Proteomic Comparison of *Arabidopsis thaliana* Under High and Low Nitrogen Fertilization

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of
Master of Science at McGill University

by

KEITH E. M. DIAS

Department of Bioresource Engineering
Faculty of Agricultural & Environmental Sciences

McGill University
St. Anne de Bellevue, Québec,
Canada

April 15th, 2012

Acknowledgement

My sincere appreciation and gratitude to Prof. Mark G. Lefsrud, for his support towards my research project. I wish to thank Drs Valérie Orsat and Viacheslav Adamchuk, at McGill University, for guidance and assisting us in this project and all others that have contributed in some form or the other to this project.

Table of Contents

List of Tables

Page

List of Figures

Page

Abstract

Nitrogen (N) and the levels of N in plants play a vital role in the physiology, regulating their development and metabolism. We grew *Arabidopsis thaliana* under agronomic conditions at low (6 mg N/L) and high (106 mg N/L) N fertilizer regimes, maintaining a constant NO_3-N to NH_4-N ratio (3:1). Using a shotgun mass spectrometry proteomics approach, multi-dimensional protein identification technology (MudPIT), we characterized a total of 2134 reproducibly identified proteins shared between the two N treatments. By statistical analysis in both treatments we found 37 differentially expressed proteins that satisfied both the AC test and the FDR *q*-value specified cutoffs, where 18 proteins were down regulated and 19 proteins were up regulated under low and high N treatments. We also found 35 differentially expressed proteins that are statistically important but did not satisfy the *q*- test. These differentially expressed proteins appear to have roles in glycolysis, metabolic, developmental, and signaling processes, or protein binding, transport and nucleic acid binding. The proteins associated with glycolysis indicate glutamine metabolism is of major importance in the plant N economy since it provides N to young developing tissues. Our study indicates that under varying N level treatments, proteins responsible for glutamate synthase (GOGAT), glutamine synthase (GS), and dehydrogenase activity (DH) that serve as enzymes to catalyze a link between carbohydrate and amino acid metabolism are up regulated. Thus, this study has enabled us to apply comparative shotgun proteomics to characterize *A. thaliana* at the proteomic level and will provide the tools necessary to provide an improved understanding of how and what up-regulates and down regulates different proteins under varying environmental conditions.

Résumé

La fertilisation en azote (N) et la teneur en N des plantes ont un rôle clef dans leur physiologie, régulant leur développement et métabolisme. Tout en gardant un rapport de NO_3-N à NH_4-N de 3:1, des plants d'*Arabidopsis thaliana* (L.) Heynh furent cultivés sous deux régimes de fertilisation: bas (6 mg N/L) et élevé (106 mg N/L). Utilisant une technique protéomique en vrac par spectrométrie de masse et une technologie d'identification multidimensionnelle des protéines (MudPIT), nous avons pu caractériser un total de 2134 protéines identifiées de façon récurrente comme apparaissant dans les deux traitements de fertilisation azotée. Une analyse statistique des deux traitements a indiqué la présence de 37 protéines différentiellement exprimées, satisfaisant à la fois le test Audic-Claverie (AC) et le seuil de valeur *q* dans l'estimation du taux d'erreur (FDR). De celles-ci, 18 protéines furent régulées à la baisse lors des traitements à haut ou bas niveau de N, et 19 furent régulées à la hausse dans les mêmes circonstances. En plus, 35 protéines différentiellement exprimées du point de vue statistique ne passèrent tout de même pas le test de la valeur *q*. Les protéines différentiellement exprimées semblent avoir des rôles dans les processus de glycolyse, de métabolisme, de développement, de signalisation, et de transport, ainsi que dans la liaison des protéines et des acides nucléiques. Une analyse des protéines associées à la glycolyse indique que le métabolisme de la glutamine est d'une importance majeur dans l'économie en N de la plante, puisqu'il fournit l'azote aux jeunes tissus en voie de développement. Notre étude indique que, sous différents niveaux de fertilisation en N, les protéines responsables pour l'activité glutamate synthétase (GOGAT), glutamine synthétase (GS), et déshydrogénase (DH), servant comme enzymes dans la catalyse du lien entre les voies de métabolisme des glucides et celui des acides aminés, sont régulés à la hausse. Ainsi, cette étude nous permettra d'utiliser une technique protéomique en vrac comparative afin de

caractériser *A. thaliana* au niveau protéomique, et nous fournira les outils nécessaires à mieux comprendre quelles protéines sont régulées à la hausse ou à la baisse sous différentes conditions environnementales et comment cette régulation est mise en œuvre.

Chapter 1

Literature Review

Deoxyribonucleic acid (DNA) carries the genetic information of a cell and consists of thousands of genes, with each gene serving as a recipe on how to build a protein molecule. Genomics is the branch of science that studies the genome (genes) of individual organisms, populations, and species, contained in the DNA, to better understand the workings of the organism, and what happens when certain genes interact with each other and the environment (Boutros and Perrimon, 2000; Tyers and Mann, 2003).

Proteins (products of genes) perform important tasks for the cell functions or serve as building blocks (Finnie, 2006; Tyers and Mann, 2003). The flow of information from the genes determines the protein composition and thereby the functions of the cell. The DNA is situated in the nucleus organized into chromosomes, therefore every cell must contain genetic information and the DNA must be duplicated before the cell divides (replication) (Boutros and Perrimon, 2000; Cho et al., 2008; Tyers and Mann, 2003; Zhang and Riechers, 2008). When proteins are required by the cell, the corresponding genes are transcribed into RNA (transcription) (Cho et al., 2008; Kaul et al., 2000; Zhang and Riechers, 2008). The process of transcription has given rise to a field of science called transcriptomics (Handrick et al., 2010; MacKay et al., 2004). Transcriptomics has provided tools to researchers that are able to use the RNA to identify the expression of the genes during growth and development (Agrawal and Rakwal, 2008; Cho et al., 2008; Handrick et al., 2010; Park, 2004; Tyers and Mann, 2003).

The term proteome was initially proposed in the year 1994 as the “PROTEin complement expressed by a genOME”(Park, 2004; Patterson and Aebersold, 2003). A proteome study is

representative of a comprehensive description of all proteins expressed in a given cell, tissue or organism at any given time in given condition (Agrawal and Rakwal, 2008; Barbier-Brygoo and Joyard, 2004; Finnie, 2006; Zhang and Riechers, 2008). The study of the proteome is very dynamic responding to environmental and cellular challenges and complex because of the number of proteins that can be produced. In contrast, the genome (DNA) is very stable and does not change to environmental or cellular changes. Proteome studies provide information on the amino acid sequence, the properties of proteins, their relative abundance, specific activity, the state of modification and association with other proteins or molecules of different types, subcellular localization, and three dimensional structures represent crucial information for the description of biological systems (Agrawal and Rakwal, 2008; Finnie, 2006; Rose, 2004; Tyers and Mann, 2003). Therefore the term proteomics is not only restricted to construction of proteins but also stands for studies of protein properties such as post translational modification (PTM), expressional levels, and also provides an integrated view of cellular processes and networks at the protein level (Colas et al., 2010).

Proteomics provides us with the capability of studying PTMs, to analyze the number of biochemical and physical changes to proteins caused by biotic or abiotic stresses (Barbier-Brygoo and Joyard, 2004), and aids in better understanding of signaling pathways in plants(Agrawal and Rakwal, 2008; Finnie, 2006). Quantitative (profiling) analysis of global proteins levels, termed as “quantitative proteomics” is required for the system-based understanding of the molecular function of each protein component and is expected to provide insights into molecular mechanisms of various biological processes and systems. (Barbier-Brygoo and Joyard, 2004; Finnie, 2006; Tyers and Mann, 2003; von Mering et al., 2002).

Plant genomes express many proteins since they contain tens of thousands of genes (AGI, 2000; Haynes and Roberts, 2007). Plants are not only incredibly complex living systems consisting of interdependent organs but also consists of thousands of tissues within each of those organs, and specialized organelles and compartments within each individual cells. An ideal proteomic approach would include a highly sensitive, high through-put, analytically robust technique, with the ability to differentiate between differentially expressed proteins, and analyze thousands of proteins in a given sample.

1.1 Methodologies for analysis:

Technologies and methods are routinely employed for differential proteomics studies such as validation of regulated proteins, biomarkers and targets. Two-dimensional gel electrophoresis (2-DE) with immobilized pH gradients (IPGs) combined with protein identification by mass spectrometry (MS) is currently the workhorse for the majority of ongoing proteome projects. Although, alternative and complementary technologies such as: Multi-dimensional Protein Identification Techniques (MudPit) have emerged. Below is the description of two most widely used methods.

1.1.2 2-D (SDS-PAGE) (Sodium dodecyl sulfate two dimensional polyacrylamide gel electrophoresis):

This method was developed in 1970 and is still very much used to separate proteins in complex protein mixtures.(Bjellqvist et al., 1982; Görg et al., 2000; Haynes and Roberts, 2007; Kenrick and Margolis, 1970; Schulze and Usadel, 2010). Two-dimensional electrophoresis (2-DE) couples isoelectric focusing (IEF) in the first dimension and sodium dodecyl sulfate

polyacrylamide gel electrophoresis (SDS-PAGE) in the second dimension to separate proteins according to two independent parameters, i.e., isoelectric point (p *I*) in the first dimension and molecular mass in the second.

This method permits separation of thousands of proteins under optimal conditions. Gel spots also called protein spots are excised and digested for MS or MS/MS analysis. The spectra are then searched against different protein sequence databases to identify the proteins.

An advantage of SDS-PAGE is, when analyzing different samples from different time points a differential comparison of protein comparison can be visualized. This method provides information about proteins p*I* values (Görg et al., 2000), protein modifications and isoforms (Haynes and Roberts, 2007). However there are at least two major limitations of SDS-PAGE. First, there is a restriction of detection of 30% of all cellular proteins, which is especially deleterious in plant cells that are heavily populated by membranous structures due to a lack of representation of basic, hydrophobic and membrane spanning proteins. Second, the lack of automation in running 2-D gels has made it less favorable in comparison to MudPIT.

1.1.3 MudPIT (Multi-dimensional Protein Identification Techniques):

In this method all proteins are digested into peptides before separation. The peptides are separated (by two orthogonal properties, charge and hydrophobicity) via two dimensional chromatography, two columns are packed with reversed phase (RP) resins and cation exchange (SCX) resins. The peptide samples are directly loaded onto bi-phasic back column. The back column consisted of strong cation exchange chromatographic resin with a C_{18} reverse - phase packing material loaded sequentially. The back column is then connected between the liquid

chromatography (LC) pump and a 15 cm C_{18} analytical column sprayed into the mass spectrometer with an electric current applied to provide ionization to act like an ion source and transport ions into the mass spectrometer. Since, the proteins are digested into peptides, the original protein must be reconstructed computationally from the measured peptides into theoretical proteins that could be undistinguishable from the actual proteins due to shared peptide sequences (Nesvizhskii and Aebersold, 2005).

The resulting tandem mass spectrometry data are searched using the SEQUEST algorithm, which interprets the tandem mass spectra (MS/MS) generated and identifies the peptide sequence from which it was generated, resulting in the determination of the protein content of the original sample. This software provides functionality by incorporating a probability model to ascertain that the protein assembly is correct (Peng et al., 2003; Washburn, 2004; Wolters et al., 2001).

The advantages of MudPIT are: efficient detection of low abundance and hydrophobic proteins, the resolution of peptides, and generation of tandem mass spectra that can be achieved simultaneously using the same sample (Aebersold and Mann, 2003; Glinski and Weckwerth, 2006; Michael P. Washburn, 2001; Schulze and Usadel, 2010; Weckwerth et al., 2004; Wienkoop et al., 2004). Effectively, there are two ways to carry out a MudPIT or MudPIT-like experiments. Basically, SCX can either be run offline or online, with MudPIT being an online approach. In offline MudPIT, larger amounts of complex peptide mixture generated from a biological sample can be loaded onto the SCX back column, before attaching the back column to the front column, this offline method was developed to allow multiple flushing of the back column to allow an increase in the number of proteins identified from the sample (Froehlich et al., 2003). Online MudPIT offers minimal sample handling with the entire digested sample

directly injected into the front column for separation (no back column is required); this method works very well with limited sample handling, volume, or run time (Haynes and Roberts, 2007).

Disadvantages of MudPIT are detergents, clogging, ion suppression and large amounts of data production. Directly loading a sample can result in the loading of detergents from the sample preparation stage. Detergents are used to isolate hydrophobic proteins, and cannot be introduced into the MS, because when they are ionized and can cause interferences in the spectrum (Drexler et al., 2006). Direct loading of biological samples can lead to column clogging or slow deterioration in column performance, caused from containments and undigested protein in the sample. MS suffers from ion suppression effects which hinder the detection of low abundance ions co-eluting with ions of much higher abundance. The data comprises of thousands to millions of mass spectrums which needs expensive high powered computing and vast data storage space, and software such as SEQUEST, MASCOT, or de novo sequencing programs to deduce the amino acid sequence from spectra (Schulze and Usadel, 2010).

1.2 Quantitative methods of analysis using MS:

The methods described in the previous section continue to be an important tool in the field of proteomics, they suffer from some drawbacks. For example, MS relies on ionization of the peptides for detection, because ionization efficiency is affected by a number of factors, peak intensities of the same peptide from separate LC-MS/MS experiments are difficult to compare, making it exceedingly unlikely to detect PTMs or truncated forms of proteins (Agrawal and Rakwal, 2008; Finnie, 2006; Salvato and de Carvalho, 2010). One solution to the problem of properly identifying and quantifying proteins is using of quantitative methods in tandem with

mass spectrometry such as: isotope coded affinity tags (ICAT), immobilized metal affinity chromatography (IMAC), isobaric tags for relative and absolute quantitation (iTRAQ), difference gel electrophoresis (DIGE), label-free quantitation; example spectral counts (Peck, 2005; Rampitsch and Srinivasan, 2006).

1.2.1 Isotope Coded Affinity Tags (ICAT):

Isotope Coded Affinity Tags is a method that uses a reagent with specificity towards sulfhydryl groups, an eightfold deuterated linker, and a biotin affinity tag. The ICAT method is able to measure proteins representing two different cell states, one tagged with light and heavy isotopically tagged ICAT reagents, respectively. The samples are combined and enzymatically cleaved to generate peptide fragments, some of which are tagged. The tagged peptides are isolated by avidin affinity chromatography and the isolated peptides are analyzed by micro- LC-MS/MS. In the last step both the quantity and sequence identity of the proteins from which the tagged peptides originated are determined by MS/MS. The ratios of the original amounts of proteins from the two cell states are strictly maintained in the peptide fragments. The relative quantification is determined by the ratio of the peptide pairs. Every second scan is devoted to fragmenting and then recording sequence information about the eluting peptide (tandem mass spectrum). The protein is identified by computer matching the MS/MS data against a protein databases (Gygi et al., 1999; Peck, 2005; Schulze and Usadel, 2010; Thelen and Peck, 2007b)

1.2.2 Isobaris Tag for Relative and Absolute Quantitation (iTRAQ):

This method is similar to ICAT (Isotope Coded Affinity Tags) but involves chemical derivatization of the primary amines (peptides at the N terminus and the lysine side chains) of the

proteolytic peptide mixtures using a multiplexed set of amine- reactive reagents with distinct isotopic mass designs (Wiese et al., 2007). iTRAQ labeling, peptides are linked to isobaric tags consisting of a mass balance group and a reporter group. Differentially labeled peptides therefore appear as single peaks in MS scans. Relative quantitative information on iTRAQ- tagged peptides is obtained in MS/MS scans liberating the reporter group as distinct isotope- encoded fragments. Due to the specific mass design of the label and isotopic labeling at the peptide level, the classical iTRAQ represents a shot- gun approach that provides qualitative and quantitative information on the proteins simultaneously (Gygi et al., 1999; Wu et al., 2006).

1.2.3 Immobilized metal affinity chromatography (IMAC):

Immobilized metal affinity chromatography is a method that is also called metal chelate chromatography, which has been widely used for purification of proteins since its introduction by Porath et al. (2005). Phosphopeptides are usually masked due to ionization effects and non-phosphorylated proteins, making them invisible in complex sample mixtures. IMAC is a method that removes this problem is using an immobilized metal to enrich phosphopeptides prior to LC/MS analysis. Transition metal ions such as TiO_2^+ , Fe^{3+}, Ga^{3+} and ZrO_2 are often used in combination with cation exchange (SCX) chromatography as the first dimension of separation. The metal ions bind with the negatively charged phosphate group, or specific amino acid side chains (particularly those of histidine, cystine and tryptophan) and become weakly bound to the chelating groups of a chromatographic resin resulting in retention of proteins on the chromatographic column. IMAC has been used to examine the relationship between amino acid side chain surface topography of proteins and specific binding selectivity. It is a very useful

technique to purify recombitant proteins containing histidine tags (Oeljeklaus et al., 2009; Peck, 2005; Porath et al., 1975; Powell and Flurkey, 2006).

1.2.4 Difference gel electrophoresis (DIGE):

The difference gel electrophoresis method involves pre-incubating protein samples with activated fluorescent dyes (cyanine [Cy] dyes) to label the lysine (Lys) or cysteine (Cys) residues with a sensitive tag that can be used to quantify the abundance of that protein in solution (Thelen and Peck, 2007b). These Cy dye fluors (spectrally distinct fluorescent tags) covalently modify the eta- amino group of lysine in the proteins via amide linkages. Consequently, the same protein labeled with any of the fluors will migrate nearly to the same isoelectric point on a 2D gel as the unlabeled protein and produce similar 2-DE reference maps as traditional staining methods. In a typical protocol the controlled and the treated samples are separately labeled with two different charge-matched Lys-reactive dyes Cy3 and Cy5, having charges of +3 and +5 respectively, while a mixture containing equal amounts of controlled and treated samples is labeled with Cy2 (charge +2, Lys-reactive dye). The labeled samples are combined and run in a single 2D gel to allow better spot matching and minimize gel to gel variations. This method allows analysis of up to three protein samples on the same 2D gel (Wu et al., 2006).

1.2.5 Label-free quantification:

The signal intensity of peptide ions within an MS scan can be compared en mass from multiple liquid chromatography–mass spectrometry (LC-MS) analyses(Thelen and Peck, 2007a). This peak integration method is referred to as label-free quantification because no isotopic label is introduced into the proteins or peptides. Though this method is still in infancy, the reproducibility of online chromatographic separation of peptides combined with the high mass

accuracy of the latest generation of mass spectrometers machines offers renewed promise for this method. An example of this method is spectral counts, is discussed below.

1.2.6 Spectral Counts:

An alternative form of label-free quantification is spectral counting. Unlike peak integration, which calculates peak ion intensity from mass spectrometry (MS) scans, spectral counting tabulates the number of tandem (MS/MS) mass spectrometry (multiple steps of mass spectrometry selection, with some form of fragmentation occurring in between the stages) scans that are attributed to the same precursor ion (i.e., peptide in this case) (Salvato and de Carvalho, 2010; Schulze and Usadel, 2010; Thelen and Peck, 2007a). The frequency of these MS/MS scans (in theory) reflects the abundance of this peptide in the sample. Spectral counting is an approach that appeals to another developing characteristic of contemporary mass spectrometers: speed of data acquisition. For example, if 10 scans can be acquired per second on a mass spectrometer, a 2-h analytical gradient would yield >50,000 MS/MS scans, assuming two of the 10 scans are MS acquisitions (Oeljeklaus et al., 2009; Salvato and de Carvalho, 2010; Schulze and Usadel, 2010; Thelen and Peck, 2007a). This information from a simple LC-MS/MS run represents an un-mined reservoir of expression data comparable in number to EST DNA sequence reads from a cDNA library screen. However, at this point, it is unclear whether dynamic exclusion rules frequently applied during mass spectral acquisitions invalidate the spectral counting approach. Dynamic exclusion is used to maximize the number of peptides sequenced during tandem MS acquisitions. Individual peptides elute from a reversed-phase analytical column in the time scale of minutes, while a mass spectrometer collects data on the second or millisecond scale (Oeljeklaus et al., 2009; Salvato and de Carvalho, 2010; Schulze and Usadel, 2010; Thelen and Peck, 2007b). Therefore, rather than constantly re-sequencing an abundant peptide, dynamic

exclusion can be applied to ignore ions for which MS/MS spectra have already been acquired(Oeljeklaus et al., 2009; Salvato and de Carvalho, 2010; Schulze and Usadel, 2010; Thelen and Peck, 2007a).

Table L1. Quantitative Comparison: Advantages and disadvantages of the different protein quantitative methods (Peck, 2005; Thelen and Peck, 2007b)

METHOD	DESCRIPTION	ADVANTAGES	DISADVANTGES
DIGE* (Difference gel electrophoresis)	2-D gel analysis fluorescent based	Accurate pairwise comparison	Poor hydrophobic, basic, and large molecular mass proteins
ICAT* (Isotope coded affinity tags)	Isotopic Cys tagging residues based on LC-MS/MS	Few proteins captured and analyzed hence good for highly complex samples	One is seven proteins do not contain Cys residues
IMAC* (Immobilized metal affinity chromatography)	Enriches phosphopeptides in complex peptide mixtures	Eliminates/ decreases suppression effects on phosphorylated peptides in LC-MS/MS, by decreasing sample complexity	Methyl esterification of acidic residues may eliminate non-specific binding
iTRAQ* (Isobaris tag for relative and absolute quantitation)	Isobaris tagging of all primary amines; is LC-MS/MS based quantitative method	Compares PTM's and subproteomes; allows comparison of 4 samples simultaneously	Difficulty in detection with low level of proteins
Label-Free	Peak integration method; the signal intensity of peptides ions within an MS scan can be compared en mass from multiple (LC-MS) analyses	High mass accuracy	Unclear whether dynamic exclusion rules applied during mass spectral acquisition invalidate spectral counting
Spectral counts	Calculates peak ion intensity from MS scans , tabulates the frequency of these MS/MS (of the same precursor ion) scans that reflect the abundance of peptides in sample	More sensitive	Unclear whether dynamic exclusion rules applied during mass spectral acquisition invalidate spectral counting

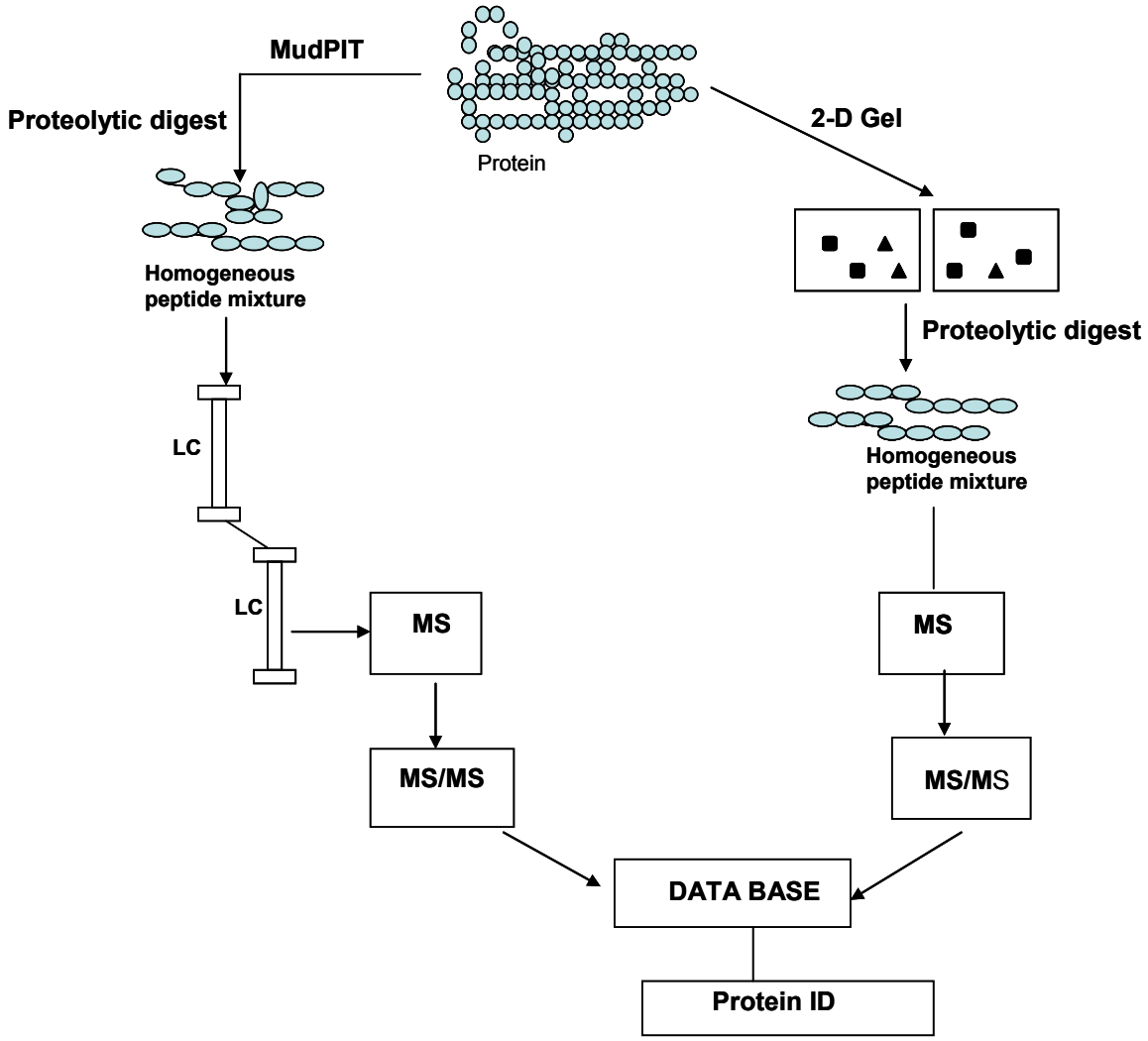

Figure L1. Represents MudPIT versus 2-D Gel methodologies for large scale protein identification. The 2D gel based methodology separates proteins from cell lysate by iso- electric focusing and molecular weight, gel spots of interest are excised, digested for MS or MS/MS analysis. MudPIT methodology involves digestion of proteins from cell lysate and separation of peptides via 2 dimensional chromatography and MS/MS analysis (Rose, 2004).

1.3 Plants:

The plants and the animal kingdoms evolved independently from unicellular eukaryotes and represent highly contrasting life forms. The nematode worm, *Caenorhabditis elegans,* and the fruitfly, *Drosophila melanogaster,* genomes sequenced revealed metazoans share a great deal of genetic information required for physiological and developmental processes, but theses genome sequence represented very limited information of multi-cellular animal kingdom organisms (AGI, 2000; Boutros and Perrimon, 2000; Walbot, 2000). Plants have unique organizational and physiological properties in addition to their ancestral features conserved between plants and animals. An insight into these features and properties can be studied by plant genomics that provide a basic understanding between the differences of genetics of plants and other eukaryotes, and a foundation for a characterization of genes of plants (AGI, 2000). The advantages and implications of studying plants are not only relevant to plant biologists, but also have applications for evolutionary biology, molecular medicine, combinatorial chemistry, functional and comparative genetics (Meinke et al., 1998).

1.3.1 *Arabidopsis thaliana*:

A. thaliana is a small flowering plant, commonly considered a weed in nature, and has a relatively short life cycle. It is in the same family as many food plants such as canola, cabbage, cauliflower, broccoli, turnip, rutabaga, kale, Brussel sprouts, kohlrabi and radish. It is a member of the mustard family (Brassicaceae) with broad natural distribution throughout Europe, Asia, and North America (Martienssen and McCombie, 2001). Arabidopsis plants are very small in size, with over one thousand able to grow and reproduce in the space of this page. Arabidopsis's small size is advantageous to researchers and plant biologists (Walbot, 2000), making it

convenient to be cultivated in laboratory conditions. Arabidopsis is one of the most widely used model organisms for studying the biology of higher plants. Its small genome (approximately 130Mb) and low amounts of repetitive DNA mean that it is well suited for genetic and physical mapping. This simple angiosperm has served as not only a model for plant biology but also for other eukaryotes, as it has played a major role in the understanding of basic biological principles relevant to many other organisms including human (Martienssen and McCombie, 2001; Meinke et al., 1998). Most developmental and physiological processes in Arabidopsis, as well as genes controlling them have counterparts in crop plants (O'Neill and Bancroft, 2000). The Arabidopsis genome sequence provides a valuable resource for identifying and evaluating sets of candidate genes that may account for complex traits in other organisms. The Arabidopsis genome sequence information is applicable to major crops including corn and soybeans, because of the similarities among the genomes of all flowering plants (Lukens et al., 2003; Paterson et al., 2001).

1.3.2 Environmental effects on plants:

Environmental factors play a very significant role in the development and growth of plants. The three most ecologically important factors effecting plant growth are light, temperature, and water. Plant growth and development are controlled by internal regulators that are modified according to environmental conditions, they possess a finite capacity to acclimate to physical and chemical mechanisms to protect themselves against stressful environmental conditions (Dat et al., 2000; Mahan et al., 1995; Neilson et al., 2010). In our experiment we studied the effect of nitrogen on the growth of *A. thaliana*.

1.3.3 Effect of Nitrogen:

Nitrogen plays a critical role in plant growth. Nitrogen uptake in plants occurs in two forms: ammonium ion (NH_4^+) and the ion nitrate (NO_3^-). Nitrate is the principle form of N acquired by plants when present in adequate amounts, and it can be both actively and passively absorbed (Taiz, 1998). Plants reduce NO_3^- to nitrite in the cytosol via the light-dependent enzyme nitrate reductase. Nitrite can then be stored in the vacuole or converted via NH_4^+ into organic molecules. Soil NH_4^+ passively diffuses across plant membranes, and is then converted directly into organic compounds (Taiz, 1998). However, NH_4^+ must be rapidly converted into organic molecules, since free NH_4^+ can damage redox reactions in the photosynthetic pathway. Plants tolerate much higher levels of substrate NO_3^- than NH_4^+ (Marschner, 1997). Most plants obtain the nitrogen they need as inorganic nitrate from the soil/water matrix. Ammonium is used less by plants for uptake because in large concentrations it is extremely toxic.

Limiting nitrogen will restrict the growth of plant organs through a reduction of protein synthesis, since it is an essential component for amino acid, proteins, nucleic acid, and enzymes. This effect will largely account for stunting (have reduced chlorophyll a and chlorophyll b, resulting in chlorisis) of plants (Arney, 1952; Pidwirny, 2006). Conversely, high amounts of nitrogen may result in toxicity. Toxicity has been reported in cabbage at nitrogen rates of 602 mg N/L or higher, which results in severe yield reduction. Thus, nitrogen management in plants is essential to achieve maximum growth rates (Huett, 1989). The type as well as the amount of nitrogen plays an important role, since the nitrate ion is tolerated at a higher level in comparison to ammonium ion (Lefsrud, 2006).

1.4 Objective of the research:

The objective of our experiment was to study and monitor plant growth of *A. thaliana* grown under agronomic conditions at low and high N fertilization regimes, to better understand how different proteins are up-regulates and down regulates under different environmental conditions and how this impacts the plants growth. Using comparative shotgun proteomics to characterize *A. thaliana* at the proteomic level it should be possible to develop a whole plant physiological study combined with gene, protein, and metabolite profiling to build up a comprehensive picture depicting the different levels of N uptake and assimilation.

1.5 Future Perspective:

This research will provide an avenue in future to further explore the following: plant responses to N as being essential to elucidate the regulation of N-use efficiency, adaptive responses of plants by highlighting master traits controlling growth under each nutritional condition and provide key target selection criteria for breeders and monitoring tools for farmers for conducting a reasoned fertilization protocol.

In future forth coming research, we can compare Arabidopsis to kale (*Brassica oleracea L.var. acephala*); both the plants belong to the same family: Brassicaceae. Previous studies of proteins and mitochondrial DNA have predicted that, Brassica and Arabidopsis have evolved 16 to 19 Mya (million years ago). While a perfect organism may not exist Brassica seems well placed due to its closeness to Arabidopsis have shown as average of 87% of conservation in the coding region. Previous studies have also indicated that *B. oleracea* linkage maps and *A. thaliana* genome identified numerous one to one segmental relationships, and apparent genome duplication, in addition to genome triplication (Katari et al., 2005; Koch et al., 2000; Quiros et

al., 2001; Town et al., 2006; Yang et al., 1999). Arabidopsis one of the most closely related plant to the genus Brassica; therefore it is an obvious choice for evaluating comparative genomic approaches to understanding and manipulating biological processes and traits in crops.

1.6 References

Aebersold, R., and M. Mann. 2003. Mass spectrometry-based proteomics. *Nature*. 422:198(110).

AGI. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 408:796-815.

Arney, S.E. 1952. Some effects of nitrogen nutrition on the morphology and anatomy of marrow-stem kale. *Annals of Applied Biology*. 39:266-276.

Barbier-Brygoo, H., and J. Joyard. 2004. Focus on plant proteomics. *Plant Physiol Biochem*. 42:913-917.

Bjellqvist, B., K. Ek, P. Giorgio Righetti, E. Gianazza, A. Görg, R. Westermeier, and W. Postel. 1982. Isoelectric focusing in immobilized pH gradients: Principle, methodology and some applications. *Journal of Biochemical and Biophysical Methods*. 6:317-339.

Colas, I., O. Koroleva, and P.J. Shaw. 2010. Mass spectrometry in plant proteomic analysis. *Plant Biosystems*. 144:703-714.

Dat, J., S. Vandenabeele, E. Vranová, M. Van Montagu, D. Inzé , and F. Van Breusegem. 2000. Dual action of the active oxygen species during plant stress responses. *Cellular and Molecular Life Sciences*. 57:779-795.

David W. Meinke, J. Michael Cherry, Caroline Dean, S.D.R. and, and M. Koornneef. 1998. *Arabidopsis thaliana*: A Model Plant for Genome Analysis. *Science*. 282:662-682.

Drexler, D., D.J. Barlow, P. Falk, J. Cantone, D. Hernandez, A. Ranasinghe, M. Sanders, B. Warrack, and F. McPhee. 2006. Development of an on-line automated sample clean-up method and liquid chromatography–tandem mass spectrometry analysis: application in an in vitro proteolytic assay. *Analytical & Bioanalytical Chemistry*. 384:1145-1154.

Fahey, J.W. 2003. Brassicas. *In* Encyclopedia of Food Sciences and Nutrition. C. Benjamin, editor. Academic Press, Oxford. 606-615.

Froehlich, J.E., C.G. Wilkerson, W.K. Ray, R.S. McAndrew, K.W. Osteryoung, D.A. Gage, and B.S. Phinney. 2003. Proteomic Study of the Arabidopsis thaliana Chloroplastic Envelope Membrane Utilizing Alternatives to Traditional Two-Dimensional Electrophoresis. *Journal of Proteome Research*. 2:413-425.

Glinski, M., and W. Weckwerth. 2006. The role of mass spectrometry in plant systems biology. *Mass Spectrometry Reviews*. 25:173-214.

Görg, A., C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, and W. Weiss. 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients.*Electrophoresis*. 21:1037-1053.

Gygi, S.P., B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, and R. Aebersold. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*. 17:994-999.

Haynes, P.A., and T.H. Roberts. 2007. Subcellular shotgun proteomics in plants: Looking beyond the usual suspects. *Proteomics*. 7:2963-2975.

Huett, D.O. 1989. Effect of nitrogen on the yield and quality of vegetables. *Acta Hort. (ISHS)*. 247:205-210.

Katari, M.S., V. Balija, R.K. Wilson, R.A. Martienssen, and W.R. McCombie. 2005. Comparing low coverage random shotgun sequence data from *Brassica oleracea* and *Oryza sativa* genome sequence for their ability to add to the annotation of *Arabidopsis thaliana*. *Genome Research*. 15:496-504.

Kenrick, K.G., and J. Margolis. 1970. Isoelectric focusing and gradient gel electrophoresis: A two-dimensional technique. *Analytical Biochemistry*. 33:204-207.

Koch, M.A., B. Haubold, and T. Mitchell-Olds. 2000. Comparative Evolutionary Analysis of Chalcone Synthase and Alcohol Dehydrogenase Loci in Arabidopsis, Arabis, and Related Genera (*Brassicaceae*). *Molecular Biology and Evolution*. 17:1483-1498.

Lefsrud, M.G. 2006. Environmental manipulation to increase the nutritional content in leafy vegetables. *In* Department of Plant Sciences Vol. Ph.D. The University of Tennessee, Knoxville.

Ma, B., K. Zhang, C. Hendrie, C. Liang, M. Li, A. Doherty-Kirby, and G. Lajoie. 2003. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*. 17:2337-2342.

MacCoss, M.J., W.H. McDonald, A. Saraf, R. Sadygov, J.M. Clark, J.J. Tasto, K.L. Gould, D. Wolters, M. Washburn, A. Weiss, J.I. Clark, and J.R.Y. Iii. 2002. Shotgun Identification of Protein Modifications from Protein Complexes and Lens Tissue. *Proceedings of the National Academy of Sciences of the United States of America*. 99:7900-7905.

MacKay, V.L., X.H. Li, M.R. Flory, E. Turcott, G.L. Law, K.A. Serikawa, X.L. Xu, H. Lee, D.R. Goodlett, R. Aebersold, L.P. Zhao, and D.R. Morris. 2004. Gene expression analyzed by high-resolution state array analysis and quantitative proteomics - Response of yeast to mating pheromone. *Molecular & Cellular Proteomics*. 3:478-489.

Mahan, J.R., B.L. McMichael, and D.F. Wanjura. 1995. Methods for reducing the adverse effects of temperature stress on plants: A review. *Environmental and Experimental Botany*. 35:251-258.

Martienssen, R., and W.R. McCombie. 2001. The First Plant Genome. *Cell*. 105:571-574.

Michael P. Washburn, D.W., & John R. Yates III. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*. 19:242-247.

Neilson, K.A., C.G. Gammulla, M. Mirzaei, N. Imin, and P.A. Haynes. 2010. Proteomic analysis of temperature stress in plants. *Proteomics*. 10:828-845.

Nesvizhskii, A.I., and R. Aebersold. 2005. Interpretation of Shotgun Proteomic Data. *Molecular & Cellular Proteomics*. 4:1419-1440.

O'Neill, C.M., and I. Bancroft. 2000. Comparative physical mapping of segments of the genome of *Brassica oleracea* var. *alboglabra* that are homoeologous to sequenced regions of chromosomes 4 and 5 of Arabidopsis *thaliana. The Plant Journal*. 23:233-243.

Oeljeklaus, S., H.E. Meyer, and B. Warscheid. 2009. Advancements in plant proteomics using quantitative mass spectrometry. *Journal of Proteomics*. 72:545-554.

Peck, S.C. 2005. Update on Proteomics in Arabidopsis. Where Do We Go from Here? *Plant Physiology*. 138:591-599.

Pidwirny, M. 2006. The Nitrogen Cycle. *In* Fundamentals of Physical Geography. Vol. *2nd Edition.* Date Viewed. March 18th 2011. http://www.physicalgeography.net/fundamentals/9s.html.

Porath, J., J. Carlsson, I. Olsson, and G. Belfrage. 1975. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature*. 258:598-599.

Powell, A.D., and W.H. Flurkey. 2006. Immobilized metal affinity chromatography (IMAC) of mushroom tyrosinase. *Abstracts of Papers of the American Chemical Society*. 231:718-CHED.

Quiros, C.F., F. Grellet, J. Sadowski, T. Suzuki, G. Li, and T. Wroblewski. 2001. Arabidopsis and Brassica Comparative Genomics: Sequence, Structure and Gene Content in the ABI1-Rps2-Ck1 Chromosomal Segment and Related Regions. *Genetics*. 157:1321-1330.

Rampitsch, C., and M. Srinivasan. 2006. The application of proteomics to plant biology: a review. *Canadian Journal of Botany-Revue Canadienne De Botanique*. 84:883-892.

Schulze, W.X., and B. Usadel. 2010. Quantitation in Mass-Spectrometry-Based Proteomics. *Annual Review of Plant Biology, Vol 61*. 61:491-516.

Shevchenko, A., S. Sunyaev, A. Loboda, A. Shevchenko, P. Bork, W. Ens, and K.G. Standing. 2001. Charting the Proteomes of Organisms with Unsequenced Genomes by MALDI-Quadrupole Time-of-Flight Mass Spectrometry and BLAST Homology Searching. *Analytical Chemistry*. 73:1917-1926.

Tabb, D.L., W.H. McDonald, and J.R. Yates. 2002. DTASelect and Contrast: Tools for Assembling and Comparing Protein Identifications from Shotgun Proteomics. *Journal of Proteome Research*. 1:21-26.

Taiz, L., and E. Zeiger. 1998. Plant Physiology. Sinauer Associates, Inc., Sunderland, Mass.

Thelen, J.J., and S.C. Peck. 2007. Quantitative proteomics in plants: Choices in abundance. *Plant Cell*. 19:3339-3346.

Town, C.D., F. Cheung, R. Maiti, J. Crabtree, B.J. Haas, J.R. Wortman, E.E. Hine, R. Althoff, T.S. Arbogast, L.J. Tallon, M. Vigouroux, M. Trick, and I. Bancroft. 2006. Comparative Genomics of *Brassica oleracea* and *Arabidopsis thaliana* Reveal Gene Loss, Fragmentation, and Dispersal after Polyploidy. *The Plant Cell Online*. 18:1348-1359.

Tyers, M., and M. Mann. 2003. From genomics to proteomics. *Nature*. 422:193(195).

von Mering, C., R. Krause, B. Snel, M. Cornell, S.G. Oliver, S. Fields, and P. Bork. 2002. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*. 417:399-403.

Weckwerth, W., K. Wenzel, and O. Fiehn. 2004. Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics*. 4:78-83.

Wienkoop, S., M. Glinski, N. Tanaka, V. Tolstikov, O. Fiehn, and W. Weckwerth. 2004. Linking protein fractionation with multidimensional monolithic reversed-phase peptide chromatography/mass spectrometry enhances protein identification from complex mixtures even in the presence of abundant proteins. *Rapid Communications in Mass Spectrometry*. 18:643-650.

Wiese, S., K.A. Reidegeld, H.E. Meyer, and B. Warscheid. 2007. Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *Proteomics*. 7:340-350.

Wu, W.W., G.H. Wang, S.J. Baek, and R.F. Shen. 2006. Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. *Journal of Proteome Research*. 5:651-658.

Yang, Y.W., K.N. Lai, P.Y. Tai, and W.H. Li. 1999. Rates of Nucleotide Substitution in Angiosperm Mitochondrial DNA Sequences and Dates of Divergence Between *Brassica*; and Other Angiosperm Lineages. *Journal of Molecular Evolution*. 48:597-604.

Zhang, Q., and D.E. Riechers. 2008. Proteomics: An emerging technology for weed science research. *Weed Science*. 56:306-31

Connecting Statement

The literature review from Chapter 1 epitomizes the quantitative methods used in plant proteomics and provides a perspective of the research done in Chapter 2. In Chapter 2, a study was performed to monitor plant growth of *A. thaliana* grown under agronomic conditions at low and high N fertilization regimes, to better understand how different proteins are up-regulates and down regulates under different environmental conditions, and how this impacts the plants growth.

The co -authors contributing to the research in Chapter 2 and their corresponding physical addresses are as follows:

Dean Kopsell[2] provided help on plant culture, Robert L Hettich[3] provided guidance and laboratory and mass spectrometry time for sample analysis, Manesh Shah[3] and Eric provided guidance on data analysis, Paul Abraham[3] provided guidance on sample preparation; Nathan C VerBerkmoes[3] provided guidance on mass spectrometry and sample preparation.

Mark G Lefsrud[1] provided overall guidance, approbation and supervision.

[1]McGill University, Bioresource Engineering, 21,111 Lakeshore, Ste-Anne-de-Bellevue, QC H9X 3V9, Canada. [2]Plant Science Department, University of Tennessee, 2431 Joe Johnson Drive, 252 Ellington Building, Knoxville, TN, 37996, [3]Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

Chapter 2

Proteomic Comparison of *Arabidopsis thaliana* Under High and Low Nitrogen Fertilization

2.1 Introduction:

The goal of proteomics is the ability to identify and monitor protein synthesis within growing organisms. Plants are one of the most challenging groups of organisms to measure at a protein level because of the production and synthesis of highly abundant proteins, such as ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and the high level of chlorophyll. RuBisCo accounts for an average of 56% of the protein within wheat (*Triticum aestivum* L.) (Pal et al., 2005). With this large amount of a single protein overwhelming many analytical techniques, methods have been developed to remove RuBisCO (Kim et al., 2001) or focused on cellular subtractions / parts of the plant that have no or limited RuBisCO production (roots, pollen grains, seeds, mitochondria) (Agrawal et al., 2008; Basu et al., 2006; Carpentier et al., 2008; Cho et al., 2008; Robinson et al., 1988; Sheoran et al., 2007; Wen et al., 2007).

The use of proteomics to separate, identify and quantify proteins from higher organisms have relied on SDS-PAGE or IPG-IEF gels (Espagne et al., 2007; Geisler-Lee et al., 2007; Jamet et al., 2008; Lee et al., 2007; Lei et al., 2005). In the gel method proteins are separated based on size (1-D) and/or charge (2-D) (Carpentier et al., 2008). After staining to identify gel locations, the spots are excised, digested into peptides and positively identified by liquid chromatography mass spectrometry / mass spectrometry (LC-MS/MS). Multidimensional Protein Identification Technology (Agrawal and Rakwal, 2008; Finnie, 2006), with on-line electrospray on rapid scanning tandem mass spectrometers coupled with robust informatics tools allows for the rapid separation and identification of complex peptide mixtures derived from proteomes. Through the use of shotgun proteomics, which refers to the global analysis of the digested products of protein mixtures such as tissues, cells, or protein complexes, these proteins are proteolytically reduced to peptides. Peptide separation occurs by loading a peptide sample onto a SCX (strong cation exchange) phase column which is then connected to a reverse phase separation column. Two-dimensional separation uses a high-pressure liquid chromatography pump which separates the

peptide mixture by charge and hydrophobicity. The liquid chromatography system is directly coupled to the tandem mass spectrometer (LC-MS/MS) where intact peptide masses and rapid data dependent tandem mass spectrometer spectra are obtained (McDonald et al., 2002; Washburn et al., 2002). These acquired tandem mass spectrometer spectra are searched against a predicted protein database using a computer algorithm such as SEQUEST (Eng et al., 1994a); Mascot(Perkins et al., 1999), or X! Tandem (Brosch et al., 2008).

The LC-MS/MS system has the ability to simplify the identification and quantification of protein expression within all organisms. In general, proteomics allows the assignment of proteins to an organism because the genes that encode them have been sequenced and annotated for specific organisms and specific proteins. Developing these DNA databases is dependent on properly sequencing and identifying DNA from individual model organisms. This procedure has been used to investigate microbial communities with unheralded protein identification (Denef et al., 2007; Lo et al., 2007; VerBerkmoes et al., 2009; Wilmes and Bond, 2006).

The family Brassicaceae *(*Cruciferae*)* is defined by their sulfur containing plant compounds called glucosinolates (Judd, 1999). Within this family is the tribe Brassiceae and Arabidae for which the *Brassica* and *Arabis* genus exists. The *Arabis* genus is also known as rockcress, encompasses the standard model plant *Arabidopsis thaliana*. *A. thaliana* has little direct impact on agriculture, but as a model plant has extended the understandings of genetic, cellular, and molecular biology of flowering plants. *A. thaliana* was the first genetically sequenced plant (AGI, 2000) and has a relatively small genetic size of five chromosomes and around 157 million base pairs (Bennett et al., 2003). *A. thaliana* has 27,250 genes which encode 34,522 proteins. The *Brassica* genus encompasses a diverse group of plants that grow all over the world and is comprised of a number of common plants such as cabbage (*Brassica oleracea* var *capitata*), cauliflower (*B. oleracea* var *botrytis*), canola/rapeseed (*B. napus* L.), mustard (*B. juncea*), and collards/kale (*B. oleracea* var *acephala*). It is estimated that the ancestor of the *Brassica* and *Arabis* genus diverged over 28 Myr ago (Wroblewski et al., 2000).

Nitrogen (N) plays an important role in plant growth. It is an essential component for the production of amino acid, proteins, nucleic acid, and enzymes in plants. The starvation of N will restrict the growth of plant organs through a reduction of protein synthesis. This effect will

largely account for stunting (have reduced chlorophyll *a* and chlorophyll *b*, resulting in chlorosis) of plants (Arney, 1952; Pidwirny, 2006). Conversely, high amounts of N may account for toxicity. A balanced N levels in plants is essential to achieve maximum growth rates (Huett, 1989). Nitrogen uptake in plants occurs in two forms: ammonium ion (NH_4^+) and the ion nitrate (NO_3^-). Most plants obtain the N they need as inorganic nitrate from the soil solution. Plants reduce NO_3^- to nitrite via light dependent enzyme reductase. Ammonium is used less by plants for uptake because in large concentrations it is extremely toxic and NH_4^+ is rapidly converted to organic molecules, since free NH_4^+ can damage redox activity in photosynthesis pathways (Pidwirny, 2006; Taiz, 1998). Hence, the type as well as the amount of N plays an important role since the NO_3^- ion is tolerated at a higher level in comparison to NH_4^+ ion (Lefsrud, 2006). Since N uptake can be controlled by producers, it is not very clear what would be the effect of N on secondary plant compounds, other cellular interactions and genetic differences.

The objective of our experiment was to study and monitor plant growth of *A. thaliana* grown under agronomic conditions at low and high N fertilization regimes, to better understand how and what up-regulates and down regulates different proteins under different environmental conditions and how this impacts the plants growth using comparative shotgun proteomics. This research will allow for development of a whole plant physiological study combined with gene, protein, and metabolite profiling to build up a comprehensive picture depicting the different levels of N uptake and assimilation.

2.2 Material and Methods:

2.2.0 Samples and Sample Preparation

A. thaliana seeds were planted into rockwool growing cubes (Grodan A/S, Dk-2640, Hedehusene, Denmark) and germinated in a growth chamber (E15, Conviron; Winnipeg, Manitoba) under cool white fluorescent (160W) and incandescent (60W) bulbs. The growth chamber light intensity photosynthetic active radiation (PAR) was measured at 275 ± 10 μmol m^2 s^{-1} (Model QSO-ELEC, Apogee Instruments; Logan, UT) and the temperature held at 20 ± 1 °C throughout the experiment. Peter's 20N-6.9P-16.6K water-soluble fertilizer (Scotts, Marysville, OH) was applied every five days at a rate of 200 mg/L. After 2 weeks, the plants were transferred to 11 L plastic containers (Rubbermaid Inc., Wooster, OH). Six plants were

placed into 2 cm round holes at 10.6 by 9.5 cm spacing in each container lid. The containers were placed into the growth chambers. The plants were grown in 10 L of half strength modified Hoagland nutrient solution (Hoagland and Arnon, 1950).

Plants were grown under two N treatment levels of 6 and 106 mg/L. The ratio of NO_3-N to NH_4-N was kept constant at 3:1 and solutions were changed every two weeks. Elemental concentrations of the nutrient solutions were (mg/L): P (15.3), K (117), Ca (80.2), Mg (24.6), S (32.0), Fe (0.5), B (0.25), Mo (0.005), Cu (0.01), Mn (0.25), and Zn (0.025). The electrical conductance of the starting nutrient solution was 7S m^{-1} and pH was measured at 5.6. Solutions were aerated with an aquarium air pump (MK-1504, Wal-Mart, Bentonville, AR) connected to air stones. Deionized water was added daily to maintain 10 L in each container. Nitrogen treatments were randomized and replicated three times in space.

Plants were harvested after 4 weeks in the hydroponic system. At harvest, shoot and root tissues were separated. The whole *A. thaliana* plant was removed and all plants from a treatment replicate were combined for protein analysis. Plant samples were stored at -80 °C prior to protein extraction.

2.2.1 Cells Lysis and Protein Extraction:

Proteins were obtained from the leaf tissue biomass after grinding tissue in liquid nitrogen with a mortar and pestle. Ground frozen plant powder (1g) was collected in a 50 ml centrifuge tube, then 5 ml of guanidine mixture (6M guanidine, 10mM DTT and 50mM Tris pH 7.6 10mM CaCl) was added to the tube and placed in a 60 °C water bath for 1 h with the sample vortexed every 15 min for 1 min. The sample was diluted 5 fold with tris buffer, with a final pH between 7 and 8 then 40 µg of trypsin was added to each tube and rocked at 37 °C for 12 h, where another 40 µg was added and rocked for 12 h. Finally 10 mM DTT was added to the sample and rocked for 1 h, before being spun down at 10,000 g_n for 5 min. The supernatant was removed and peptides desalted and concentrated with Solid Phase extraction (C18 Sep-Pak Plus filters (Waters Corporation, Milford, MA)).The final sample was filtered through an Ultrafree-MC 45 µm spin filter (Millipore, Billerica, MA). The peptide sample was stored at -80°C prior to analysis on the mass spectrometer.

2.2.3 LC-MS/MS and Informatics:

An overview of the 2D LC-MS/MS process is provided by(Wilmes et al., 2008). The peptide samples were directly loaded onto an in house packed bi phasic back column. The back column consisted of strong cation exchange chromatographic resin with a C_{18} reverse - phase packing material loaded sequentially (~3-5cm SCX and 3-5cm C_{18}) (Polymicro technologies, Phoenix, AZ). The back column was then connected between the LC pump and a 15 cm C_{18} analytical column (packed in house) sprayed into the mass spectrometer with an electric current applied to provide ionization and transport into the mass spectrometer, a linear ion trap LTQ (Thermo Fisher Scientific, San Jose, CA, USA). The LTQ settings were as follows: all data-dependent MS/MS in LTQ (top five), two microscans for both full and MS/MS scans, centroid data for all scans and two microscans averaged for each spectrum, dynamic exclusion set at 1. 12 solvent gradient steps, a total time of 23 h using water, acetonitrile and a salt pulse (ammonium acetate) were used in this 2-D method (Wilmes et al., 2008; Wilmes and Bond, 2006). During the entire chromatographic run the mass and charge of the precursor ions and product ions were collected to output raw mass spectrum MS/MS data which was then converted into a MS2 data file. Two N treatments with MudPIT technical replicates (n=4) were run.

According to the SEQUEST (Eng et al., 1994a; Eng et al., 1994b) cross-correlation score (XCorr) and the SEQUEST normalized difference in cross-correlation score (DeltaCN), all the MS/MS spectra validity of matches were assessed in DTA Select (Tabb et al., 2002). The search results were grouped by charge state (+1, +2, and +3) and tryptic status (fully tryptic, half-tryptic, and nontryptic). Only proteins identified with two fully tryptic peptides were retained. Containment proteins (human keratin, IgG receptors) were included in the search database. The distribution of XCorr and DeltaCN values for the direct and decoy database hits were obtained, and the subsets obtained were separated by quadratic discriminant analysis. For positive protein identification, a minimum of 2 peptides were required per protein (Ram et al., 2005).

False discovery rates were estimated by reverse database searching. Reverse protein sequences of *A. thaliana* were included in the search database to estimate the overall false-positive rates of protein identification using the reverse database method (false-positive rate =

2[nrev/(nrev + nreal)] *100, with nrev = the number of peptides identified from the reverse database; and nreal = the number of peptides identified from the real database. Differentially expressed proteins were calculated based on the spectral counts using PatternLab software (Carvalho et al., 2008b).

2.2.4 Data Parsing using Pattern Lab:

The experimental data files (The DTA Select filtered files which contain the spectral counting information) were converted into PatternLab's native data format (index and sparse matrix). The index file lists all identified proteins within all the project's assays and assigns to each a unique protein identification (PID) integer. The sparse matrix file contains rows were each one corresponds to an assay and follows the schema: class label, PID_1: $value_1$... PID_n: $value_n$, where n is the number of identified proteins for that assay. PID_i and $value_i$ correspond, respectively, to the ith protein's identification integer and its spectral count for the respective assay. The resulting sparse matrix has 6 rows obtained from each condition in triplicates (Fischer et al., 2011; Moresco et al., 2010; Salvato and de Carvalho, 2010).

2.2.5 Differentially Expressed Proteins according to AC Fold:

The PatternLab **AC Fold module** was used to pinpoint differentially expressed proteins between the two different N treatments (6 mg/L and 106 mg/L). Proteins having an absolute fold change greater than 2.5 (2.5 times more spectral counts for all peptides in the compared protein) and a Student's t test p-value of 0.01 were considered as differentially expressed. The fold change cut off of 2.5 was obtained through the AC Fold procedure as to maximize the number of proteins that statistically satisfy both the FDR and the AC test criteria. The Benjamini-Hochberg theoretical false discovery rate estimator (BH-FDR) (q-value of 0.1) and AC test p-value of 0.05 were specified (Benjamini and Hochberg, 1995). The AC Fold method was chosen because it can be applied even if the assays are not technical replicates, to search for differential protein patterns in shotgun proteomics by considering information from protein fold changes, statistical AC test and a FDR estimator.

2.2.6 Selection of unique proteins:

Proteins that were only identified in either the 6mg N/L or the 106 mg N/L treatments were assessed using PatternLab's approximately area-proportional Venn diagram (AAPVD). A selection criterion was imposed to only consider proteins that were identified in both assays for the treatment and found in no assays for the other treatments (Fischer et al., 2011; Moresco et al., 2010; Salvato and de Carvalho, 2010).

2.2.7 Gene Ontology Explorer Analysis:

PatternLab's Gene Ontology Explorer (GOEx) module (Carvalho et al., 2008a; Carvalho et al., 2009; Carvalho et al., 2008c) was used to interpret the data. Our data analysis used the gene ontology database (OBO v1.2; accessed 2 June 2010) and the human annotation file (GOA; Accessed 30 May 2010) from The Arabidopsis Information Resource (TAIR) database (Lukens et al., 2003) of genetic and molecular biology data for the model higher plant *A. thaliana.* The Gene Ontology Explorer specialist mode was used to search proteins in the Venn diagram and AC Fold results were mapped to keywords such as negative apoptosis regulation and cellular growth.

2.3 Results and Discussion:

We were able to use MudPIT to identify *A. thaliana* proteins at a global scale where physiochemical properties of the identified proteins could be studied. We identified 1387 proteins in treatment 6 mg N/L and 1652 proteins in treatment 106 mg N/L. In 6 mg N/L treatment, 5 reverse sequence protein were identified, which gives false discovery rate of 0.7%. In 106 mg N/L treatment, only 1 reverse sequence protein found, which gives false discovery of 0.1%.

2.3.1 Selection of differentially expressed proteins according to the AC Fold Methodology:

A total of 37 proteins that satisfied both the AC test and the FDR *q*-value specified cutoffs were considered differentially expressed between the two N treatments, according to the AC fold methodology (Table 1). We observed18 proteins (negative fold indicates a greater expression) in the 6 mg N/L treatment and 19 (positive fold indicates a greater expression) in the 106 mg N/L treatment. Figure 1 shows the graphical representation of the AC fold analysis, it

maps proteins according to their *p*-value for the differential expression and their fold change. Table 5 lists 35 differentially expressed proteins that are statistically important, but did not satisfy the *q*-test.

2.3.2 Selection of unique proteins:

A Venn diagram (AAPVD) analysis was generated using PatternLab to pinpoint proteins that were uniquely identified in the 6 mg N/L and 106 mg N/L treatments (Figure 2). A total of 2134 proteins were shared between the two N treatments, 567 proteins were only observed in the 6 mg N/L treatment, while 263 proteins were observed in 106 mg N/L treatment. Unique protein identified in the 106 mg N/L and 6 mg N/L treatments having average spectral counts above six were102 and 38 proteins, respectively.

2.3.3 Gene Ontology classification:

The GO pie charts represent general distribution of differentially expressed proteins (Figure 3). Differentially expressed proteins distribution of biological process, cellular component and molecular function are presented in Figures 4, 5, and 6, respectively. We identified proteins localized to cell walls, plasma membranes, chloroplast, cytoplasm, mitochondria. Proteins involved in a range of biological and molecular processes such as metabolic, developmental, and signaling processes, or protein binding, transport, nucleic acid binding, and kinase activity were also detected. However, we were not able to detect the same proportion of proteins with unknown classifications (combined with “the other function” category), which tend to make up the largest part of the distribution (Figure 3). This large number of “other” proteins stresses clearly the necessity for more functional data to categorize and interpret large scale proteomic datasets.

Differentially expressed proteins were grouped according to their biological processes, which is further subdivided into different subgroups as shown in Figure 4 and Table 4. The major subclasses were cellular processes, catabolic processes and metabolic processes. Among these metabolically active proteins were phosphoglycerate kinase (IPI00534991), pyruvate dehydrogenase (IPI00538502) and 6-phosphogluconate dehydrogenase family protein (IPI00516481, IPI00519564, IPI00545955). These proteins are related to different pathways such

as amino acid synthesis (phosphoglycerate dehydrogenase). Thereby, an influence of the respective phosphorylation on the protein function of the individual proteins and the related pathways is possible.

Besides their classification regarding the association to biological processes, PatternLab also enabled a grouping according to the molecular function and cellular function of the identified proteins. A total of 37 proteins (Figure 1 as indicated by blue dots) were selected as differentially expressed because they satisfied both the AC test and the FDR *q*-value specified cutoffs. These differentially expressed proteins were annotated by GO terms (Figure 6 and Table 2). Several proteins have been attributed to more than one single subgroup. The majority of proteins show catalytic functions of various kinds within the chloroplasts. Among these are proteins involved in the photosynthetic pathway as well as proteins such as fructose-bisphosphate aldolase or phosphoglycerate kinase. Fructose-bisphosphate aldolase is involved in glycolytic reactions and catalyzes the reaction of D-fructose 1,6-bisphosphate to glycerone phosphate and D-glyceraldehyde 3-phosphate. During this process, a reversible binding of ATP occurs in the catalytic domain. Phosphoglycerate kinase in turn catalyzes the reaction of ATP + 3-phospho-D-glycerate to ADP + 3-phospho-D-glyceroyl phosphate and again ATP has to bind to the catalytic domain. Reversible phosphorylation processes of these proteins are mandatory during their catalytic activity by binding ATP. In addition, the reversible phosphorylation of distinct amino acid could also regulate the catalytic activity itself.

Besides catalytic processes, the other main categories found for molecular function annotation by GO were; protein binding and metal ion binding. Among the proteins within this group are RNA binding proteins (IPI00542179, IPI00542840, and IPI00535689) and ion binding proteins like oxidoreductase-like protein (IPI00656816, IPI00518426, IPI00516481, IPI00519564, IPI00545955) which is predicted to be set up to the metal-containing alcohol dehydrogenase family.

Here after, in this manuscript we have discussed differentially expressed proteins that were significantly up regulated during the two N level treatments, low (6 mg N/L) and high (106 mg N/L). We have also discussed proteins (Table 5) that are statistically important.

2.3.4 Differentially expressed protein in *A. thaliana* for the 6 mg N/L treatment:

Glycine-rich RNA-binding proteins (GR-RBPs) have been implicated to play roles in post-transcriptional regulation of gene expression in plants under various stress conditions, but the functional roles of GR-RBPs under stress conditions remain to be verified. Here, we have observed IPI00538866, IPI00520057, IPI00517521 proteins (Table 1) that were up-regulated during low levels of nitrogen treatment. These proteins have biological roles of a GR-RBP, in *A. thaliana* under stress (low N) conditions (Kim et al., 2005; Vega-Garcia et al., 2010). However, no direct experimental evidence in the literature has yet been reported to clearly define the functional role of GR-RBP's under stress conditions. GR-RBP's play important roles in seed germination, seedling growth, and contributes to the enhancement of freezing tolerance in *A. thaliana* plants. As indicated in the literature, the observation that phenotypes of the wild type, *atRZ-1a* mutants, and atRZ-1a-overexpressing transgenic plants are identical throughout the Arabidopsis life cycle under normal growth conditions indicates that atRZ-1a may not be a molecular regulator displaying fundamental and basic roles during growth and development of Arabidopsis plants during normal growth conditions. This is in conformation with the goal of our study to monitor plant growth in real time that would eventually contribute to a better understanding of growth.

2.3.5 Differentially expressed protein in *A. thaliana* for the 106 mg N/L treatment:

The gene IPI00521970 in Table 1, a NADH-dependent glutamate synthase protein (UniProtKB/Swiss-Prot), was up regulated with higher levels of N, with the 106 mg N/L having a fold change of 3.51. NADH-glutamate synthase is important in the primary assimilation of ammonia, whether produced by N_2 fixation, nitrate reduction or from direct uptake and also plays an important role in the re-assimilation of ammonia during N remobilization and transport, particularly into the flower and developing seed , since it is associated in senescence associated processes (Chaffei-Haouari et al., 2011; Lea and Miflin, 2003; Miflin and Lea, 1976).

IPI00539225 is a Ferredoxin (Fd)-dependent glutamate synthase protein (Table 5). Fd-glutamate synthase activity is affected by the availability of N sources, which may be related to the availability of light (Ireland, 1999; Suárez et al., 2002). Fd-glutamate synthase activity will increase with the onset of photosynthesis and photorespiration. According to previous studies,

Fd-glutamate synthase (encoded by *GLU1*) plays a crucial role in the re-assimilation of ammonia released from glycine decarboxylation during photorespiration. The differential expression of the two glutamate synthase proteins (genes) IPI00521970 and IPI00539225 may suggest they have specific roles in N metabolism. With a 3.5 fold increase between the two treatments for glutamate synthase, N fertilization clearly impacts the production of these proteins (Lea and Miflin, 2003; Miflin and Lea, 1976).

Glycolysis is the fundamental metabolic pathway found in virtually all organisms, where hexose sugars are converted to ATP, pyruvate and substrates for various anabolic reactions (Plaxton, 1996). Our experimental results identified four proteins (IPI00531385, IPI00656816, IPI00528534, IPI00518426) that were up-regulated with higher levels of N. These proteins are involved in glycolate oxidase activity, i.e. glycolysis. Plant glycolysis utilizes sucrose and starch as principal substrates, taking place in either the plastid or the cytosol (Plaxton, 1996). A hallmark of plant cytosolic glycolysis is its flexibility to switch between alternative enzymatic reactions using ATP or pyrophosphate (PPi) as energy donors. This is believed to be modulated by factors such as tissue type, the developmental stage of the plant and various environmental stresses (at optimum levels of N in our experiment) (Plaxton, 1996)Ito et al., 2010). As with glycolysis, the pentose phosphate pathway (PPP) is a related and central metabolic pathway found in most organisms generating reductant (NADPH) and pentose sugars by two respective stages; oxidative (OPPP) and nonoxidative (Kruger and von Schaewen, 2003). NADPH is used by plants for reductive biosynthetic reactions including fatty acid synthesis and the assimilation of inorganic nitrogen and to protect against oxidative stress (Neuhaus and Emes, 2000). Pentose sugars are utilized as carbon skeletons for the synthesis of many important molecules including nucleotides, aromatic amino acids, phenylpropanoids and lignin (Allen et al., 2009; Herrmann and Weaver, 1999). In a number of plant species, both oxidative and non-oxidative stages occur in the plastid and while the oxidative stage occurs in the cytosol, it is not clear if the non-oxidative stage also takes place in the cytosol (Debnam, 1999; Schnarrenberger et al., 1995). Intermediates of PPP can be exchanged between the cytosol and plastid through a family of pentose phosphate translocators across the plastid inner envelope, which may compensate for any absence of the nonoxidative stage in the cytosol (Eicks et al., 2002; Ito et al., 2010).

In this study, we have identified one IPI00521970; NADPH dependent protein producing enzymes of oxidative PPP (pentose phosphate pathways) that also generate glucono-δ-lactone-6′-phosphate and ribulose-5′-phosphate respectively. We have identified three proteins (IPI00516481, IPI00519564, IPI00545955) that are up-regulated with high levels of N that are involved in 6-phosphogluconate dehydrogenase, which serve as rate limiting enzymes for PPP (Hou et al., 2007; Ito et al., 2010; Wakao et al., 2008) and eventually providing information on molecular mechanisms of PPP in plant responses to abiotic stresses (Hou et al., 2007). Exposed to abiotic stresses, the PPP can be enhanced through the transcript accumulation and activity increase of 6- phosphogluconate dehydrogenase (6PGDH) so as to provide more NADPH, precursors or co-factors for the biosynthesis of ligin, aromatic amino acid, phytoalexins, nucleotide acid, sugar derivatives and co-enzymes (Hauschild and von Schaewen, 2003). These products might play roles in many aspects when a plant is stressed by various abiotic factors. NADPH generated by PPP could be used for the reduction of dihydroxyacetone phosphate (DHAP), which is essential for the synthesis of glycerol (Liska et al., 2004). NADPH and G-6-P are required in lipid synthesis (Hutchings et al., 2005). NADPH may also supply electrons with biosynthesis of energy (ATP) through cytochromes or other systems, ensuring that plants can use ATP for synthesis of stress-responsive proteins. Other intermediates derived from PPP such as ribose-5-phosphate and erythrose-4-phosphate are precursors for biosynthesis of aromatic amino acids, nucleic acids and coenzymes, which are potentially involved in plant resistance or tolerance to the stresses. However, the precise function of PPP is largely not clear in plant responses to abiotic stresses. The 6PGDH genes may serve as good targets for improving plant tolerance to abiotic stresses by enhancing plant pentose phosphate pathway.

2.3.6 Differentially expressed proteins that are statistically important but did not satisfy the *q*- test:

IPI00518961, is a heat shock protein 70 (Hsp 70) indicated in Table 5. Located in the mitochondria, Hsp is usually produced only in response to environmental stress (Debel et al., 1997; Downs and Heckathorn, 1998; Lenne et al., 1995; Lenne and Douce, 1994; Vierling, 1991). It contains a conserved C-terminal ‘heat-shock domain' as well as a second conserved domain that exhibits roughly 70% amino acid homology among all known members of mitochondrial low molecular weight heat shock protein (Lund et al., 1998) and strongly

associates in a temperature-dependent manner with the matrix side of the mitochondrial inner membrane (Borovskii and Voinikov, 1993; Voinikov et al., 1998). The onset of heat stress is concomitant with increasing stress imposed by radical oxygen species (oxidative stress). Mitochondria are one of the earliest targets for these compounded effects (Richter and Kass, 1991). Polla et al. (1996) have established that the induction of heat-shock proteins protects mitochondrial function from both heat stress and oxidative injury, it acts as a determinant of the thermotolerance of oxidative phosphorylation (Downs and Heckathorn, 1998; Polla et al., 1996).

2.4 Conclusions:

Using shotgun proteomics, a high-throughput analysis method, we were able to identify *A. thaliana* proteins at a global scale where physiochemical properties of the identified proteins were unbiased. The main goal of this study was to apply a comparative shotgun proteomic approach to characterize *A. thaliana* at the proteomic level to determine the effect of N on the protein expression in plants and use individual proteins and metabolites as biomarkers to extract information about the various biochemical pathways.

We identified 37 differentially expressed proteins in *A. thaliana* grown under different N levels that were associated with NADH, TCA, some secondary metabolism, and glycolysis. The proteins associated with glycolysis indicate glutamine metabolism is of major importance in plant N economy since it provides N to young developing organs. Glutamine metabolism is usually associated with the induction of expression of enzymes such as specific proteases, the cytosolic glutamine synthases isoforms and glutamine dehydrogenases. Our study indicates that under high N level treatments, proteins responsible for glutamate synthase (GOGAT), glutamine synthase (GS), and dehydrogenase activity (DH) that serve as enzymes to catalyze a link between carbohydrate and amino-acid metabolism are up-regulated. It is therefore possible that N could be one of the factors that may affect the levels of glucosinolates in Arabidopsis. However, due to the limitation of AC fold test and the *q*-value specified cut offs, to detect differentially expressed proteins with low fold changes and low number of detected spectra, the difference detected in our analysis may be restricted to the overall proteome expression patterns of *A. thaliana*. This research will provide an avenue in future to further explore the following:

plant responses to N as being essential to elucidate the regulation of N-use efficiency, adaptive responses of plants by highlighting master traits controlling growth under each nutritional condition and provide key target selection criteria for breeders and monitoring tools for farmers for conducting a reasoned fertilization protocol.

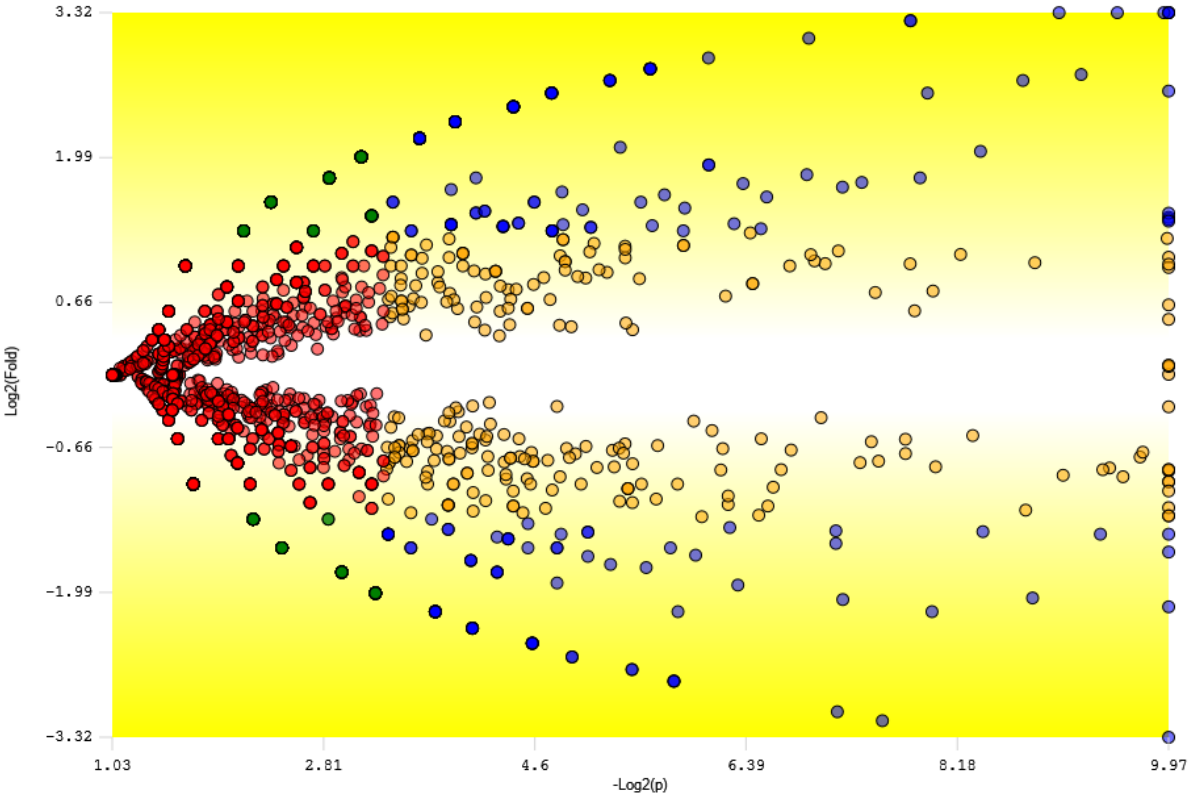

Figure 1: **Fold change versus AC test probability plot**. This plot was obtained using PatternLab's ACFold algorithm and displays the results (both *A. thaliana* nitrogen treatments 6 mg/L and 106mg/L) obtained with shotgun proteomic. Each protein (represented as a dot) was mapped according to its log2 (fold change) on the ordinate (y) axis and -log2 (1-(AC test *p*-value)) on the abscissa (x) axis. A total of 35 proteins (blue dots) were selected as differentially expressed because they satisfied both the AC test and the FDR *q*-value specified cutoffs. 48 proteins (orange dots) did not meet the fold change cutoff but were indicated as statistically differentially expressed, therefore deserving further analysis. 625 proteins (green dots) met the fold change cutoff, but the AC test indicated that this happened by chance. 1422 proteins (red dots) were pinpointed as not differentially expressed between classes because they failed both the AC test and the fold change cutoffs. The number of dots does not match the number of identified proteins due to the many overlaps. Avg. fold is 0.54, B-H FDR is 12 (0.6%), p-value 0.01

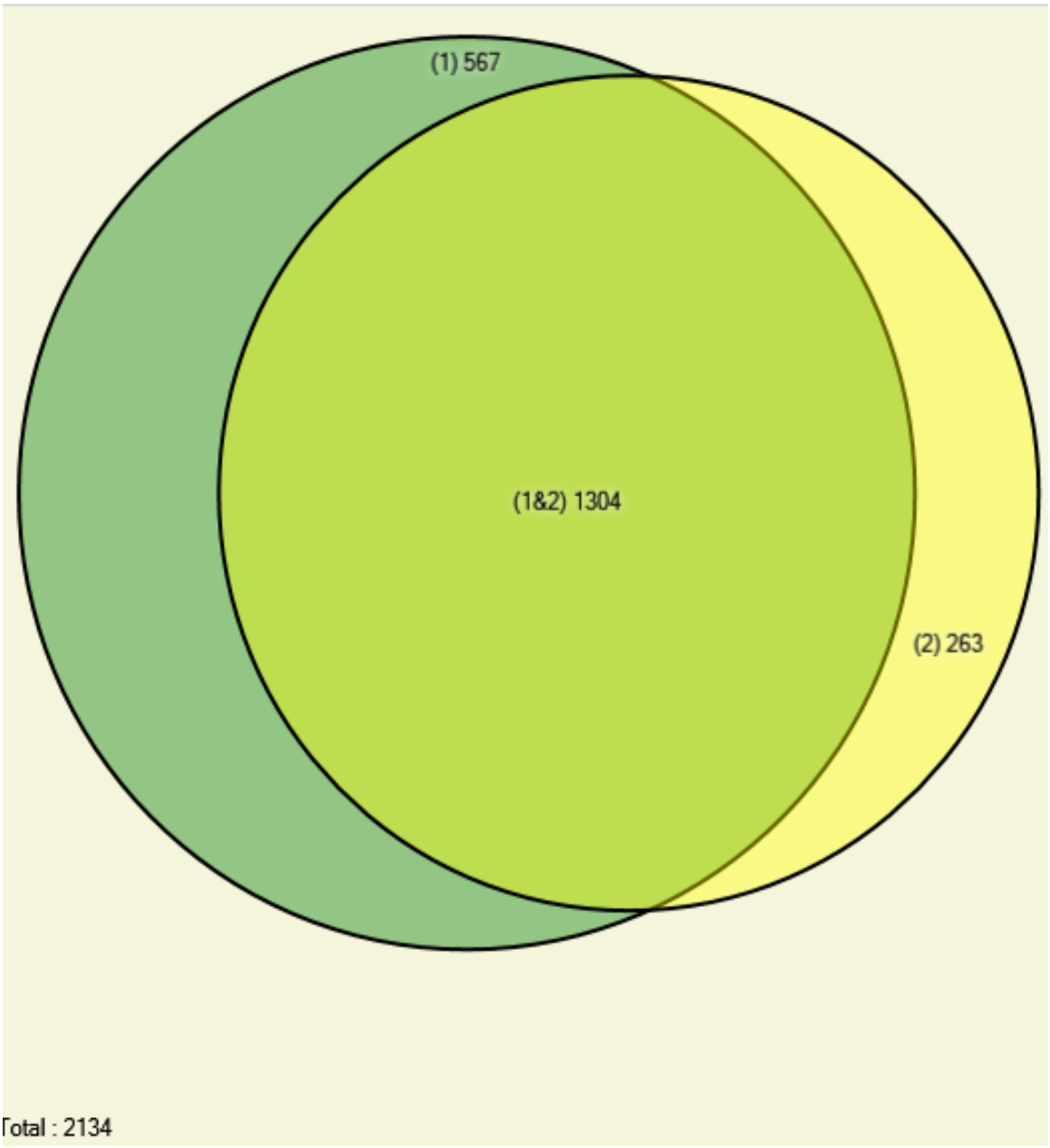

Figure 2: **Protein Expression Venn Diagram.** Approximate area-proportional Venn diagram (AAPVD) of *A*. *thaliana* provides a bird's eye view of the numbers of protein present in both (states: 106 mg N/L and 6 mg N/L) nitrogen treatments. The light green represents the number of proteins detected in both treatments. The dark green represents the number of protein detected only in 106mg/L of nitrogen treatment. The yellow represents the number of protein detected only in 6 mg N/L.

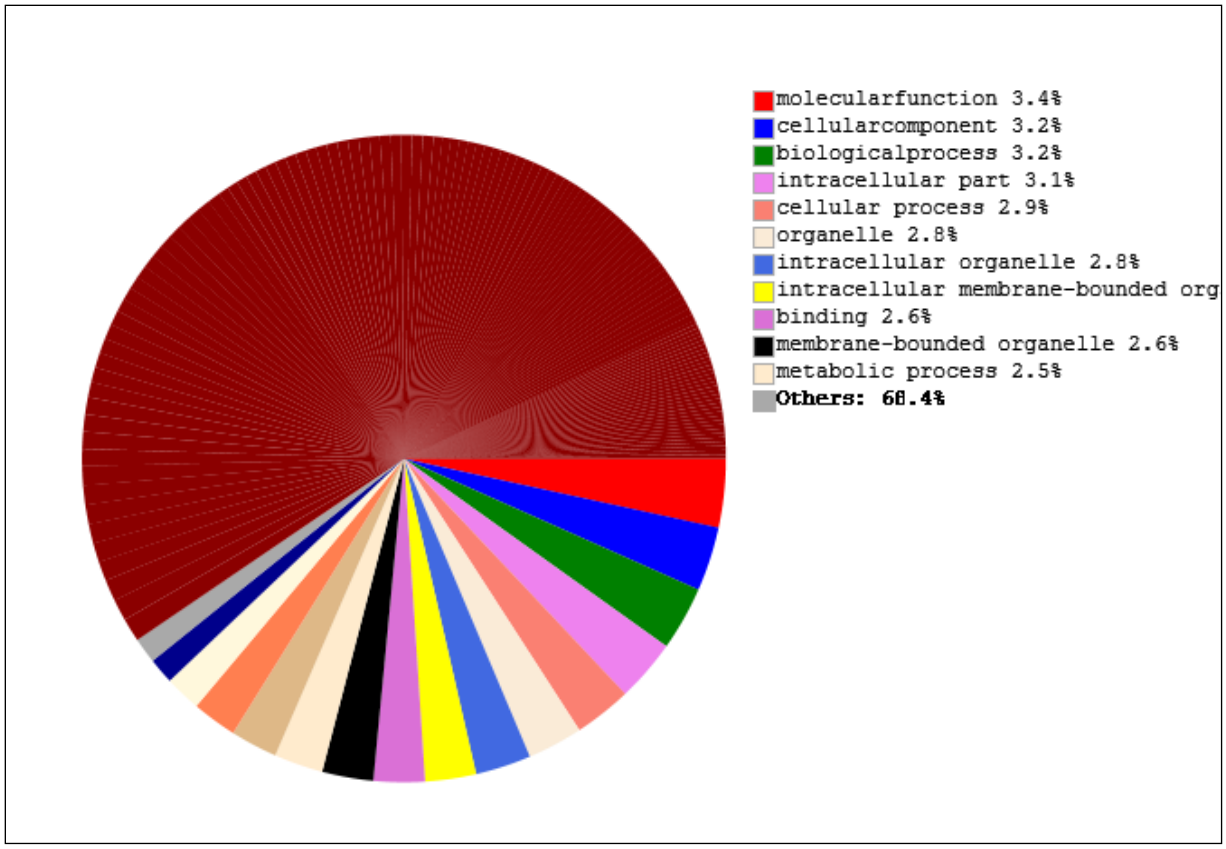

Figure 3: ***A. thaliana*** **gene ontology of the distribution of differentially expressed proteins present in both (states: 106 mg N/L and 6 mg N/L) nitrogen treatments**.

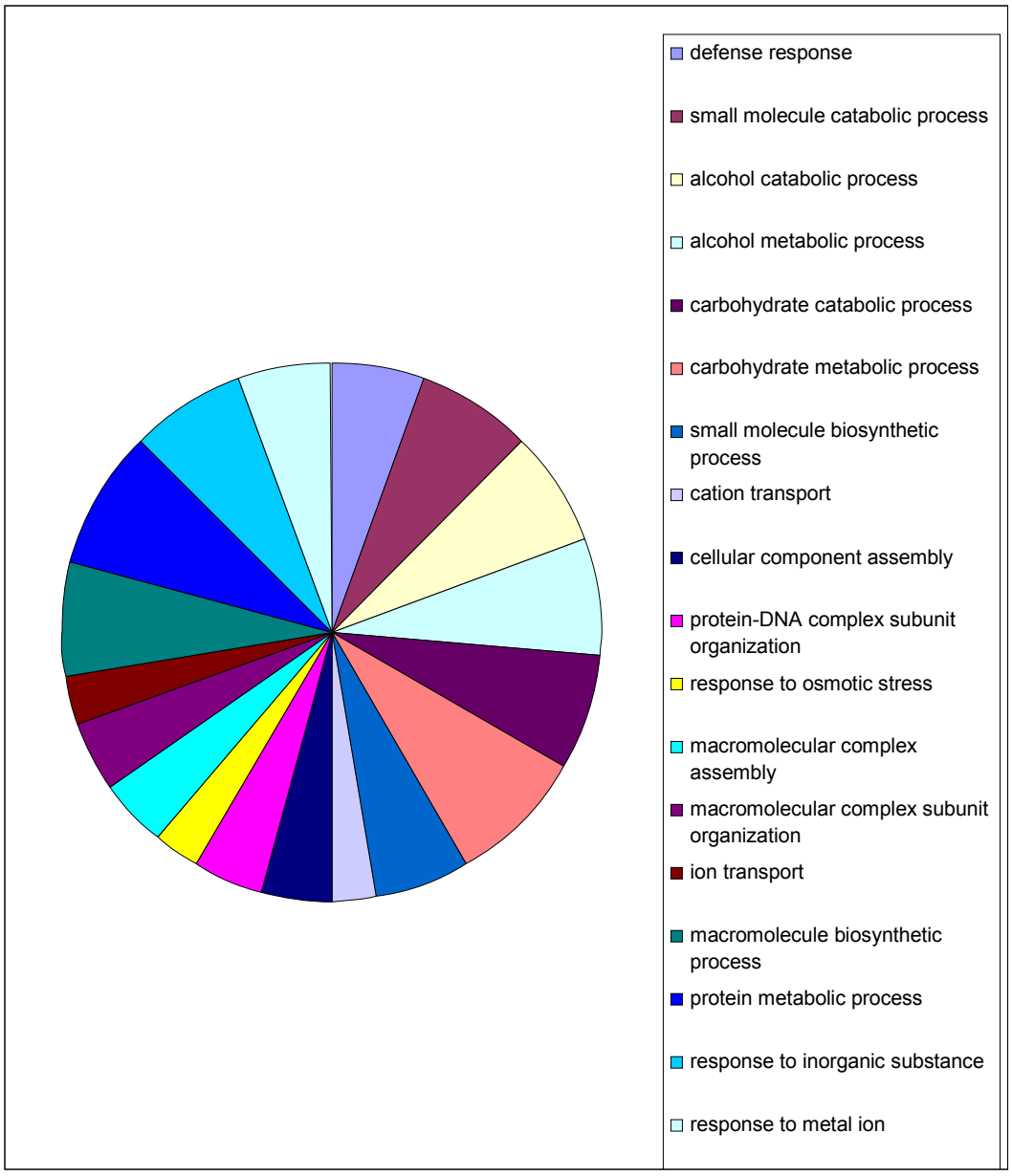

Figure 4: ***A. thaliana*** **gene ontology of biological component, of differentially expressed proteins present in both (states: 106 mg N/L and 6 mg N/L) nitrogen treatments**.

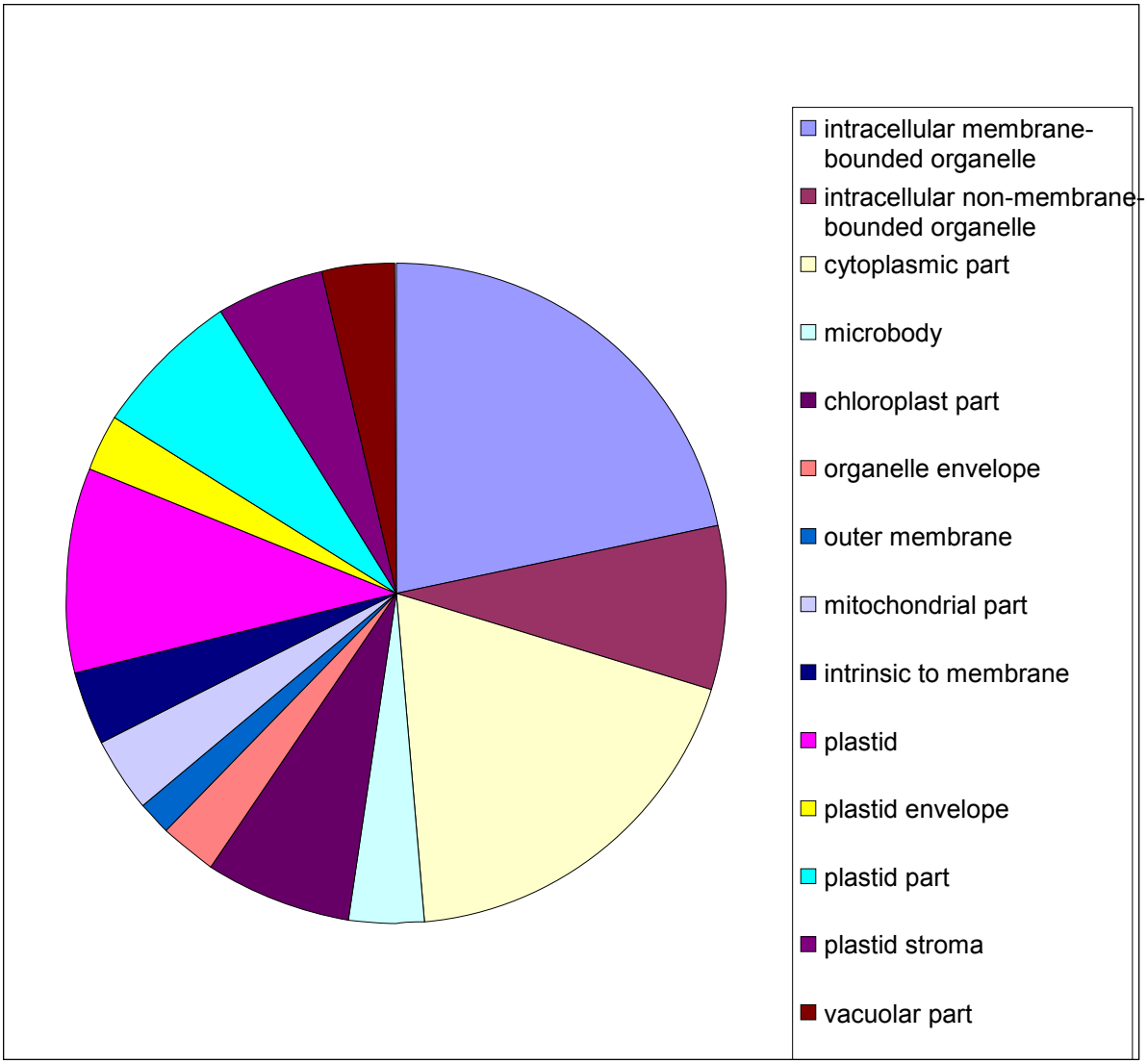

Figure 5: ***A. thaliana*** **gene ontology of cellular component, of differentially expressed proteins present in both (states: 106 mg N/L and 6 mg N/L) nitrogen treatments.**

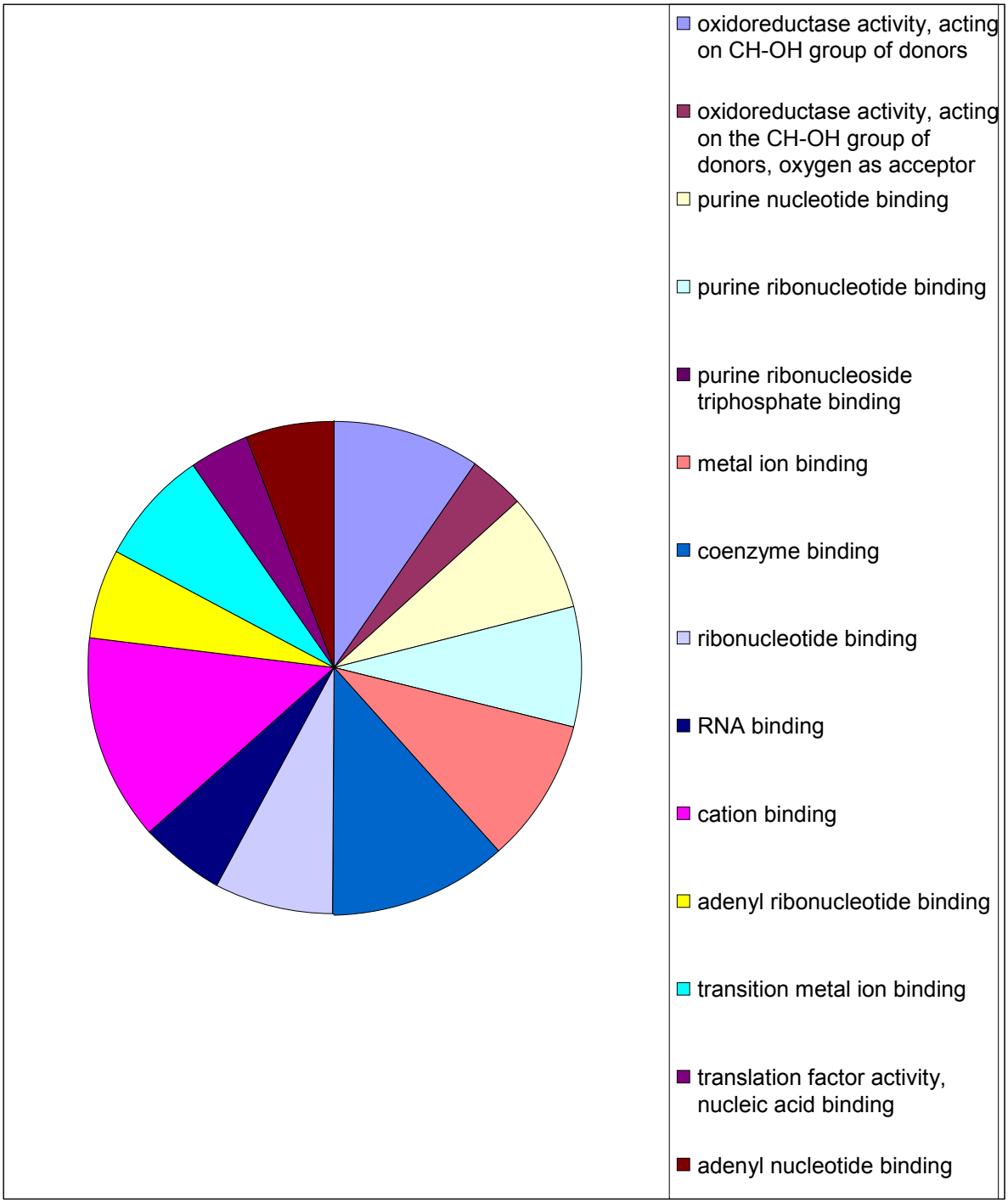

Figure 6: ***A. thaliana*** **gene ontology of molecular function, of differentially expressed proteins present in both (states: 106 mg N/L and 6 mg N/L) nitrogen treatments.**

Table1: **Differentially expressed proteins in *A. thaliana* selected by the AC Fold analysis.** Represented by blue dots in figure 1, present in both (states: 106 mg N/L and 6 mg N/L) nitrogen treatments.

#Locus	Fold Change	pValue	Description
Up regulated proteins at 6 mg/L of nitrogen			
IPI00538866	-3.62	0.006385	T22F8.160 Glycine-rich RNA-binding protein 8 (GRP8) (CCR1)
IPI00520057	-3.27	0.003067	T22F8.160 Isoform 1 of Glycine-rich RNA-binding protein 8
IPI00540225	-2.69	0.007484	T9A14.20 Peptidyl-prolyl cis-trans isomerase CYP18-3
IPI00519788	-2.75	0.001632	T20K14.130 40S ribosomal protein S19-2
IPI00534882	-2.70	0.003211	MXC20.6 Probable NADH dehydrogenase [ubiquinone], mitochondrial precursor
IPI00516334	-9	0.005591	MJB20.12 High mobility group protein gamma (HMGgamma)
IPI00542147	-2.91	0.007484	MNA5.8 Histone H3-like 5
IPI00517521	-4.16	0.00711	T22F8.160 Glycine-rich RNA-binding protein 8 (GRP8) (CCR1)
IPI00542179	-3.08	6.06E-05	T24P13.1 Eukaryotic translation initiation factor 5A-2
IPI00532033	-2.75	0.000248	F11F8.5 Histone H2B.5
IPI00542793	-4.5	0.004233	TOM9-2 Mitochondrial import receptor subunit TOM22 homolog 2
IPI00545932	-12.5	0.000607	T8E24.8 60S ribosomal protein L29-1
IPI00544229	-9	0.005591	F20C19.24 Major latex protein-related
IPI00542840	-13.5	0.000322	MTI20.13 Eukaryotic translation initiation factor 4F
IPI00531114	-4.12	0.002369	F2K15.4 Peroxidase 33 precursor
IPI00527934	-4.36	0.000146	MYF5.3 Ribosomal protein S21 family protein
IPI00538502	-8.5	0.007265	K16E14.6 Pyruvate dehydrogenase E1 component subunit beta, mitochondrial precursor
IPI00520650	-3.08	6.06E-05	F26F24.28; Polyubiquitin (UBQ10) (SEN3), senescence-associated protein

Up regulated proteins at 106 mg/L of nitrogen			
IPI00531385	4.14	0.002846	F28J12.20 (S)-2-hydroxy-acid oxidase
IPI00529487	3.4	0.005691	F9N11.40 ATPase 2, plasma membrane-type
IPI00521970	3.57	0.007951	MYN8.7 NADH-dependent glutamate synthase
IPI00516481	6.5	0.002225	F22C12.5 6-phosphogluconate dehydrogenase family protein
IPI00519564	6.75	0.001577	AT3G02360 6-phosphogluconate dehydrogenase family protein
IPI00539634	6	0.003911	F5A18.24 Glutamate:glyoxylate aminotransferase 2 (GGT2)
IPI00535689	9.5	0.004352	RPL22 Chloroplast 50S ribosomal protein L22
IPI00546372	9.5	0.004352	F2G14.30 Heavy-metal-associated domain-containing protein, Pfam profile PF00403
IPI00545955	12	0.000974	MBK23.20 6-phosphogluconate dehydrogenase family protein
IPI00524670	8.5	0.007939	T6A9.32 GTP-binding protein (TOC33)
IPI00537023	11	0.001809	APL1 Glucose-1-phosphate adenylyltransferase large subunit 1
IPI00656816	6.07	1.31E-07	MOA2.2 Similar to (S)-2-hydroxy-acid oxidase
IPI00518426	2.69	0.000307	MOA2.13 Probable peroxisomal (S)-2-hydroxy-acid oxidase 1
IPI00528534	2.72	0.000218	MOA2.2 Probable peroxisomal (S)-2-hydroxy-acid oxidase 2
IPI00523164	3.5	0.004035	F12E4.30 Adenosine kinase 2
IPI00546691	2.66	2.76E-05	F17O14.5 ADP,ATP carrier protein 1, mitochondrial precursor
IPI00532969	3.3	0.006388	MMN10.22 Probable histone H2A.7
IPI00534991	2.80	6.61E-07	F14G9.19 Phosphoglycerate kinase
IPI00527768	13	0.000522	F8L15.14 Lipoic acid synthase-like protein

Table 2: **GO of differentially expressed proteins in *A. thaliana* for N treatment of 106 mg/L and 6 mg/L; molecular component**. Numbers in parentheses represent fold change

Term ID	Protein Count	IP's	Description
oxidoreductase activity, acting on CH-OH group of donors	5	IPI00656816(6.1) IPI00518426(2.7) IPI00516481(6.5) IPI00519564(6.8) IPI00545955(12)	"Catalysis of an oxidation-reduction (redox) reaction in which a CH-OH group act as a hydrogen or electron donor and reduces a hydrogen or electron acceptor."
oxidoreductase activity, acting on the CH-OH group of donors, oxygen as acceptor	2	IPI00656816(6.1) IPI00518426(2.7)	"Catalysis of an oxidation-reduction (redox) reaction in which a CH-OH group acts as a hydrogen or electron donor and reduces an oxygen molecule."
purine nucleotide binding	4	IPI00523164(3.5) IPI00529487(3.4) IPI00524670(8.5) IPI00537023(11)	"Interacting selectively and non-covalently with purine nucleotides, any compound consisting of a purine nucleoside esterified with (ortho) phosphate."
purine ribonucleotide binding	4	IPI00523164(3.5) IPI00529487(3.4) IPI00524670(8.5) IPI00537023(11)	"Interacting selectively and non-covalently with a purine ribonucleotide, any compound consisting of a purine ribonucleoside that is esterified with (ortho) phosphate or an oligophosphate at any hydroxyl group on the ribose moiety."
purine ribonucleoside triphosphate binding	2	IPI00523164(3.5) IPI00529487(3.4)	"Interacting selectively and non-covalently with a purine ribonucleoside triphosphate, a compound consisting of a purine base linked to a ribose sugar esterified with triphosphate on the sugar."
metal ion binding	5	IPI00531114(-4.1) IPI00546691(2.7) IPI00523164(3.5)	"Interacting selectively and non-covalently with any metal ion."

		TIPI00529487(3.4) IPI00521970(3.6)	
coenzyme binding	6	IPI00656816(6.1) IPI00518426(2.7) IPI00521970(3.6) IPI00516481(6.5) IPI00519564(6.8) IPI00545955(12)	"Interacting selectively and non-covalently with a coenzyme, any of various nonprotein organic cofactors that are required, in addition to an enzyme and a substrate, for an enzymatic reaction to proceed."
ribonucleotide binding	4	IPI00523164(3.5) IPI00529487(3.4) IPI00524670(8.5) IPI00537023(11)	"Interacting selectively and non-covalently with a ribonucleotide, any compound consisting of a ribonucleoside that is esterified with (ortho) phosphate or an oligophosphate at any hydroxyl group on the ribose moiety."
RNA binding	3	IPI00542179(-3.1) IPI00542840(-13.5) IPI00535689(9.5)	"Interacting selectively and non-covalently with an RNA molecule or a portion thereof."
cation binding	7	IPI00531114(-4.1) IPI00546691(2.7) IPI00523164(3.5) IPI00529487(3.4) IPI00521970(3.6) IPI00546372(9.5) IPI00524670(8.5)	"Interacting selectively and non-covalently with cations, charged atoms or groups of atoms with a net positive charge."
adenyl ribonucleotide binding	3	IPI00523164(3.5) IPI00529487(3.4) IPI00537023(11)	"Interacting selectively and non-covalently with an adenyl ribonucleotide, any compound consisting of adenosine esterified with (ortho) phosphate or an oligophosphate at any hydroxyl group on the ribose moiety."
transition metal ion binding	4	IPI00531114(-4.1) IPI00546691(2.7) IPI00523164(3.5) IPI00521970(3.6)	"Interacting selectively and non-covalently with a transition metal ions; a transition metal is an element whose atom has an incomplete d-subshell of extranuclear electrons, or which gives rise to a cation or cations with an

			incomplete d-sub"
translation factor activity, nucleic acid binding	2	IPI00542179(-3.1) IPI00542840(-13.5)	"Functions during translation by binding nucleic acids during polypeptide synthesis at the ribosome."
adenyl nucleotide binding	3	IPI00523164(3.5) IPI00529487(3.4) IPI00537023(11)	"Interacting selectively and non-covalently with adenyl nucleotides, any compound consisting of adenosine esterified with (ortho) phosphate."

Table 3: **GO of differentially expressed proteins in *A. thaliana* for N treatment of 106 mg/L and 6 mg/L; cellular component**. Numbers in parentheses represent fold change

Term ID	Protein Count	IP's	Description
intracellular membrane-bounded organelle	24	IPI00534882(-2.7) IPI00656816(6.1) IPI00518426(2.7) IPI00542147(-2.9) IPI00542179(-3.1) IPI00532033(-2.8) IPI00542793(-4.5) IPI00542840(-13.5) IPI00531114(-4.1) IPI00546691(2.7) IPI00532969(3.3) IPI00527934(-4.4) IPI00538502(-8.5) IPI00520650(-3.1) IPI00521970(3.6) IPI00516481(6.5) IPI00519564(6.8) IPI00539634(6) IPI00535689(9.5) IPI00546372(9.5) IPI00545955(12) IPI00524670(8.5) IPI00537023(11) IPI00527768(13)	"Organized structure of distinctive morphology and function, bounded by a single or double lipid bilayer membrane and occurring within the cell. Includes the nucleus, mitochondria, plastids, vacuoles, and vesicles. Excludes the plasma membrane."

intracellular non-membrane-bounded organelle	9	IPI00519788(-2.8) IPI00656816(6.1) IPI00542147(-2.9) IPI00532033(-2.8) IPI00545932(-12.5) IPI00546691(2.7) IPI00532969(3.3) IPI00527934(-4.4) IPI00538502(-8.5)	"Organized structure of distinctive morphology and function, not bounded by a lipid bilayer membrane and occurring within the cell. Includes ribosomes, the cytoskeleton and chromosomes."
microbody	4	IPI00656816(6.1) IPI00518426(2.7) IPI00519564(6.8) IPI00539634(6)	"Cytoplasmic organelles, spherical or oval in shape, that are bounded by a single membrane and contain oxidative enzymes, especially those utilizing hydrogen peroxide (H2O2)." [ISBN:0198506732 "Oxford Dictionary of Biochemistry and Molecular Biology"]
outer membrane	2	IPI00542793(-4.5) IPI00524670(8.5)	"The external membrane of Gram-negative bacteria or certain organelles such as mitochondria and chloroplasts; freely permeable to most ions and metabolites."
cytoplasmic part	21	IPI00519788(-2.8) IPI00534882(-2.7) IPI00656816(6.1) IPI00518426(2.7) IPI00542793(-4.5) IPI00545932(-12.5) IPI00531114(-4.1) IPI00546691(2.7) IPI00527934(-4.4) IPI00538502(-8.5) IPI00529487(3.4) IPI00521970(3.6) IPI00516481(6.5) IPI00519564(6.8)	"Any constituent part of the cytoplasm, all of the contents of a cell excluding the plasma membrane and nucleus, but including other subcellular structures."

		IPI00539634(6) IPI00535689(9.5) IPI00546372(9.5) IPI00545955(12) IPI00524670(8.5) IPI00537023(11) IPI00527768(13)	
chloroplast part	8	IPI00656816(6.1) IPI00518426(2.7) IPI00546691(2.7) IPI00521970(3.6) IPI00539634(6) IPI00546372(9.5) IPI00524670(8.5) IPI00537023(11)	"Any constituent part of a chloroplast, a chlorophyll-containing plastid with thylakoids organized into grana and frets, or stroma thylakoids, and embedded in a stroma."
organelle envelope	3	IPI00546691(2.7) IPI00524670(8.5) IPI00537023(11)	"A double membrane structure enclosing an organelle, including two lipid bilayers and the region between them. In some cases, an organelle envelope may have more than two membranes."
mitochondrial part	4	IPI00534882(-2.7) IPI00542793(-4.5) IPI00546691(2.7) IPI00538502(-8.5)	"Any constituent part of a mitochondrion, a semiautonomous, self-replicating organelle that occurs in varying numbers, shapes, and sizes in the cytoplasm of virtually all eukaryotic cells. It is notably the site of tissue respiration."

intrinsic to membrane	4	IPI00542793(-4.5) IPI00546691(2.7) IPI00529487(3.4) IPI00524670(8.5)	"Located in a membrane such that some covalently attached portion of the gene product, for example part of a peptide sequence or some other covalently attached group such as a GPI anchor, spans or is embedded in one or both leaflets of the membrane."
plastid envelope	3	IPI00546691(2.7) IPI00524670(8.5) IPI00537023(11)	"The double lipid bilayer enclosing a plastid and separating its contents from the rest of the cytoplasm; includes the intermembrane space."
Plastid	11	IPI00534882(-2.7) IPI00656816(6.1) IPI00518426(2.7) IPI00546691(2.7) IPI00527934(-4.4) IPI00521970(3.6) IPI00516481(6.5) IPI00539634(6) IPI00535689(9.5) IPI00546372(9.5) IPI00545955(12)	"Any member of a family of organelles found in the cytoplasm of plants and some protists, which are membrane-bounded and contain DNA. Plant plastids develop from a common type, the proplastid."
plastid part	8	IPI00656816(6.1) IPI00518426(2.7) IPI00546691(2.7) IPI00521970(3.6) IPI00539634(6) IPI00546372(9.5) IPI00524670(8.5) IPI00537023(11)	"Any constituent part of a plastid, a member of a family of organelles found in the cytoplasm of plants and some protists, which are membrane-bounded and contain DNA. Plant plastids develop from a common type, the proplastid."

plastid stroma	6	IPI00656816(6.1) IPI00518426(2.7) IPI00521970(3.6) IPI00539634(6) IPI00546372(9.5) IPI00537023(11)	"The proteinaceous ground substance of plastids."
vacuolar part	4	IPI00542793(-4.5) IPI00546691(2.7) IPI00538502(-8.5) IPI00529487(3.4)	"Any constituent part of a vacuole, a closed structure, found only in eukaryotic cells, that is completely surrounded by unit membrane and contains liquid material."

Table 4: **GO of differentially expressed proteins in *A. thaliana* for N treatment of 106 mg/L and 6 mg/L; biological component.** Numbers in parentheses represent fold change

Term ID	Protein count	IP's	Description
defense response	4	IPI00540225(-2.7) IPI00542179(-3.1) IPI00531114(-4.1) IPI00538502(-8.5)	"Reactions, triggered in response to the presence of a foreign body or the occurrence of an injury, which result in restriction of damage to the organism attacked or prevention/recovery from the infection caused by the attack."
small molecule catabolic process	5	IPI00534991(2.8) IPI00538502(-8.5) IPI00516481(6.5) IPI00519564(6.8) IPI00545955(12) IPI00524670(8.5)	"The chemical reactions and pathways resulting in the breakdown of small molecules, any monomeric molecule of small relative molecular mass."
alcohol catabolic process	5	IPI00534991(2.8) IPI00538502(-8.5) IPI00516481(6.5) IPI00519564(6.8) IPI00545955(12)	"The chemical reactions and pathways resulting in the breakdown of alcohols, any of a class of compounds containing one or more hydroxyl groups attached to a saturated carbon atom."
alcohol metabolic process	5	IPI00534991(2.8) IPI00538502(-8.5) IPI00516481(6.5) IPI00519564(6.8) IPI00545955(12)	"The chemical reactions and pathways involving alcohols, any of a class of compounds containing one or more hydroxyl groups attached to a saturated carbon atom."
carbohydrate catabolic process	5	IPI00534991(2.8) IPI00538502(-8.5) IPI00516481(6.5) IPI00519564(6.8) IPI00545955(12)	"The chemical reactions and pathways resulting in the breakdown of carbohydrates, any of a group of organic compounds based of the general formula Cx(H2O)y."
carbohydrate metabolic process	6	IPI00534991(2.8) IPI00538502(-8.5) IPI00516481(6.5) IPI00519564(6.8) IPI00545955(12)	"The chemical reactions and pathways involving carbohydrates, any of a group of organic compounds based of the general formula Cx(H2O)y. Includes the formation of carbohydrate derivatives by the addition of a

		IPI00537023(11)	carbohydrate residue to another molecule."
small molecule biosynthetic process	4	IPI00542179(-3.1) IPI00523164(3.5) IPI00529487(3.4) IPI00521970(3.6)	"The chemical reactions and pathways resulting in the formation of small molecules, any monomeric molecule of small relative molecular mass."
cation transport	2	IPI00529487(3.4) IPI00546372(9.5)	"The directed movement of cations, atoms or small molecules with a net positive charge into, out of or within a cell, or between cells, by means of some agent such as a transporter or pore."
cellular component assembly	3	IPI00542147(-2.9) IPI00532033(-2.8) IPI00532969(3.3)	"The aggregation, arrangement and bonding together of a cellular component."
protein-DNA complex subunit organization	3	IPI00542147(-2.9) IPI00532033(-2.8) IPI00532969(3.3)	"Any process in which macromolecules aggregate, disaggregate, or are modified, resulting in the formation, disassembly, or alteration of a protein-DNA complex."
response to osmotic stress	2	IPI00516481(6.5) IPI00519564(6.8)	"A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus indicating an increase or decrease in the concentration of solutes outside the organism or cell."
macromolecular complex assembly	3	IPI00542147(-2.9) IPI00532033(-2.8) IPI00532969(3.3)	"The aggregation, arrangement and bonding together of a set of macromolecules to form a complex."
macromolecular complex subunit organization	3	IPI00542147(-2.9) IPI00532033(-2.8) IPI00532969(3.3)	"Any process in which macromolecules aggregate, disaggregate, or are modified, resulting in the formation, disassembly, or alteration of a macromolecular complex."
ion transport	2	IPI00529487(3.4) IPI00546372(9.5)	"The directed movement of charged atoms or small charged molecules into, out of or within a cell, or between cells, by means of some agent

			such as a transporter or pore."
macromolecule biosynthetic process	5	IPI00519788(-2.8) IPI00545932(-12.5) IPI00527934(-4.4) IPI00535689(9.5) IPI00537023(11)	"The chemical reactions and pathways resulting in the formation of a macromolecule, any molecule of high relative molecular mass, the structure of which essentially comprises the multiple repetitions of units derived, actually or conceptually, from molecules of low relative molecular mass."
protein metabolic process	6	IPI00540225(-2.7) IPI00519788(-2.8) IPI00542179(-3.1) IPI00545932(-12.5) IPI00527934(-4.4) IPI00535689(9.5)	"The chemical reactions and pathways involving a specific protein, rather than of proteins in general. Includes protein modification."
response to inorganic substance	5	IPI00540225(-2.7) IPI00542179(-3.1) IPI00531114(-4.1) IPI00521970(3.6) IPI00545955(12)	"A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of an inorganic substance stimulus."
response to metal ion	4	IPI00540225(-2.7) IPI00542179(-3.1) IPI00521970(3.6) IPI00545955(12)	"A change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a metal ion stimulus."

Table 5 Differentially expressed proteins in *A. thaliana* that are statistically significant but do not satisfy the AC fold test. Represented by orange dots in figure 1, present in both (states: 106 mg N/L and 6 mg N/L) nitrogen treatments

#Orange p-value cutoff: 0.005		
#Locus	pValue	Description
IPI00525776	1.00E-007	ATPB ATP synthase subunit beta
IPI00544292	1.00E-007	T5A14.11 RuBisCO large subunit-binding protein subunit beta, chloroplast precursor
IPI00530817	0.00163946	F28J12.140 Magnesium-chelatase subunit chlI, chloroplast precursor
IPI00526733	1.00E-007	T5I7.18 Isoform Long of Ribulose bisphosphate carboxylase/ oxygenase activase, chloroplast precursor
IPI00518163	1.00E-007	T5I7.18 Isoform Short of Ribulose bisphosphate carboxylase/ oxygenase activase, chloroplast precursor
IPI00520309	1.00E-007	T5I7.18 Ribulose bisphosphate carboxylase/oxygenase activase
IPI00517861	0.000453967	F8J2.100 Fructose-bisphosphate aldolase
IPI00535490	0.000521808	MBK21.15 Phosphoglycerate kinase
IPI00548616	9.49E-005	C17L7.100 Oxygen-evolving enhancer protein 3-2, chloroplast precursor
IPI00541680	0.003745087	F13A11.3 Glyceraldehyde-3-phosphate dehydrogenase B, chloroplast precursor
IPI00524965	0.001399407	T22F8.100 40S ribosomal protein S25-4

IPI00534845	0.004391822	ATPE ATP synthase epsilon chain
IPI00547451	0.000815219	F3G5.26 Histone H2B.4
IPI00535114	1.00E-007	RBCL Ribulose bisphosphate carboxylase large chain precursor
IPI00529426	0.000301266	MMN10.15 Histone H2B.11
IPI00521247	0.000495756	F16L2.190 Histone H2B.6
IPI00517188	0.003477364	T9L3.40 Carbonic anhydrase 2
IPI00545332	0.001488819	F24B9.10 Histone H2B.1
IPI00530695	4.23E-006	F14G9.19 Phosphoglycerate kinase, chloroplast precursor
IPI00526519	0.000348648	T11P11.3 Histone H2B.3
IPI00530497	0.002110907	MRN17.11 Histone H2B.10
IPI00531287	0.001391594	CPHSC70-1 Heat shock protein 70
IPI00656734	0.004921473	T9L3.40 Similar to carbonic anhydrase 1, chloroplast
IPI00537303	2.98E-005	MLJ15.3 Glyceraldehyde-3-phosphate dehydrogenase A, chloroplast precursor
IPI00539517	0.002531258	AT2G47470 Probable protein disulfide-isomerase A6 precursor
IPI00516738	0.001938886	F25A4.6 30S ribosomal protein S9, chloroplast precursor

IPI00524715	0.004131523	F5O8.29 Quinone oxidoreductase-like protein At1g23740, chloroplast precursor
IPI00527484	0.000388617	T22P11.160 Putative histone H2B.9
IPI00518961	0.00030607	K9P8.5 Heat shock protein 70
IPI00544582	2.90E-005	F13K23.15 Glyceraldehyde 3-phosphate dehydrogenase
IPI00530685	0.001685487	AT4G34555 40S ribosomal protein S25-3
IPI00535301	0.000996956	F4P12.350 Putative histone H2B.8
IPI00657073	0.003791704	MCD7.27 Similar to chaperonin
IPI00532889	0.000899304	3702 Gene_Symbol=F20H23.19 5-methyltetrahydropteroyltriglutamate—homocysteine methyltransferase
IPI00539225	0.003075287	F21E1.60 Isoform Long of Ferredoxin-dependent glutamate synthase 1, chloroplast precursor

2.5 Reference

Aebersold, R., and M. Mann. 2003. Mass spectrometry-based proteomics. *Nature*. 422:198(110).

AGI. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis *thaliana*. *Nature*. 408:796-815.

Agrawal, G.K., and R. Rakwal. 2008. Plant Proteomics: Technologies, Strategies, and Applications. In: *Wiley-Interscience series on Mass Spectrometry*. 764 pp.

Agrawal, L., S. Chakraborty, D.K. Jaiswal, S. Gupta, A. Datta, and N. Chakraborty. 2008. Comparative proteomics of tuber induction, development and maturation reveal the complexity of tuberization process in potato (*Solanum tuberosum L.*). *Journal of Proteome Research*. 7:3803-3817.

Allen, D.K., I.G.L. Libourel, and Y. Shachar-Hill. 2009. Metabolic flux analysis in plants: coping with complexity. *Plant, Cell & Environment*. 32:1241-1257.

Arney, S.E. 1952. Some effects of nitrogen nutrition on the morphology and anatomy of marrow-stem kale. *Annals of Applied Biology*. 39:266-276.

Aukerman, M., S.M. Allen, D. Loussaert, S. Luck, H. Sakai, and S.V. Tingey. 2010. Plant genes associated with improved tolerance of low nitrogen levels and their use in improving nitrogen utilization. E. I. Du Pont de Nemours and Company, USA; Pioneer Hi-Bred International, Inc. . 112pp.

Barbier-Brygoo, H., and J. Joyard. 2004. Focus on plant proteomics. *Plant Physiol Biochem*. 42:913-917.

Basu, U., J.L. Francis, R.M. Whittal, J.L. Stephens, Y. Wang, O.R. Zaiane, R. Goebel, D.G. Muench, A.G. Good, and G.J. Taylor. 2006. Extracellular proteomes of *Arabidopsis thaliana* and *Brassica napus* roots: analysis and comparison by MudPIT and LC-MS/MS. *Plant and Soil*. 286:357-376.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate - A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological*. 57:289-300.

Bennett, M.D., I.J. Leitch, H.J. Price, and J.S. Johnston. 2003. Comparisons with *Caenorhabditis* (similar to 100 Mb) and *Drosophila* (similar to 175 Mb) using flow cytometry show genome size in Arabidopsis to be similar to 157 Mb and thus similar to 25 % larger than the Arabidopsis genome initiative estimate of similar to 125 Mb. *Annals of Botany*. 91:547-557.

Bjellqvist, B., K. Ek, P. Giorgio Righetti, E. Gianazza, A. Görg, R. Westermeier, and W. Postel. 1982. Isoelectric focusing in immobilized pH gradients: Principle, methodology and some applications. *Journal of Biochemical and Biophysical Methods*. 6:317-339.

Borovskii, G.B., and V.K. Voinikov. 1993. Localization of low-molecular-weight heat shock proteins on the surface of and inside corn mitochondria. *Russian Plant Physiology (English translation)*. 40:524.

Boutros, M., and N. Perrimon. 2000. *Drosophila* genome takes flight. *Nat Cell Biol*. 2:E53-E54.

Brosch, M., S. Swamy, T. Hubbard, and J. Choudhary. 2008. Comparison of mascot and X!Tandem performance for low and high accuracy mass spectrometry and the development of an adjusted Mascot threshold. *Molecular & Cellular Proteomics*. 7:962-970.

Carpentier, S.C., B. Panis, A. Vertommen, R. Swennen, K. Sergeant, J. Renaut, K. Laukens, E. Witters, B. Samyn, and B. Devreese. 2008. Proteome analysis of non-model plants: A challenging but powerful approach. *Mass Spectrometry Reviews*. 27:354-377.

Carvalho, P.C., J.S. Fischer, E.I. Chen, J.R. Yates, and V.C. Barbosa. 2008a. PatternLab for proteomics: a tool for differential shotgun proteomics. *BMC Bioinformatics*. 9.

Carvalho, P.C., J.S.G. Fischer, E.I. Chen, G.B. Domont, M.G.C. Carvalho, W.M. Degrave, J.R. Yates, and V.C. Barbosa. 2009. GO Explorer: A gene-ontology tool to aid in the interpretation of shotgun proteomics data. *Proteome Science*. 7.

Carvalho, P.C., J.S.G. Fischer, E.I. Chen, J.R. Yates, and V.C. Barbosa. 2008b. PatternLab for proteomics: a tool for differential shotgun proteomics. *BMC Bioinformatics*. 9:316-329.

Carvalho, P.C., J. Hewel, V.C. Barbosa, and J.R. Yates. 2008c. Identifying differences in protein expression levels by spectral counting and feature selection. *Genetics and Molecular Research*. 7:342-356.

Chaffei-Haouari, C., A. Hajjaji-Nasraoui, E. Carrayol, M. Lelendais, M.H. Ghorbel, and H. Gouia. 2011. Role of glutamate availability on cadmium induced changes of nitrogen and glutamate in tomato plants (*Solanum lycopersicon*). *Acta Botanica Gallica*. 158:57-69.

Cho, K., J. Shibato, G.K. Agrawal, Y.H. Jung, A. Kubo, N.S. Jwa, S. Tamogami, K. Satoh, S. Kikuchi, T. Higashi, S. Kimura, H. Saji, Y. Tanaka, H. Iwahashi, Y. Masuo, and R. Rakwal. 2008. Integrated Transcriptomics, Proteomics, and Metabolomics Analyses To Survey Ozone Responses in the Leaves of Rice Seedling. *Journal of Proteome Research*. 7:2980-2998.

Colas, I., O. Koroleva, and P.J. Shaw. 2010. Mass spectrometry in plant proteomic analysis. *Plant Biosystems*. 144:703-714.

Dat, J., S. Vandenabeele, E. Vranová, M. Van Montagu, D. Inzé, and F. Van Breusegem. 2000. Dual action of the active oxygen species during plant stress responses. *Cellular and Molecular Life Sciences*. 57:779-795.

Debel, K., W. Sierralta, H. Braun, U. Schmitz, and K. Kloppstech. 1997. The 23-kDa light-stress-regulated heat-shock protein of *Chenopodium rubrum L*. is located in the mitochondria. *Planta*. 201:326-333.

Debnam, P.M.E., M. J. 1999. Subcellular distribution of enzymes of the oxidative pentose phosphate pathway in root and leaf tissues. *J. Exp. Bot*. 50:1653- 1661.

Denef, V.J., M.B. Shah, N.C. VerBerkmoes, R.L. Hettich, and J.F. Banfield. 2007. Implications of strain- and species-level sequence divergence for community and isolate shotgun proteomic analysis. *Journal of Proteome Research*. 6:3152-3161.

Downs, C.A., and S.A. Heckathorn. 1998. The mitochondrial small heat-shock protein protects NADH:ubiquinone oxidoreductase of the electron transport chain during heat stress in plants. *FEBS Letters*. 430:246-250.

Drexler, D., D.J. Barlow, P. Falk, J. Cantone, D. Hernandez, A. Ranasinghe, M. Sanders, B. Warrack, and F. McPhee. 2006. Development of an on-line automated sample clean-up method and liquid chromatography–tandem mass spectrometry analysis: application in an in vitro proteolytic assay. *Analytical & Bioanalytical Chemistry*. 384:1145-1154.

Eicks, M., V. Maurino, S. Knappe, U.-I. Flügge, and K. Fischer. 2002. The Plastidic Pentose Phosphate Translocator Represents a Link between the Cytosolic and the Plastidic Pentose Phosphate Pathways in Plants. *Plant Physiology*. 128:512-522.

Eng, J., A. McCormack, and J. Yates. 1994a. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*. 5:976-989.

Eng, J.K., A.L. McCormack, and J.R. Yates. 1994b. An approach to correlate tandem mass-spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*. 5:976-989.

Espagne, C., A. Martinez, B. Valot, T. Meinnel, and C. Giglione. 2007. Alternative and effective proteomic anaiysis in Arabidopsis. *Proteomics*. 7:3788-3799.

Finnie, C. 2006. Plant Proteomics. *In:Annual Plant Revews*. *Vol* 28.Blackwell Publishing. 253 pp.

Fischer, J.D.D., P.C. Carvalho, C.O. da Fonseca, L.J. Liao, W.M. Degrave, M.D.D. Carvalho, J.R. Yates, and G.B. Domont. 2011. Chemo-Resistant Protein Expression Pattern of Glioblastoma Cells (A172) to Perillyl Alcohol. *Journal of Proteome Research*. 10:153-160.

Froehlich, J.E., C.G. Wilkerson, W.K. Ray, R.S. McAndrew, K.W. Osteryoung, D.A. Gage, and B.S. Phinney. 2003. Proteomic Study of the *Arabidopsis thaliana* Chloroplastic Envelope Membrane Utilizing Alternatives to Traditional Two-Dimensional Electrophoresis. *Journal of Proteome Research*. 2:413-425.

Geisler-Lee, J., N. O'Toole, R. Ammar, N.J. Provart, A.H. Millar, and M. Geisler. 2007. A predicted interactome for Arabidopsis. *Plant Physiology*. 145:317-329.

Glinski, M., and W. Weckwerth. 2006. The role of mass spectrometry in plant systems biology. *Mass Spectrometry Reviews*. 25:173-214.

Görg, A., C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, and W. Weiss. 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*. 21:1037-1053.

Guiboileau, A., K. Yoshimoto, F. Soulay, M.P. Bataille, J.C. Avice, and C. Masclaux-Daubresse. 2012. Autophagy machinery controls nitrogen remobilization at the whole-plant level under both limiting and ample nitrate conditions in Arabidopsis. *New Phytol.* 194:732-740.

Gygi, S.P., B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, and R. Aebersold. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*. 17:994-999.

Halkier, B. 1999. Glucosinolates. *In* Naturally Occurring Glycosides: Chemistry, Distribution and Biological Properties. R. Ikan, editor. John Wiley & Sons.

Han, S., B. Yu, Y. Wang, and Y. Liu. 2011. Role of plant autophagy in stress response. *Protein & Cell*. 2:784-791.

Handrick, V., T. Vogt, and A. Frolov. 2010. Profiling of hydroxycinnamic acid amides in *Arabidopsis thaliana* pollen by tandem mass spectrometry. *Analytical and Bioanalytical Chemistry*. 398:2789-2801.

Hauschild, R., and A. von Schaewen. 2003. Differential regulation of glucose-6-phosphate dehydrogenase isoenzyme activities in potato. *Plant Physiology*. 133:47-62.

Haynes, P.A., and T.H. Roberts. 2007. Subcellular shotgun proteomics in plants: Looking beyond the usual suspects. *Proteomics*. 7:2963-2975.

Herrmann, K.M., and L.M. Weaver. 1999. The shikimate pathway *Annual Review of Plant Physiology and Plant Molecular Biology*. 50:473-503.

Hou, F.-Y., J. Huang, S.-L. Yu, and H.-S. Zhang. 2007. The 6-phosphogluconate Dehydrogenase Genes Are Responsive to Abiotic Stresses in Rice. *Journal of Integrative Plant Biology*. 49:655-663.

Huett, D.O. 1989. Effect of nitrogen on the yield and quality of vegetables. *Acta Hort. (ISHS)*. 247:205-210.

Hutchings, D., S. Rawsthorne, and M.J. Emes. 2005. Fatty acid synthesis and the oxidative pentose phosphate pathway in developing embryos of oilseed rape (*Brassica napus L.*). *Journal of Experimental Botany*. 56:577-585.

Ireland, R.J.a.L., P.J. 1999. The enzymes of glutamine, glutamate, asparagine and aspartate metabolism. Marcel Dekker, New York.

Ito, J., T.S. Batth, C.J. Petzold, A.M. Redding-Johanson, A. Mukhopadhyay, R. Verboom, E.H. Meyer, A.H. Millar, and J.L. Heazlewood. 2010. Analysis of the Arabidopsis Cytosolic Proteome Highlights Subcellular Partitioning of Central Plant Metabolism. *Journal of Proteome Research*. 10:1571-1582.

Jamet, E., C. Albenne, G. Boudart, M. Irshad, H. Canut, and R. Pont-Lezica. 2008. Recent advances in plant cell wall proteomics. *Proteomics*. 8:893-908.

Judd, W.S., C. S. Campbell, E. A. Kellogg, and P. F. Stevens. 1999. Plant Systematics: A Phylogentic Approach. Sinauer Associates, Inc. , Sunderland, MA.

Katari, M.S., V. Balija, R.K. Wilson, R.A. Martienssen, and W.R. McCombie. 2005. Comparing low coverage random shotgun sequence data from *Brassica oleracea* and *Oryza sativa* genome sequence for their ability to add to the annotation of *Arabidopsis thaliana*. *Genome Research*. 15:496-504.

Kaul, S., H.L. Koo, J. Jenkins, M. Rizzo, T. Rooney, L.J. Tallon, T. Feldblyum, W. Nierman, M.I. Benito, X.Y. Lin, C.D. Town, J.C. Venter, C.M. Fraser, S. Tabata, Y. Nakamura, T. Kaneko, S. Sato, E. Asamizu, T. Kato, H. Kotani, S. Sasamoto, J.R. Ecker, A. Theologis, N.A. Federspiel, C.J. Palm, B.I. Osborne, P. Shinn, A.B. Conway, V.S. Vysotskaia, K. Dewar, L. Conn, C.A. Lenz, C.J. Kim, N.F. Hansen, S.X. Liu, E. Buehler, H. Altafi, H. Sakano, P. Dunn, B. Lam, P.K. Pham, Q. Chao, M. Nguyen, G.X. Yu, H.M. Chen, A. Southwick, J.M. Lee, M. Miranda, M.J. Toriumi, R.W. Davis, R. Wambutt, G. Murphy, A. Dusterhoft, W. Stiekema, T. Pohl, K.D. Entian, N. Terryn, G. Volckaert, M. Salanoubat, N. Choisne, M. Rieger, W. Ansorge, M. Unseld, B. Fartmann, G. Valle, F. Artiguenave, J. Weissenbach, F. Quetier, R.K. Wilson, M. de la Bastide, M. Sekhon, E. Huang, L. Spiegel, L. Gnoj, K. Pepin, J. Murray, D. Johnson, K. Habermann, N. Dedhia, L. Parnell, R. Preston, L. Hillier, E. Chen, M. Marra, R. Martienssen, W.R. McCombie, K. Mayer, O. White, M. Bevan, K. Lemcke, T.H. Creasy, C. Bielke, B. Haas, D. Haase, R. Maiti, S. Rudd, J. Peterson, H. Schoof, D. Frishman, B. Morgenstern, et al. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 408:796-815.

Kenrick, K.G., and J. Margolis. 1970. Isoelectric focusing and gradient gel electrophoresis: A two-dimensional technique. *Analytical Biochemistry*. 33:204-207.

Kim, S.T., K.S. Cho, Y.S. Jang, and K.Y. Kang. 2001. Two-dimensional electrophoretic analysis of rice proteins by polyethylene glycol fractionation for protein arrays. *Electrophoresis*. 22:2103-2109.

Kim, Y.-O., J.S. Kim, and H. Kang. 2005. Cold-inducible zinc finger-containing glycine-rich RNA-binding protein contributes to the enhancement of freezing tolerance in *Arabidopsis thaliana. The Plant Journal*. 42:890-900.

Koch, M.A., B. Haubold, and T. Mitchell-Olds. 2000. Comparative Evolutionary Analysis of Chalcone Synthase and Alcohol Dehydrogenase Loci in Arabidopsis, Arabis, and Related Genera (*Brassicaceae*). *Molecular Biology and Evolution*. 17:1483-1498.

Krapp, A., R. Berthomé, M. Orsel, S. Mercey-Boutet, A. Yu, L. Castaings, S. Elftieh, H. Major, J.P. Renou, and F. Daniel-Vedele. 2011. Arabidopsis Roots and Shoots Show Distinct Temporal Adaptation Patterns toward Nitrogen Starvation. *Plant Physiology*. 157:1255-1282.

Kruger, N.J., and A. von Schaewen. 2003. The oxidative pentose phosphate pathway: structure and organisation. *Current Opinion in Plant Biology*. 6:236-246.

Kwon, S., and O. Park. 2008. Autophagy in plants. *Journal of Plant Biology*. 51:313-320.

Lea, P.J., and B.J. Miflin. 2003. Glutamate synthase and the synthesis of glutamate in plants. *Plant Physiology and Biochemistry*. 41:555-564.

Lee, J., W.M. Garrett, and B. Cooper. 2007. Shotgun proteomic analysis of *Arabidopsis thaliana* leaves. *Journal of Separation Science*. 30:2225-2230.

Lefsrud, M.G. 2006. Environmental manipulation to increase the nutritional content in leafy vegetables. *In* Department of Plant Sciences Vol. Ph.D. The University of Tennessee, Knoxville.

Lei, Z., A.M. Elmer, B.S. Watson, R.A. Dixon, P.J. Mendes, and L.W. Sumner. 2005. A two-dimensional electrophoresis proteomic reference map and systematic identification of 1367 proteins from a cell suspension culture of the model legume *Medicago truncatula*. *Molecular & Cellular Proteomics*. 4:1812-1825.

Lenne, C., M.A. Block, J. Garin, and R. Douce. 1995. Sequence and expression of the mRNA encoding HSP22, the mitochondrial small heat-shock protein in pea leaves. *Biochemical Journal*. 311:805-813.

Lenne, C., and R. Douce. 1994. A low molecular mass heat-shock protein is localized to higher plant mitochondria. *Plant Physiology*. 105:1255-1261.

Liska, A.J., A. Shevchenko, U. Pick, and A. Katz. 2004. Enhanced photosynthesis and redox energy production contribute to salinity tolerance in Dunaliella as revealed by homology-based proteomics. *Plant Physiology*. 136:2806-2817.

Lo, I., V.J. Denef, N.C. VerBerkmoes, M.B. Shah, D. Goltsman, G. DiBartolo, G.W. Tyson, E.E. Allen, R.J. Ram, J.C. Detter, P. Richardson, M.P. Thelen, R.L. Hettich, and J.F. Banfield. 2007. Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature*. 446:537-541.

Lukens, L., F. Zou, D. Lydiate, I. Parkin, and T. Osborn. 2003. Comparison of a *Brassica oleracea* Genetic Map With the Genome of *Arabidopsis thaliana*. *Genetics*. 164:359-372.

Lund, A.A., P.H. Blum, D. Bhattramakki, and T.E. Elthon. 1998. Heat-Stress Response of Maize Mitochondria. *Plant Physiology*. 116:1097-1110.

MacKay, V.L., X.H. Li, M.R. Flory, E. Turcott, G.L. Law, K.A. Serikawa, X.L. Xu, H. Lee, D.R. Goodlett, R. Aebersold, L.P. Zhao, and D.R. Morris. 2004. Gene expression analyzed by high-resolution state array analysis and quantitative proteomics - Response of yeast to mating pheromone. *Molecular & Cellular Proteomics*. 3:478-489.

Mahan, J.R., B.L. McMichael, and D.F. Wanjura. 1995. Methods for reducing the adverse effects of temperature stress on plants: A review. *Environmental and Experimental Botany*. 35:251-258.

Marschner, H. 1997. Mineral nutrition of higher plants. Academic Press Inc., New York. 231-254 pp.
Martienssen, R., and W.R. McCombie. 2001. The First Plant Genome. *Cell*. 105:571-574.
McDonald, W.H., R. Ohi, D.T. Miyamoto, T.J. Mitchison, and J.R. Yates. 2002. Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *International Journal of Mass Spectrometry*. 219:245-251.
Meinke, D., J. Michael Cherry, Caroline Dean, S.D.R. and, and M. Koornneef. 1998. *Arabidopsis thaliana*: A Model Plant for Genome Analysis. *Science*. 282:662-682.
Michael P. Washburn, D.W., & John R. Yates III. (2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*. 19:242-247.
Miflin, B.J., and P.J. Lea. 1976. The pathway of nitrogen assimilation in plants. *Phytochemistry*. 15:873-885.
Moresco, J.J., P.C. Carvalho, and J.R. Yates Iii. 2010. Identifying components of protein complexes in C. elegans using co-immunoprecipitation and mass spectrometry. *Journal of Proteomics*. 73:2198-2204.
Neilson, K.A., C.G. Gammulla, M. Mirzaei, N. Imin, and P.A. Haynes. 2010. Proteomic analysis of temperature stress in plants. *Proteomics*. 10:828-845.
Nesvizhskii, A.I., and R. Aebersold. 2005. Interpretation of Shotgun Proteomic Data. *Molecular & Cellular Proteomics*. 4:1419-1440.
Neuhaus, H.E., and M.J. Emes. 2000. Nonphotosynthetic metabolism in plastids *Annual Review of Plant Physiology and Plant Molecular Biology*. 51:111-140.
O'Neill, C.M., and I. Bancroft. 2000. Comparative physical mapping of segments of the genome of *Brassica oleracea var. alboglabra* that are homoeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. *The Plant Journal*. 23:233-243.
Oeljeklaus, S., H.E. Meyer, and B. Warscheid. 2009. Advancements in plant proteomics using quantitative mass spectrometry. *Journal of Proteomics*. 72:545-554.
Pal, M., L.S. Rao, V. Jain, A.C. Srivastava, R. Pandey, A. Raj, and K.P. Singh. 2005. Effects of ekevated CO2 nitrogen on wheat growth and photosynthesis. *Biologia Plantarum*. 49:467-470.
Park, O.K. 2004. Proteomic studies in plants. *Journal of Biochemistry and Molecular Biology*. 37:133-138.
Paterson, A.H., T.H. Lan, R. Amasino, T.C. Osborn, and C. Quiros. 2001. *Brassica* genomics: a complement to, and early beneficiary of, the Arabidopsis sequence. *Genome Biology [NLM - MEDLINE]*. 2:n/a-REVIEWS1011.
Patterson, S.D., and R.H. Aebersold. 2003. Proteomics: the first decade and beyond. *Nature Genetics*. 33:311-323.
Peck, S.C. 2005. Update on Proteomics in Arabidopsis. Where Do We Go from Here? *Plant Physiology*. 138:591-599.
Peng, J., J.E. Elias, C.C. Thoreen, L.J. Licklider, and S.P. Gygi. 2003. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome. *J. Proteome Res.* 2:43-50.
Perkins, D.N., D.J.C. Pappin, D.M. Creasy, and J.S. Cottrell. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 20:3551-3567.

Pidwirny, M. 2006. The Nitrogen Cycle. *In* Fundamentals of Physical Geography. Vol. *2nd Edition.* Date Viewed. March 18th 2011. http://www.physicalgeography.net/fundamentals/9s.html.

Plaxton, W.C. 1996. The organization and regulation of plant glycolysis *Annual Review of Plant Physiology and Plant Molecular Biology*. 47:185-214.

Polla, B.S., S. Kantengwa, D. François, S. Salvioli, C. Franceschi, C. Marsac, and A. Cossarizza. 1996. Mitochondria are selective targets for the protective effects of heat shock against oxidative injury. *Proceedings of the National Academy of Sciences of the United States of America*. 93:6458-6463.

Porath, J., J. Carlsson, I. Olsson, and G. Belfrage. 1975. Metal chelate affinity chromatogrphy, a new approach to protein fractionation. *Nature*. 258:598-599.

Powell, A.D., and W.H. Flurkey. 2006. Immobilized metal affinity chromatography (IMAC) of mushroom tyrosinase. *Abstracts of Papers of the American Chemical Society*. 231:718-CHED.

Quiros, C.F., F. Grellet, J. Sadowski, T. Suzuki, G. Li, and T. Wroblewski. 2001. Arabidopsis and Brassica Comparative Genomics: Sequence, Structure and Gene Content in the ABI1-Rps2-Ck1 Chromosomal Segment and Related Regions. *Genetics*. 157:1321-1330.

Ram, R.J., N.C. VerBerkmoes, M.P. Thelen, G.W. Tyson, B.J. Baker, R.C. Blake, M. Shah, R.L. Hettich, and J.F. Banfield. 2005. Community proteomics of a natural microbial biofilm. *Science*. 308:1915-1920.

Rampitsch, C., and M. Srinivasan. 2006. The application of proteomics to plant biology: a review. *Canadian Journal of Botany-Revue Canadienne De Botanique*. 84:883-892.

Reumann, S., O. Voitsekhovskaja, and C. Lillo. 2010. From signal transduction to autophagy of plant cell organelles: lessons from yeast and mammals and plant-specific features. *Protoplasma*. 247:233-256.

Richter, C., and G.E.N. Kass. 1991. Oxidative stress in mitochondria: Its relationship to cellular Ca2+ homeostasis, cell death, proliferation and differentiation. *Chemico-Biological Interactions*. 77:1-23.

Robinson, S.P., V.J. Streusand, J.M. Chatfield, and A.R. Portis. 1988. Purification and assay of ribisco activase from leaves. *Plant Physiology*. 88:1008-1014.

Rose, J.K.C.C. 2004. Tackling the plant proteome: practical approaches, hurdles and experimental tools. Vol. 39, The plant journal. 715-733.

Salvato, F., and M. de Carvalho. 2010. Methods and strategies in proteomics and their applications in plants. *Ciencia Rural*. 40:727-734.

Schnarrenberger, C., A. Flechner, and W. Martin. 1995. Enzymatic Evidence for a Complete Oxidative Pentose Phosphate Pathway in Chloroplasts and an Incomplete Pathway in the Cytosol of Spinach Leaves. *Plant Physiology*. 108:609-614.

Schulze, W.X., and B. Usadel. 2010. Quantitation in Mass-Spectrometry-Based Proteomics. *Annual Review of Plant Biology, Vol 61*. 61:491-516.

Sheoran, I.S., A.R.S. Ross, D.J.H. Olson, and V.K. Sawhney. 2007. Proteomic analysis of tomato (*Lycopersicon esculentum*) pollen. *Journal of Experimental Botany*. 58:3525-3535.

Suárez, M.F., C. Avila, F. Gallardo, F.R. Cantón, A. García-Gutiérrez, M.G. Claros, and F.M. Cánovas. 2002. Molecular and enzymatic analysis of ammonium assimilation in woody plants. *Journal of Experimental Botany*. 53:891-904.

Tabb, D.L., W.H. McDonald, and J.R. Yates. 2002. DTASelect and Contrast: Tools for Assembling and Comparing Protein Identifications from Shotgun Proteomics. *Journal of Proteome Research*. 1:21-26.

Taiz, L., and E. Zeiger. 1998. Plant Physiology. Sinauer Associates, Inc., Sunderland, Mass.

Thelen, J.J., and S.C. Peck. 2007a. Quantitative Proteomics in Plants: Choices in Abundance. *The Plant Cell Online*. 19:3339-3346.

Thelen, J.J., and S.C. Peck. 2007b. Quantitative proteomics in plants: Choices in abundance. *Plant Cell*. 19:3339-3346.

Town, C.D., F. Cheung, R. Maiti, J. Crabtree, B.J. Haas, J.R. Wortman, E.E. Hine, R. Althoff, T.S. Arbogast, L.J. Tallon, M. Vigouroux, M. Trick, and I. Bancroft. 2006. Comparative Genomics of *Brassica oleracea* and *Arabidopsis thaliana* Reveal Gene Loss, Fragmentation, and Dispersal after Polyploidy. *The Plant Cell Online*. 18:1348-1359.

Tyers, M., and M. Mann. 2003. From genomics to proteomics. *Nature*. 422:193(195).

UniProtKB/Swiss-Prot. *Date Accessed*: April 18th 2011, http://www.uniprot.org/uniprot/Q9LV03.

Vázquez-Nin, G.H., M.L. Escobar, M. De Felici, O.M. Echeverría, F.G. Klinger, and O.M. Echeverría. 2011. Autophagy Cell Death in Mammalian Ovary.Springer Netherlands., Springer Netherlands. 81-102.

Vega-Garcia, M.O., G. Lopez-Espinoza, J.C. Ontiveros, J.J. Caro-Corrales, F.D. Vargas, and J.A. Lopez-Valenzuela. 2010. Changes in Protein Expression Associated with Chilling Injury in Tomato Fruit. *J. Am. Soc. Hortic. Sci.* 135:83-89.

VerBerkmoes, N.C., V.J. Denef, R.L. Hettich, and J.F. Banfield. 2009. Systems Biology Functional analysis of natural microbial consortia using community proteomics. *Nat. Rev. Microbiol.* 7:196-205.

Vierling, E. 1991. The roles of heat shock proteins in plants. *Annual Review of Plant Physiology and Plant Molecular Biology*. 42:579-620.

Voinikov, V., T. Pobezhimova, A. Kolesnichenko, N. Varakina, and G. Borovskii. 1998. Stress protein 310 kD affects the energetic activity of plant mitochondria under hypothermia. *Journal of Thermal Biology*. 23:1-4.

von Mering, C., R. Krause, B. Snel, M. Cornell, S.G. Oliver, S. Fields, and P. Bork. 2002. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*. 417:399-403.

Wakao, S., C. Andre, and C. Benning. 2008. Functional Analyses of Cytosolic Glucose-6-Phosphate Dehydrogenases and Their Contribution to Seed Oil Accumulation in Arabidopsis. *Plant Physiology*. 146:277-288.

Walbot, V. 2000. A green chapter in the book of life. *Nature*. 408:794.

Washburn, M.P. 2004. Utilisation of proteomics datasets generated via multidimensional protein identification technology (MudPIT). *Briefings in Functional Genomics & Proteomics*. 3:280-286.

Washburn, M.P., R. Ulaszek, C. Deciu, D.M. Schieltz, and J.R. Yates. 2002. Analysis of Quantitative Proteomic Data Generated via Multidimensional Protein Identification Technology. *Analytical Chemistry*. 74:1650-1657.

Weckwerth, W., K. Wenzel, and O. Fiehn. 2004. Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics*. 4:78-83.

Wen, F.S., H.D. VanEtten, G. Tsaprailis, and M.C. Hawes. 2007. Extracellular proteins in pea root tip and border cell exudates. *Plant Physiology*. 143:773-783.

Wienkoop, S., M. Glinski, N. Tanaka, V. Tolstikov, O. Fiehn, and W. Weckwerth. 2004. Linking protein fractionation with multidimensional monolithic reversed-phase peptide chromatography/mass spectrometry enhances protein identification from complex mixtures even in the presence of abundant proteins. *Rapid Communications in Mass Spectrometry*. 18:643-650.

Wiese, S., K.A. Reidegeld, H.E. Meyer, and B. Warscheid. 2007. Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *Proteomics*. 7:340-350.

Wilmes, P., A.F. Andersson, M.G. Lefsrud, M. Wexler, M. Shah, B. Zhang, R.L. Hettich, P.L. Bond, N.C. VerBerkmoes, and J.F. Banfield. 2008. Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *ISME J.* 2:853-864.

Wilmes, P., and P.L. Bond. 2006. Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends in Microbiology*. 14:92-97.

Wittstock, U., and B.A. Halkier. 2002. Glucosinolate research in the Arabidopsis era. *Trends in Plant Science*. 7:263-270.

Wolters, D.A., M.P. Washburn, and J.R. Yates, 3rd. 2001. An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem*. 73:5683-5690.

Wroblewski, T., S. Coulibaly, J. Sadowski, and C.F. Quiros. 2000. Variation and phylogenetic utility of the *Arabidopsis thaliana* Rps2 homolog in various species of the tribe Brassiceae. *Molecular Phylogenetics and Evolution*. 16:440-448.

Wu,W.W.,G.H. Wang, S.J. Baek, and R.F. Shen. 2006. Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. *Journal of Proteome Research*. 5:651-658.

Yang,Y.W., K.N. Lai, P.Y. Tai, and W.H. Li. 1999. Rates of Nucleotide Substitution in Angiosperm Mitochondrial DNA Sequences and Dates of Divergence Between Brassica& and Other Angiosperm Lineages. *Journal of Molecular Evolution*. 48:597-604.

Zhang, Q., and D.E. Riechers. 2008. Proteomics: An emerging technology for weed science research. *Weed Science*. 56:306-313.

Chapter 3

General Conclusion

Arabidopsis one of the most closely related plants to the genus *Brassica*; therefore it is an obvious choice for evaluating comparative genomic approaches to understanding and manipulating biological processes and traits in crops.

Our experiment was to study and monitor plant growth of *A. thaliana* grown under agronomic conditions at low and high N fertilization regimes, to better understand how different proteins are up-regulates and down regulates under different environmental conditions and how this impacts the plants growth. Using comparative shotgun proteomics to characterize *A. thaliana* at the proteomic level it should be possible to develop a whole plant physiological study combined with gene, protein, and metabolite profiling to build up a comprehensive picture depicting the different levels of N uptake and assimilation. Since, Arabidopsis has long been known to contain aliphatic, aromatic and indole glucosinolates and has therefore served as a valuable model plant for the investigation of the biosynthesis of all three classes of glucosinolates (Halkier, 1999; Wittstock and Halkier, 2002). Since, glucosinolates (mustard oil glucosides) are nitrogen- and sulfur-containing natural plant products found mainly in the order Capparales, which includes agriculturally important crop plants of the Brassicaceae family such as oilseed rape (*Brassica napus*) and *Brassica* fodder and vegetables, and the model plant Arabidopsis (Halkier, 1999;

Wittstock and Halkier, 2002) used in the research as aforementioned the previous chapters in this thesis.

We identified 37 differentially expressed proteins in *A. thaliana* grown under different N levels that were associated with NADH, TCA, some secondary metabolism, and glycolysis. The proteins associated with glycolysis indicate glutamine metabolism is of major importance in plant N economy since it provides N to young developing organs. Glutamine metabolism is usually associated with the induction of the expression of enzymes such as specific proteases, cytosolic glutamine synthases isoforms and glutamine dehydrogenases. Our study indicates that under high N level treatments, proteins responsible for glutamate synthase (GOGAT), glutamine synthase (GS), and dehydrogenase activity (DH) that serve as enzymes to catalyze a link between carbohydrate and amino-acid metabolism are up-regulated. From this research, we have shown at a protein level that N could be one of the factors that may influence the levels of glucosinolates in Arabidopsis.

Future Perspective:

This research opens up various avenues in future research to explore the following: plant responses to N as being essential to elucidate the regulation of N-use efficiency, adaptive responses of plants by highlighting master traits controlling growth under each nutritional condition, and provide key target selection criteria for breeders and monitoring tools for farmers for conducting a reasoned fertilization protocol. It also aids in better understanding of a very important process of autophagy activity in plants. Autophagy plays an important role in nitrogen management at the whole-plant level through the control of nitrogen remobilization (Guiboileau et al., 2012; Han et al., 2011; Reumann et al., 2010). This role explains the importance of

autophagy in adaptation to nutrient restriction and in plant longevity (Han et al., 2011; Kwon and Park, 2008; Reumann et al., 2010; Vázquez-Nin et al., 2011). Nitrogen stress in plants has an effect globally in nitrogen use efficiency at the whole plant level and also has a strong effect on autophagy mutation on nitrogen remobilization efficiency (Aukerman et al., 2010; Guiboileau et al., 2012; Krapp et al., 2011).