylpropanoids in maize. Journal of proteomics 93, 254-275.

Robertson, D., Mitchell, G.P., Gilroy, J.S., Gerrish, C., Bolwell, G.P., Slabas, A.R., 1997. Differential extraction and protein sequencing reveals major differences in patterns of primary cell wall proteins from plants. Journal of Biological Chemistry 272, 15841-15848.

Rodriguez-del-Bosque, L., 1996. Impact of agronomic factors on aflatoxin contamination in preharvest field corn in northeastern Mexico. Plant disease 80, 988-993.

Rogers, M., Graham, J., 2004. 2 Dimensional Electrophoresis Gel Registration Using Point Matching and Local Image-Based Refinement.

Rohr, K., Cathier, P., Worz, S., 2004. Elastic registration of gel electrophoresis images based on landmarks and intensities, Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on, IEEE, pp. 1451-1454.

Rohr, K., Fornefett, M., Stiehl, H.S., 1999. Approximating thin-plate splines for elastic registration: Integration of landmark errors and orientation attributes, Information Processing in Medical Imaging, Springer, pp. 252-265.

Rosengren, A.T., Salmi, J.M., Aittokallio, T., Westerholm, J., Lahesmaa, R., Nyman, T.A., Nevalainen, O.S., 2003. Comparison of PDQuest and Progenesis software packages in the analysis of two-dimensional electrophoresis gels. Proteomics 3, 1936-1946.

Roy, A., Zhang, Y., 2012. Protein Structure Prediction. eLS.

Rye, M.B., Alsberg, B.K., 2008. A multivariate spot filtering model for two-dimensional gel electrophoresis. Electrophoresis 29, 1369-1381.

Rye, M.B., Færgestad, E.M., Alsberg, B.K., 2008. A new method for assigning common spot boundaries for multiple gels in two-dimensional gel electrophoresis. Electrophoresis 29, 1359-1368.

Sadana, J., Asrani, R., Pandita, A., 1992. Effect of dietary aflatoxin B1 on the growth response and haematologic changes of young Japanese quail. Mycopathologia 118, 133-137.

Sali, A., Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. Journal of molecular biology 234, 779-815.

Sanchez, R., Sali, A., 1997. Evaluation of comparative protein structure modeling by MODELLER-3. Proteins Structure Function and Genetics 29, 50-58.

Santoni, V., Kieffer, S., Desclaux, D., Masson, F., Rabilloud, T., 2000a. Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. Electrophoresis 21, 3329-3344.

Santoni, V., Molloy, M., Rabilloud, T., 2000b. Membrane proteins and proteomics: un amour impossible? Electrophoresis 21, 1054-1070.

Schulz, G.E., Schirmer, R.H., 1979. Principles of protein structure. Springer-Verlag KG.

Serna, A., Maitz, M., O'Connell, T., Santandrea, G., Thevissen, K., Tienens, K., Hueros, G., Faleri, C., Cai, G., Lottspeich, F., 2001. Maize endosperm secretes a novel antifungal protein into adjacent maternal tissue. The Plant Journal 25, 687-698.

Sidman, C., 1981. Two-dimensional gel electrophoresis. Immunological methods 2, 57-74.

Skolnick, M.M., 1986. Application of morphological transformations to the analysis of two-dimensional electrophoretic gels of biological materials. Computer vision, graphics, and image processing 35, 306-332.

Skylas, D., Van Dyk, D., Wrigley, C., 2005. Proteomics of wheat grain. Journal of cereal science 41, 165-179.

Smilansky, Z., 2001. Automatic registration for images of two-dimensional protein gels. Electrophoresis 22, 1616-1626.

Snelders, E., Karawajczyk, A., Schaftenaar, G., Verweij, P.E., Melchers, W.J., 2010. Azole resistance profile of amino acid changes in Aspergillus fumigatus CYP51A based on protein homology modeling. Antimicrobial agents and chemotherapy 54, 2425-2430.

Sonka, M., Hlavac, V., Boyle, R., 2008. Image processing, analysis, and machine vision. Thomson Toronto.

Strasters, K.C., Gerbrands, J.J., 1991. Three-dimensional image segmentation using a split, merge and group approach. Pattern Recognition Letters 12, 307-325.

Tajkarimi, M., Shojaee, M.H., Yazdanpanah, H., Ibrahim, S.A., 2011. Aflatoxin in Agricultural Commodities and Herbal Medicine. Aflatoxins–Biochemistry and Molecular Biology, 367-396.

Tenaillon, M.I., Charcosset, A., 2011. A European perspective on maize history. Comptes rendus biologies 334, 221-228.

Timms, J.F., Cramer, R., 2008. Difference gel electrophoresis. Proteomics 8, 4886-4897.

Tompa, P., Fersht, A., 2010. Structure and function of intrinsically disordered proteins. CRC Press.

Ünlü, M., Morgan, M.E., Minden, J.S., 1997. Difference gel electrophoresis. A single gel method for detecting changes in protein extracts. Electrophoresis 18, 2071-2077.

Van Breemen, R.B., Snow, M., Cotter, R.J., 1983. Time-resolved laser desorption mass spectrometry. I. Desorption of preformed ions. International Journal of Mass Spectrometry and Ion Physics 49, 35-50.

Van Loon, L., Van Strien, E., 1999. The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiological and molecular plant pathology 55, 85-97.

Varjo, J., 2014. Implementing Texture Analysis Software Frame for Magnetic Resonance Image Data in MATLAB.

Vincent, L., 1993. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. Image Processing, IEEE Transactions on 2, 176-201.

Vincent, L., Soille, P., 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE transactions on pattern analysis and machine intelligence 13, 583-598.

Virgilio, M.d., Lombardi, A., Caliandro, R., Fabbrini, M.S., 2010. Ribosome-inactivating proteins: from plant defense to tumor attack. Toxins 2, 2699-2737.

Wagacha, J., Muthomi, J., 2008. Mycotoxin problem in Africa: current status, implications to food safety and health and possible management strategies. International Journal of Food Microbiology 124, 1-12.

Wang, W., Ackermann, D., Mehlich, A.-M., König, S., 2011. Impact of quenching failure of Cy dyes in differential gel electrophoresis. PloS one 6, e18098.

Wang, W., Vignani, R., Scali, M., Cresti, M., 2006. A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. Electrophoresis 27, 2782-2786.

Wang, Y., Staib, L.H., 1998. Elastic model based non-rigid registration incorporating statistical shape information, Medical Image Computing and Computer-Assisted Interventation—MICCAI'98, Springer, pp. 1162-1173.

Wheelock, Å.M., Buckpitt, A.R., 2005. Software-induced variance in two-dimensional gel electrophoresis image analysis. Electrophoresis 26, 4508-4520.

Wicklow, D., Donahue, J.E., 1984. Sporogenic germination of sclerotia inAspergillus flavus and A. parasiticus. Transactions of the British Mycological Society 82, 621-624.

Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., 1996. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and arnino acid analysis. Nature Biotechnology 14, 61-65.

Williams, W., 2006. Breeding for resistance to aflatoxin accumulation in maize. Mycotoxin research 22, 27-32.

Williams, W.P., Windham, G.L., 2001. Registration of Maize Germplasm Line Mp715 Registration by CSSA. Crop Sci. 41, 1374-a-1375.

Williams, W.P., Windham, G.L., 2012. Registration of Mp718 and Mp719 germplasm lines of maize. Journal of Plant Registrations 6, 200-202.

Windham, G., Williams, W., 2002. Evaluation of corn inbreds and advanced breeding lines for resistance to aflatoxin contamination in the field. Plant disease 86, 232-234.

Wold, S., Martens, H., Wold, H., 1983. The multivariate calibration problem in chemistry solved by the PLS method, Matrix pencils, Springer, pp. 286-293.

Woodward, A.M., Rowland, J.J., Kell, D.B., 2004. Fast automatic registration of images using the phase of a complex wavelet transform: application to proteome gels. Analyst 129, 542-552.

Wu, Y., Zhang, L., 2011. Comparison of two academic software packages for analyzing two-dimensional gel images. Journal of bioinformatics and computational biology 9, 775-794.

Xiong, B., Gui, C.-S., Xu, X.-Y., Luo, C., Chen, J., Luo, H.-B., Chen, L.-L., Li, G.-W., Sun, T., Yu, C.-Y., 2003. A 3D model of SARS_CoV 3CL proteinase and its inhibitors design by virtual screening. Acta pharmacologica Sinica 24, 497-504.

Yamaguchi, K., Subramanian, A.R., 2000. The Plastid Ribosomal Proteins IDENTIFICATION OF ALL THE PROTEINS IN THE 50 S SUBUNIT OF AN ORGANELLE RIBOSOME (CHLOROPLAST). Journal of Biological Chemistry 275, 28466-28482.

Yang, G.-Z., 2001. Multiresolution image registration for two-dimensional gel electrophoresis. Proteomics 1, 856-870.

Yin, Y.-n., Yan, L.-y., Jiang, J.-h., Ma, Z.-h., 2008. Biological control of aflatoxin contamination of crops. Journal of Zhejiang University Science B 9, 787-792.

Zhang, Y.J., 1996. A survey on evaluation methods for image segmentation. Pattern recognition 29, 1335-1346.

APPENDIX A

PROTEIN GEL IMAGES

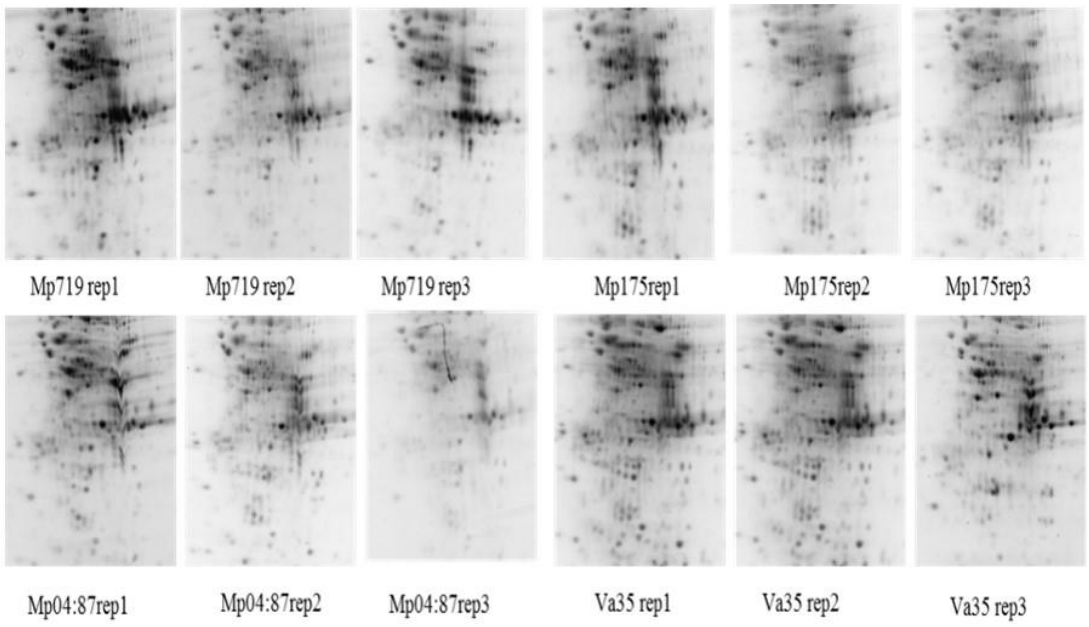

Figure A.1 Aligned and cropped images of two resistant (Mp719 and Mp715) and two susceptible (Mp04:87 and Va35) maize inbred lines.

All the gel images were inverted by using Adobe Photoshop software for illustration purpose.

The original gel images of the inverted ones were used in the Chapter III for the gel image analysis using Matlab image processing software. The original gel images has black background and white protein spots.

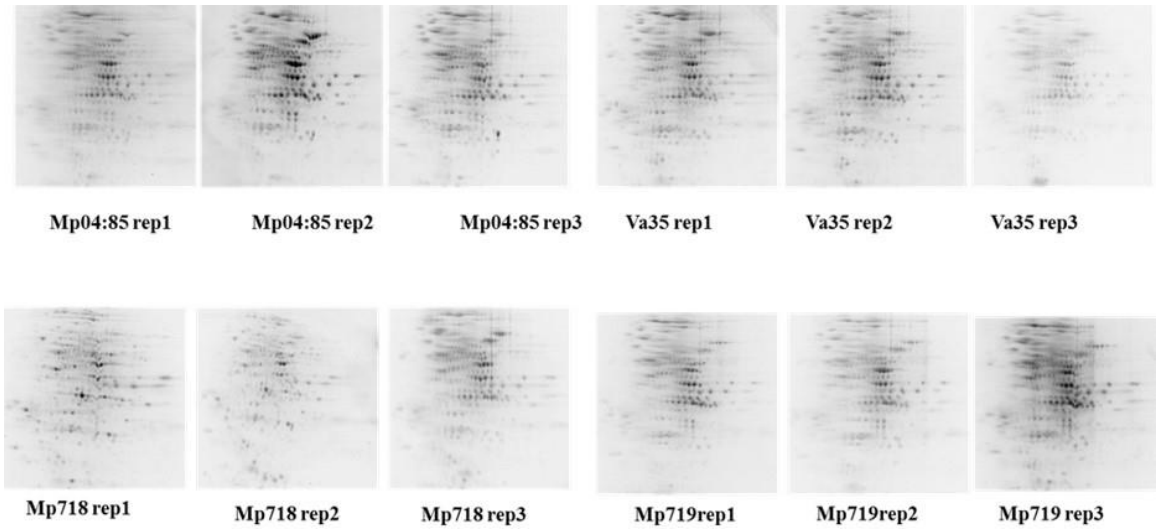

Figure A.2 Twelve aligned and cropped protein gel images of two resistant (Mp718 and Mp719) and two susceptible (Mp04:85 and Va35) maize inbred lines.

All the gel images were inverted by Gimp 2.8.14 software for the illustration purpose.