Changes in the barley stem proteome in response to drought during grain filling

Mohammed Al Slamh

University of Northern Iowa

Let us know how access to this document benefits you

Copyright ©2022 Mohammed Al Slamh

Follow this and additional works at: https://scholarworks.uni.edu/etd

Part of the Agronomy and Crop Sciences Commons, and the Plant Biology Commons

Recommended Citation

Al Slamh, Mohammed, "Changes in the barley stem proteome in response to drought during grain filling" (2022). Dissertations and Theses @ UNI. 1235.
https://scholarworks.uni.edu/etd/1235

This Open Access Thesis is brought to you for free and open access by the Student Work at UNI ScholarWorks. It has been accepted for inclusion in Dissertations and Theses @ UNI by an authorized administrator of UNI ScholarWorks. For more information, please contact scholarworks@uni.edu.

Offensive Materials Statement: Materials located in UNI ScholarWorks come from a broad range of sources and time periods. Some of these materials may contain offensive stereotypes, ideas, visuals, or language.
CHANGES IN THE BARLEY STEM PROTEOME IN RESPONSE TO DROUGHT
DURING GRAIN FILLING

An Abstract of a Thesis
Submitted
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Mohammed Al Slamh
University of Northern Iowa
May 2022
ABSTRACT

Drought is the major environmental factor that limits crop yield worldwide. Crops are particularly vulnerable to drought at the reproductive stage. In cereals, grain filling during drought depends on the contribution of current photosynthesis products from the leaf and spike and carbohydrates stored mainly in the stem and remobilized to the grain. Even though stem reserve is an important source for grain filling, changes in the proteome of barley stem under drought have not been studied. Also, the molecular events associated with differential stem reserve remobilization under drought are not well understood. In this study, barley plants were exposed to gradual drought over seven days at the grain filling stage to compare changes in the stem (penultimate) proteome of two barley varieties, Samson (drought susceptible) and Giza 132 (drought tolerant). We used isobaric tandem mass tag (TMTsixplex) and liquid chromatography-tandem mass spectrometry (LC-MS) to identify proteins in non-stressed control and drought-stressed stem in the two cultivars. A total of 2,467 proteins were identified, of which 38 proteins were significantly up-regulated and 6 proteins were down-regulation in Giza 132 in response to drought. Samson did not show any changes in its stem proteome in response to drought. The proteins that were up-regulated were grouped into proteolysis and the biosynthesis of flavonoids, polyamines, and aromatic amino acids. Down-regulated proteins include those involved in carbohydrate metabolism. Our results suggest that the stems of the drought tolerant barley (Giza 132) might contribute to improving drought resistance by reorganizing its proteome through proteolysis and increasing its cellular
concentrations of flavonoids, polyamines, and aromatic amino acids while down-regulating its carbohydrate metabolism.
CHANGES IN THE BARLEY STEM PROTEOME IN RESPONSE TO DROUGHT DURING GRAIN FILLING

A Thesis
Submitted
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Mohammed Al Slamh
University of Northern Iowa
May 2022
This Study by: Mohammed Al Slamh

Entitled: CHANGES IN THE BARLEY STEM PROTEOME IN RESPONSE TO DROUGHT DURING GRAIN FILLING

has been approved as meeting the thesis requirement

for the Degree of Masters of Science

Date

______________________________

Dr. Tilahun Abebe, Chair, Thesis Committee

Date

______________________________

Dr. Michael Walter, Thesis Committee Member

Date

______________________________

Dr. Mark Sherrard, Thesis Committee Member

Date

______________________________

Dr. Jennifer Waldron, Dean, Graduate College
DEDICATION

I dedicate my thesis work to my family. A special feeling of gratitude to my loving parents whose words of encouragement and push for determination rings in my ears. I also dedicate my thesis work to my beloved Iowa host family, the Hrubys. My siblings have never left my side while I was in graduate school and are very special. I also dedicate this work to my wonderful son Salim who fills my life with love and joy. Lastly, I dedicate this dissertation to my many friends who supported me throughout my academic journey.
ACKNOWLEDGEMENTS

A special thank you to Dr. Tilahun Abebe, my thesis committee chair and mentor for his countless hours of reflection, reading, encouragement, and most of all open-mindedness throughout the entire process. I wish to thank my thesis committee members, Dr. Michael Walter and Dr. Mark Sherrard, who were more than generous with their expertise and precious time. Thank you for agreeing to serve on my committee. I am immensely grateful to Ms. Mary McDade for her invaluable support and encouragement throughout my career at the University of Northern Iowa. I would like to thank Dr. Ira Simet, Dr. Michael Walter, and Dr. Latricia Hylton for their unlimited support and encouragement throughout my time at the University of Northern Iowa. I am grateful to my family, professors, advisors, and peers who supported me throughout my journey in graduate school. I am very grateful to the Graduate College, the College of Humanities, Arts, and Sciences (CHAS), and the Department of Biology at the University of Northern Iowa for supporting my research and helping me achieve my goal during the challenging COVID-19 pandemic.
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................ vii
LIST OF FIGURES .................................................................................................... viii

CHAPTER 1: INTRODUCTION .................................................................................. 1
Importance of Grain Cereals, Global Production, and the Challenges of Feeding the World ........................................................................................................ 1
Impact of Drought on Crop Yield .............................................................................. 3
High Throughput Proteomics ................................................................................. 4
Research Significance .............................................................................................. 10
Hypotheses and Objectives ...................................................................................... 11

CHAPTER 2: METHODS ......................................................................................... 13
Plant Growth and Drought Treatment ..................................................................... 13
Experimental Design ............................................................................................... 15
Sample Collection .................................................................................................... 15
Protein Extraction ..................................................................................................... 15
Quantification of Proteins ......................................................................................... 16
SDS-PAGE Analysis of Proteins ............................................................................. 17
Protein Digestion, Desalting, and TMT Labeling .................................................... 18
Liquid Chromatography and Mass Spectrometer Analysis ...................................... 18
Identification of Peptides and Quantification of Protein Expression using MaxQuant 19
Quality Filtering, Normalization, and Differential Expression Analysis .................. 20
Functional Classification of Differentially Expressed Proteins ............................... 22

CHAPTER 3: RESULTS ............................................................................................ 23
Quality of the Protein Samples

Normalization of TMT-Labeled Data for Analysis of Differential Expression

TMT-Based Quantitative Proteomic Analysis of Barley Penultimate Internodes at the Grain Filling Stage

Differences in the Response of the Proteome between Samson and Giza 132

Functional Classification of Differentially Expressed Proteins

CHAPTER 4: DISCUSSION

Up-Regulated Proteins

Down-Regulated Proteins

REFERENCES

APPENDIX A DROUGHT TREATMENT SCHEDULE

APPENDIX B
LIST OF TABLES

Table 1: TMT labeling of barley stem samples. Proteome coverage statistics from the LC-MS/MS analysis is also shown. ............................................................................................................................................. 32

Table 2: GO terms enriched in proteins up-regulated in Giza 132 during drought at the reproductive stage. ....................................................................................................................................................... 38

Table 3: GO terms enriched in proteins down-regulated in Giza 132 during drought at the reproductive stage. ....................................................................................................................................................... 38

Table 4: Functional classification of proteins up-regulated during drought in Giza 132. ............. 39

Table 5: Functional classification of proteins down-regulated during drought in Giza 132........... 42
LIST OF FIGURES

Figure 1. Structure of the TMTsixplex reagents. ................................................................. 7
Figure 2. Summary of the procedure for multiplex isobaric labeling using TMTsixplex reagents. 9
Figure 3: Illustration of the Internal Reference Scaling normalization procedure. .................. 21
Figure 4: SDS-PAGE showing clear separation of proteins extracted from penultimate internode samples. .................................................................................................................. 23
Figure 5. Evaluation of eight normalization methods in NormalyzerDE for their ability to remove non-biological variation between technical replicates. ........................................................................... 25
Figure 6. MDS (multidimensional scaling) and correlation plots to evaluate the eight normalization methods for their ability to remove non-biological variation between technical replicates. .............................................................. 26
Figure 7. Cluster plots, density plots, and boxplots revealed the efficacy of the normalization through SL, TMM, and IRS methods. ..................................................................................... 28
Figure 8. Multi-panel scatter plots of CyclicLoess-normalized data for Giza 132 and Samson.... 29
Figure 9. Multi-panel scatter plots of IRS-normalized data for Giza 132 and Samson. ............. 30
Figure 10. Coefficient of variation distributions for raw filtered data, Cyclic Loess normalized data, and IRS normalized data. ............................................................................................................. 31
Figure 11. Venn diagram (A) and volcano plots (right) of differentially expressed proteins between the drought susceptible Samson (B) and the drought tolerant Giza 132 (C). ...................... 34
Figure 12. Protein-protein interaction networks for proteins differentially regulated during drought in the stem of Giza 132.......................................................................................................................... 37
Figure 13. Pie charts showing the percentage of differentially expressed proteins involved in biological processes, during the grain filling stage in the penultimate internode of drought-stressed Giza 132 ..................................................................................................................... 43
Figure 14. Protein biosynthesis and degradation in plants and its regulation ......................... 47
Figure 15. Schematic representation of shikimate, phenylpropanoid, and flavonoid biosynthesis pathways. .......................................................................................................................... 49
Figure 16. Reactions of the Shikimate pathway. .................................................................... 50
Figure 17. Secondary metabolites synthesized from the three aromatic amino acids.......... 52
Figure 18. Diagram of the flavonoid biosynthetic pathway in plants based.................................54
Figure 19. Polyamine biosynthetic pathway in plants ........................................................................57
Figure 20. Key enzymes in the primary carbohydrate metabolism.................................................60
Figure 21. Cellular location of thiamine-dependent enzymes.........................................................63
Figure 22. Illustration of plant intracellular communication .........................................................64
CHAPTER 1:

INTRODUCTION

Importance of Grain Cereals, Global Production, and the Challenges of Feeding the World

Grain cereals such as wheat, barley, rice, and maize provide the nutritional basis for humans and animals worldwide. Thus, these crop plants are crucial in terms of global food security. Grain crops are cultivated in almost 700 million hectares and together supply approximately 50% of the world's caloric intake (Dunwell, 2014). Among the approximately 30 genera and 360 species belonging to the tribe Triticeae (subfamily Pooideae, family Poaceae), wheat (Triticum ssp.) and barley (Hordeum vulgare L.) are economically the most important temperate cereal crops (Schnurbusch, 2019). Barley is used globally for animal feed, brewing, and is considered a principal food in regions where other crops cannot be grown. Wheat is a staple source of nutrients for around 40% of the world’s population (Giraldo et al., 2019). On a global scale, barley ranked fourth (166 million tons) in terms of world cereal crop production (Food and Agriculture Organization Statistics Division, FAOSTAT, 2020). Cereals and bread are the major source of energy for all age groups, providing 30% for adults in the developed countries and up to 80% in developing countries (Shewry & Hey, 2015).

The world population is expected to reach 10 billion by the year 2050 (Food and Agriculture Organization, FAO, 2017) and feeding this growing population is going to be a challenge. In fact, over 800 million people are already chronically hungry and 2 billion are malnourished (FAO, 2017). Driven by rapid increases in population, meat and dairy
consumption, and biofuel use, the global demand for agricultural crops is expected to double by 2050 (Godfray et al., 2010; Tilman et al., 2011). However, the total global crop production increased by only 28% in the last two decades (Grassini et al., 2013; Ray et al., 2012). Clearly, these trends must change for us to have a sustainable food supply that can meet the projected global food demands. The fact that an increasing proportion of the grain produced is being diverted to fuel production is making it difficult to best address this challenge. Corn is the main crop used for biofuel production. However, generating fuel from corn is an energy demanding process (Gallagher et al., 2016). The United States is the world’s largest corn producer. Up until 2002, less than 10% of corn harvested in the United States was used for biofuel. From 2007 to 2009 the amount of corn used for biofuels doubled, and since then the amount of corn diverted to ethanol production for biofuels exceeds 100 million metric tons annually, representing 35%–43% of the total US corn production (Singer et al., 2019). According to the Earth Policy Institute’s estimate, the amount of grain needed to fill a 25-gallon tank of a sport utility vehicle could feed one person over an entire year (Weeks, 2015, p. 459). FAO also estimated that excessive food waste accounts to nearly a third of the food produced globally (FAO, 2011). Thus, meeting future food demand will require an integrated approach that includes reducing food waste, tackling the food vs. fuel requirement, enhancing nutritional quality of crops, and improving equity and distribution (Singer et al., 2019). One essential component of an integrated strategy is the development of new cereal cultivars with improved yields under drought conditions.
Impact of Drought on Crop Yield

Exacerbated by the rapid changes in climate, drought is the most prevalent environmental factor limiting crop yield (Basu et al., 2016). Globally, it is estimated that 1,820 tons of cereal yield (maize, rice, barley, and wheat) has been lost due to drought during the past four decades (Leng & Hall, 2019).

*Hordeum vulgare* (barley) has been cultivated in all temperate regions from the Arctic Circle to the tropics making it an excellent model for ecological adaptation. It has a wide range of morphological, physiological, and biochemical adaptations to survive drought. These adaptations give barley the ability to escape, avoid, or tolerate drought. Barley can escape drought (especially terminal drought at the grain filling stage) by altering its phenology to complete its life cycle before the onset of severe shortages of water. Common drought avoidance mechanisms in barley include minimizing evaporative water loss via stomatal control, biomass allocation to the roots to extract soil moisture, and metabolic regulation to accumulate osmolytes for osmotic regulation (Hein et al., 2016). Typical drought tolerance mechanisms in barley include synthesis of proteins and compatible solutes to detoxify reactive oxygen species (ROS) and stabilize macromolecules and membranes, and mobilization of stem reserves (e.g., glucose, fructose, sucrose, and fructans) to supply carbon for grain filling (Hein et al., 2016).

Barley is diploid with a large haploid genome of 5.3 Gbp that contains 83,105 putative genetic loci including 39,734 high-confidence loci. Additionally, barley can endure a great range of environmental stresses like drought, flood, and cold or fungal infections, either single or combined (Gürel et al., 2016). Consequently, the barley
genome contains numerous stress response alleles making it an ideal model organism for the genetics and genomics of crop species.

Understanding the mechanisms of drought resistance in barley, as a cereal model organism, can provide new strategies for engineering improved drought tolerance in cereal crop species. For example, proteins that accumulate in drought-stressed barley can be linked to metabolic pathways that are targets for improvement through breeding or genetic engineering.

**High Throughput Proteomics**

Traditional biochemical methods for studying proteins, such as western blotting and ELISA (enzyme-linked immunosorbent assay) are useful to study a small set of proteins in a sample using antibody-based assays. In the past few decades, mass spectrometry (MS) has emerged as the method of choice for studying the dynamics of the proteome (a collection of proteins in a cell or tissue) in different cellular states (Schubert et al., 2017). There are two basic MS-based proteomics (large-scale study of proteins) approaches: top-down (protein-level) and bottom-up (peptide-level) proteomics (Chait, 2006; Zhang et al., 2014). In top-down proteomics, intact proteins or large protein fragments are analyzed by MS. Due to the protein size limitation of top-down proteomics (<50 kD), bottom-up proteomics is widely used. In bottom-up proteomics, a protein sample is digested with a protease (usually trypsin) and the accurate mass of the resulting peptides is determined by MS. The digestion of a protein by an enzyme provides a specific fingerprint, which identifies the protein by comparing the peptide peak lists for best matches against a list of theoretical peptide masses generated via in silico digestion.
Bottom-up proteomics performed on a mixture of proteins is referred to as shotgun proteomics because of its analogy to shotgun genome sequencing (Zhang et al., 2014).

Sample complexity is a critical factor for peptide quantitation in bottom-up proteomics because identification and quantification rates are directly proportional to sample complexity. Affinity purifications are often performed to remove highly abundant proteins and reduce sample complexity. In addition, liquid chromatography (LC) is used to fractionate peptides before the mass spectrometry (MS) step to further reduce sample complexity. The beauty of bottom-up proteomics is that it is very simple and does not depend on protein sequencing for identification. Its major limitation is that the protein must be present in the database for identification.

Bottom-up proteomics involves compromises between sensitivity and scalability (throughput; Dupree et al., 2020). Strategies to improve the sensitivity of proteomic analysis generally require large sample quantities and multidimensional fractionation, which sacrifices throughput. On the other hand, approaches to improve sensitivity and throughput of protein quantification limit the number of features (peptides) that can be monitored. For this reason, bottom-up proteomics approaches are divided into discovery proteomics (shotgun proteomics) and targeted proteomics. Discovery proteomics optimizes protein identification by spending more time and effort per sample and reducing the number of samples analyzed. In contrast, targeted proteomics analyzes a limited number of peptides with increased sensitivity across thousands of samples by...
optimizing the MS. Targeted proteomics experiments quantify less than 100 proteins with very high precision, sensitivity, specificity, and throughput.

Two strategies are widely used in shotgun proteomics to detect differences in the abundance of proteins in response to changes in the physiological state: label-free and label-based quantitation. Label-free quantitation involves. Label-free proteomics quantifies proteins by correlating protein abundance with either the number of MS/MS spectra matched to peptides and proteins or the mass spectrometric signal intensities of peptides (Li et al., 2012) between treatment conditions without any chemical modification or isotope-labeling. Label-free quantification is widely used because proteins can be identified and quantified without a laborious and costly process of introducing isotopes into samples. However, samples need to be prepared and measured separately, giving this technique limited quantification performance in terms of accuracy, precision, and reproducibility. In contrast, label-based quantitation utilizes metabolic labeling of protein samples with SILAC (stable isotope labeling with amino acids in cell culture) or isobaric labeling such as TMT (tandem mass tag) and iTRAQ (isobaric tag for relative and absolute quantitation) or isotopic labeling such as ICAT (isotope-coded affinity tags).

In SILAC, heavy isotope-labeled essential amino acids are metabolically incorporated into proteins during translation by growing cells in tissue culture media (Chen et al., 2015). This approach allows samples grown in different states to be combined at the cell level. Combining samples using this approach allows accurate measurement of relative protein abundance ratios because any bias introduced in the
downstream sample preparation and measurement would alter protein abundances from different samples to the same extent (Li et al., 2012). In isobaric labeling, protein samples are enzymatically-digested and the resulting peptides are labeled separately with different isotopic variants for relative and absolute quantification (iTRAQ or TMT reagents). The labeled samples are then combined for LC–MS/MS analysis. Isobaric tag labeling allows sample multiplexing thereby reducing instrument runtime and cost, increases analysis throughput, and eliminates variability caused by the mass spectrometer itself (Liu & Zhang, 2018).

Figure 1. Structure of the TMTsixplex reagents. A, functional regions of the reagents with MS/MS fragmentation sites by high energy collision dissociation (HCD) and electron transfer dissociation (ETD). B, TMTsixplex reagent structures and isotope positions (*). ©2016 Thermo Fisher Scientific Inc. Pub. No. MAN0011639.
Both iTRAQ and TMT tags contain three functional parts: a reporter ion group, a mass normalization group, and an amine-reactive group (Figure 1). The tags attach to the peptides of interest through the amine-reactive group which reacts with N-terminal amine groups and epsilon-amine groups of lysine residues in the peptides (Li et al., 2012). The mass normalization group is necessary because it balances the mass difference among the reporter ion groups such that different isotopic variants of the tag have the same mass. Reporter ions of different masses are then dissociated from isolated peptides and the mass of reporter ions (each variant of the tag has a specific reporter ion mass) is used to measure the relative abundance of the labeled peptide.

TMT-labeling of peptides allows accurate comparison of the relative abundances of proteins under stress and disease conditions in plants (Erdjument-Bromage et al., 2018). Currently, up to 16 samples can be multiplexed in TMT-labeling (Figure 2). Furthermore, the co-isolation and fragmentation of the isobaric precursor ions in TMT multiplexing greatly reduces the number of missing peptide quantification values and the technical variation in the experimental workflow. Therefore, TMT-labeling provides a deep proteome coverage, in a reasonable amount of measurement time, for statistical analysis and in turn a better statistical power to identify the differentially expressed proteins (Liu & Zhang, 2018). Moreover, the TMT labeling approach is more robust and provides higher quantification reproducibility than label-free measurements. These advantages made isobaric labeling a gold standard for mass spectrometry-based proteome analysis. The only disadvantage of the TMT labeling method is its higher cost.
The workflow of the TMT experiment involves multiple stages going from protein extraction to differential analysis and visualization of protein relative abundances from complex peptide samples. After preliminary quantification, protein digestion, TMT labeling, and pooling of peptide samples, nano-liquid chromatography coupled to tandem MS (nano-LC-MS/MS) analysis is performed using a high-resolution mass spectrometer. In the experiments described in our research, a Thermo Scientific Q Exactive Hybrid Quadrupole-Orbital Mass Spectrometer was used for all TMT peptide quantitation experiments. The resulting fragmentation spectra were analyzed using the MaxQuant proteomics software (Tyanova et al., 2016; Sinitcyn et al., 2018) that is freely available at www.maxquant.org. The Andromeda peptide search engine (Cox et al., 2011), integrated into the Maxquant environment, is used to identify peptide spectra and roll them into protein identifications. Final analysis of TMT reporter ion intensities can be easily achieved by importing MaxQuant-analyzed data into R software that has a large variety of complementary packages capable of performing a wide range of statistical analyses (Erdjument-Bromage et al., 2018).
Research Significance

The cereal inflorescence (spike) contains many spikelets (florets) that are surrounded by a husk (composed of lemma and palea). The husk protects the developing seed from pathogens and insects. In addition, the husk is photosynthetically active and supplies the developing seed with assimilates (Abebe et al., 2004). Proteome analysis of lemma, palea, awn, and leaf revealed accumulation of proteins involved in photosynthesis, defense mechanisms, carbon, nitrogen, and secondary metabolism during drought (Walck, 2014). Of these, rubisco activase in the chloroplast, luminal binding protein in endoplasmic reticulum, phosphoglycerate mutase, glutathione S-transferase, heat shock proteins and enzymes in the phenylpropanoid biosynthesis are sensitive to drought (Rodziewicz et al., 2019).

Grain filling in cereals during drought depends on the contribution of current assimilates from the leaf and spike as well as carbohydrates stored in the stem and are remobilized to the grain. In wheat, stem reserves are remobilized to the grain during drought (Ehdaie et al., 2006). Using 2D polyacrylamide gel electrophoresis, Bazargani et al. (2011) identified a total of 830 proteins in the stem, 135 of which showed significant changes under drought. Photosynthetic proteins (rubisco large and small subunit, rubisco activase, and oxygen-evolving complex) were down-regulated while defense and metabolism proteins (SAM synthase 1, abscisic acid and stress inducible protein, glutathione transferase, and ROS scavengers) were up-regulated (Bazargani et al., 2011). Also, the highest number of differentially expressed proteins was observed during the grain filling stage, suggesting that those proteins play a role in stem reserve
remobilization for grain development (Bazargani et al., 2011). Additionally, the activity of sucrose synthase was found to be higher in the grain of stressed wheat, which alters the grain sink activity by increasing assimilate unloading, therefore, increasing the rate of remobilization of carbohydrates from the stem to the grain (Gupta et al., 2011). Although stem reserve is an important source for grain filling under drought conditions, changes in the proteome of barley stem under drought have not been studied and the molecular events associated with differential stem reserve remobilization are not well understood (Bazargani et al., 2011). My research provides information about the differential expression of proteins in barley stem in response to drought. The study also helps us understand the molecular mechanisms underlying stem reserve remobilization for grain filling in barley during drought. The findings are informative for developing new strategies to improve the drought tolerance of barley and related cereal crops through breeding and genetic engineering.

**Hypotheses and Objectives**

I compared changes in the stem proteome of two barley varieties, Samson (drought susceptible) and Giza 132 (drought tolerant), during drought at the grain filling stage. I had two hypotheses: (1) drought could significantly alter the proteome of the stem in both the drought susceptible (Samson) and drought tolerant (Giza 132) barley varieties, and (2) proteins involved in drought tolerance and stem reserve mobilization would be highly expressed in the drought tolerant Giza 132. I expected this would increase stem metabolites available for grain development. The findings of my research were compared with changes in the proteome of wheat stem (Ehdaie et al., 2006;
Bazargani et al., 2011) and the spike organs of barley (Walck, 2014). I had two objectives to test these Hypotheses:

1. Compare changes in the stem proteome of samson (drought susceptible) and Giza 132 (drought tolerant) in response to drought. I compared differential expression of proteins between well-watered controls and drought-stressed plants for each cultivar. This would allow determination of how the stem proteomes of the susceptible Samson and the tolerant Giza 132 respond to drought.

2. Compare differences in the type of proteins that accumulate in the stems of Samson and Giza 132 during drought. This objective would help us identify specific proteins involved in drought tolerance and reserve mobilization in the two varieties.
CHAPTER 2: METHODS

Plant Growth and Drought Treatment

We used two six-row barley (*Hordeum vulgare*) cultivars in this study, Samson and Giza 132. The seeds were obtained from the U.S. Department of Agriculture National Small Grains Collection (NSGC). Samson is a hulled barley with a pedigree Olli/M64-69//R72-181 developed by Helm et al. (1986). Samson is susceptible to drought but resistant to lodging. Giza 132 is an Egyptian hulled barley, which is tolerant to drought and fungal diseases. It was released in 2006 by the Barley Research Department, Agricultural Research Center, Giza, Egypt (Noaman et al., 2007). Giza 132 originated from the cross Rihane-05//As 46/Aths*2 Aths/Lignee 686.

We performed a preliminary study to assess the development rates of the two cultivars under controlled conditions. We determined that Giza 132 matures 10-14 days earlier than Samson. Accordingly, we planted Giza 132 two weeks later than Samson in order to apply drought and harvest plants at comparable developmental stages.

Four days before planting, we saturated soil (containing 17% topsoil, 50% Canadian peat moss, 25% vermiculite, and 8% rice hulls) with tap water in a large plastic tub. On the third day, we crushed and mixed the soil for uniformity. The following day, we filled 2-gallon pots (7.5 L capacity, 20 cm top diameter × 16 cm bottom diameter × 20 cm deep) with the saturated soil. We placed the pots on saucers, covered with cling wrap, and allowed excess water to drip overnight. At this time the soil is at field capacity. On the fourth day, we adjusted the total weight of the saturated soil to 4.8 kg between 0800
and 0900 h (8:00 – 9:00 a.m.). Then, we planted eight seeds (1.5 cm deep) evenly in a circle about 3 cm from the edge. We supplemented each pot with 5 g Osmocote® slow release fertilizer (NPK, 19-6-12; Scotts Company LLC, Marysville, OH). More fertilizer was applied on the 7th, 21st, and 35th day (for Samson) or 7th, 21st, and 28th day (for Giza 132) after planting by adding 100 ml of 4 g/liter Jack’s Professional Petunia FeED Plus with magnesium (NPK, 20-19-20) and 5 g Osmocote® as shown in Appendix A.

Plants were kept in a controlled environment growth chamber (PGR14, Conviron, Winnipeg, Canada), which was connected to a Thermoflex 10,000 chiller (Thermo Fisher Scientific, Waltham, MA) for cooling. The chamber was maintained at 22°C day/18°C night temperature, 60% relative humidity, and 16 h light period. We programmed the light intensity to increase in half-hour intervals in the morning from 0 (dark) to 219, 437, 656, and 875 μmoles m² sec⁻¹ at mature plant height. At the end of the day, light intensity was reduced in the reverse order to dark at 30 min interval.

Two weeks after planting, seedlings were thinned to six uniform plants per pot. We removed tillers at two to three-day intervals beginning from the appearance of the third tiller (Zadock stage 12) and the flag leaf (Zadock stage 38). Samson was watered to 4000 g on the 10th, 14th, 18th, 22nd, 26th, 30th, 33rd day after planting and then to 4300 g every three days between the 36th and 48th day after planting. Giza 132 was watered to 4000 g on the 10th, 14th, 18th, and 22nd day after planting and then to 4300 g every three days between the 25th and 34th day after planting.

When the tip of the awn emerged (day 43 in Samson and day 29 in Giza 132), we randomly assigned pots to either the “control” group or the “drought” group. Pots were
randomized every day during the remainder of the growing period. Drought treatment started 50 and 36 days after planting Samson and Giza 132, respectively. Control pots were watered to 4300 g total weight every day while plants in the stressed group were exposed to drought by slowly withholding water to 50% over a period of seven days as follows. Day 0: 4300 g (100%), day 1: 3870 g (90%), day 2: 3440 g (80%), day 3: 3010 g (70%), day 4: 2795 g (65%), day 5: 2580 g (60%), day 6: 2365 g (55%), and day 7: 2150 g (50%).

Experimental Design

We used a completely randomized design (CRD) with three replicates for the drought experiment. We had two factors: cultivar (Samson and Giza 132) and treatment conditions (control and drought). Each treatment was replicated three times.

Sample Collection

At the end of the stress period, we collected the penultimate internode for protein extraction. About 8 cm stems were collected, 2 cm from the bottom node, in separate, 15 ml falcon tubes and immediately frozen in liquid nitrogen. The tissue was stored at -80°C until needed for protein extraction. For consistency of results, all sample collections were performed between 1300 and 1500 (1:00 p.m. and 3:00 p.m.) CST.

Protein Extraction

For total protein extraction, we used four penultimate internodes (~1.5 g) from each control and drought-stressed plants. We ground the tissue to a fine powder in liquid nitrogen with a mortar and pestle. The frozen powder was transferred to 50 ml polypropylene tubes containing 5 ml of ice-cold extraction buffer (5 M Tris-HCl, pH 8.0,
1 M Na2-EDTA, pH 8.0, Sucrose, 2.5 M KCl, β-mercaptoethanol, 200 mM PMSF, Protease cocktail (100×, Sigma)). Samples were shaken briefly to mix and homogenized with a handheld Polytron homogenizer for 30 seconds on ice. Five milliliters of cold Tris-saturated phenol (buffered with 0.1 M Tris-HCl, pH 8.0) was added and the tubes were shaken end-over-end on a Roto-Rack (Fisher) mixer for 30 min in a cold room. The mixture was centrifuged at 10,000×g for 10 min at 4°C. The upper phenol phase was transferred to a fresh 50 ml polypropylene tube. The interface and the bottom aqueous phase were re-extracted following the same procedure and the phenol phase was transferred to the tube containing the previous phenol phase. An equal volume of extraction buffer was added to the pooled phenol phase. Samples were mixed and centrifuged at 10,000×g for 10 min at 4°C.

The upper phenol phase was transferred to a fresh 50 ml polypropylene tube. Five volumes of 100 mM ammonium acetate in pre-chilled methanol was added to the phenol extract. Samples were mixed and incubated in a -20°C freezer overnight to precipitate proteins. Proteins were pelleted by centrifugation at 10,000×g for 10 min at 4°C. The supernatant was decanted and the protein pellet was washed twice with 3 ml of ice–cold 100 mM ammonium acetate and 10 mM DTT in methanol. Finally, the protein pellet was washed twice with 3 ml ice-cold 10 mM DTT in 80% acetone, dried in a SpeedVac, and stored at -20°C.

**Quantification of Proteins**

The protein pellet was solubilized in 300 µl of 6.5 M urea and 50 mM Tris-HCl (pH 8.0). The samples were sonicated at 10-18°C to facilitate solubilization. The protein
solution was then centrifuged at 12,000 ×g for 5 min at 4°C to remove insoluble protein. The supernatant was transferred to a fresh microcentrifuge tube. We determined the protein concentration of samples using the 2-D Quant kit (GE Healthcare Life Sciences, Pittsburgh, PA). Bovine serum albumin (BSA) was used as a standard.

**SDS-PAGE Analysis of Proteins**

We checked the quality of the protein samples using Tris-glycine sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Protein samples (20 µg of 1 µg/µl protein) were mixed with 4× urea-SDS loading buffer (containing 8% SDS, 240 mM Tris-HCl, pH 6.8, 8M urea, 40% glycerol, and a few grains of bromophenol blue) in 20-30 µl. The samples were loaded onto the gel consisting of a stacking layer of 5% acrylamide and a separating gel of 10% acrylamide. Electrophoresis was performed at 150 - 200 volts using Tris-glycine running buffer (25 mm Tris, 192 mM glycine, and 0.1% SDS, pH 8.3). Kaleidoscope protein ladder (10-250 kD, Bio-Rad, Hercules, CA) was used as a molecular weight standard.

The gel was fixed in 10 volumes of a mixture of 10% acetic acid and 25% isopropanol in a glass tray for 30 min. The fixing solution was carefully decanted from the tray and 5-10 gel volumes of staining solution (0.006% Coomassie Brilliant Blue G-250 in 10% acetic acid) were added. The tray was covered with Saran Wrap to reduce evaporation during staining. The gel was soaked in the staining solution overnight with gentle agitation. Next day, the staining solution was removed and the gel was destained in 10 volumes of 10% acetic acid until the background was removed. We scanned the gel
using Gel Logic 112 imaging system with software version 5.0 (Carestream, Rochester, NY).

**Protein Digestion, Desalting, and TMT Labeling**

A pooled control sample consisting of an equal amount (10 µg) of all samples in the study was prepared. The pooled sample was used as a reference channel in the TMT labeling experiment described below. All individual samples were reduced with DTT, modified with iodoacetamide and digested with trypsin/Lys-C (cat. no V5073, Promega, Madison, WI) at a 25:1 protein:protease ratio (w/w), at 37°C overnight. Formic acid was added to stop the digestion. Samples were desalted using C18 MicroSpin Columns (Part # SEM SS18V, Nest Group, Inc., Ipswich, MA) and dried in a SpeedVac. The digested peptides were dissolved in 100 µl of 100 mM TEAB and the concentration was determined using the Pierce Colorimetric kit (cat. no. 23275, ThermoFisher Scientific, Waltham, MA).

Each digested sample (25 µl) was labeled with 10 µl of TMTsixplex reagents (cat. no. 90061, ThermoFisher Scientific, Waltham, MA), with the pooled control being labeled with TMT6-126. Equal amounts of each TMT-labeled sample (5 µg) and the control (5 µg) were pooled into a single tube and dried in a SpeedVac. The dried samples were reconstituted in 5% acetonitrile/water/0.1% formic acid.

**Liquid Chromatography and Mass Spectrometer Analysis**

Peptides (5 µg injections) were separated by liquid chromatography (EASY nLC-1200 coupled to a Nanospray FlexIon source, ThermoFisher Scientific, Waltham, MA) through a pulled glass 75 µm × 20 cm capillary emitter (part #16-2644-5, Agilent, Santa
Clara, CA). The tip of the emitter was packed with Zorbax SB-C18 C18 5-micron chromatography packing material (part #8220966-922, Agilent, Santa Clara, CA) while the remainder of the column was packed with UChrom C18 3-micron packing material (part #80002, nanoLCMS Solutions). The peptides were fragmented for analysis by MS/MS at the Iowa State University Proteome Facility using the Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer with an HCD fragmentation cell (ThermoFisher Scientific, Waltham, MA).

Identification of Peptides and Quantification of Protein Expression using MaxQuant

The LC-MS raw files were analyzed using MaxQuant (Tyanova et al., 2016; Sinitcyn et al., 2018). Peptide m/z patterns were matched against theoretical fragments of protein sequences in the Morex barley genome sequence assembly (Monat et al., 2019) using the Andromeda peptide search engine (Cox et al., 2011) that integrated into the Maxquant environment. The Morex barley reference protein sequences were downloaded from e!DAL - Plant Genomics & Phenomics Research Data Repository (Arend et al., 2016; http://edal-pgp.ipk-gatersleben.de/) under doi:10.5447/ipk/2019/8. Isotope correction factors, provided by the manufacturer (ThermoFisher Scientific, Waltham, MA) were applied to the TMT-labeled data during MaxQuant analysis. Peptide-level normalization (weighted median ratio) with and without the reference channel was applied in MaxQuant to test if it removes batch effect. Protein identification was performed in each sample group. TMT tags oxidation of methionine (M), N-terminal protein acetylation (acetyl Protein N-term), and deamidation of asparagine and glycine (NQ) were selected as variable modifications. TMT tags on free N-termini and lysine
residues and carboxymethylation of cysteine residues were selected as fixed modifications. Unique and razor peptides were selected for quantification.

**Quality Filtering, Normalization, and Differential Expression Analysis**

MaxQuant output ProteinGroups.txt file containing intensity data from all three TMT runs was imported into RStudio. Reversed sequences, potential contaminants, and proteins identified only by sites (all indicated with a “+”), and proteins that are detected less than 2 times across all 18 samples were all removed (R code, Appendix B).

Cyclic-Loess normalization was applied to the filtered data (total of 2,467 identified major proteins) using the R package NormalizerDE (Willforss et al., 2018). The impute_LS_adaptive() function (LSA) in the missMethods package (Tobias, 2022) was used for imputation of missing values.

Since we labeled the same amount of protein in each sample, we expected the total signal in each channel to be the same. However, we found that the total signal in each channel was different. To correct for differences in sample loading (SL) and labeling efficiency, we multiplied the grand total reporter ion intensity for each channel within each TMT experiment (runs 1-3) by global scaling factors. This adjusted the total intensity to the average total intensity across the six channels within each TMT experiment.

Trimmed mean of M values (TMM) normalization was applied among technical replicates (represented by separate Mass Spectrometry runs 1-3) of both the raw and the sample-loading normalized data to remove any compositional bias that may arise from the presence of highly abundant proteins which can block signals of low abundance.
proteins. The goal of applying this method was to make the centers of protein abundance distributions (medians) more similar.

Finally, internal reference scaling normalization (IRS) was applied to the TMM-normalized data to remove variations among technical replicates (Plubell et al., 2017) (Figure 3).

![Internal Reference Scaling (IRS) normalization](image)

**Figure 3**: Illustration of the Internal Reference Scaling normalization procedure (Wilmarth, 2018).

Coefficient of variation (CV) distributions of the different normalized data (raw filtered, cyclic loess, and IRS) were compared. The CVs are computed per protein, so their distribution gives us more insight into how the abundance of each protein is behaving within the same cultivar and treatment. We expect proteins from the same cultivar and treatment to have similar abundance across replicates.

Differentially expressed proteins between groups were determined by comparing the LSA-imputed, CycLoess-SL-IRS-TMM-normalized total reporter ion intensities
between groups using the Bioconductor package *limma* (3.46.0) (Ritchie et al., 2015). A minimum fold change of 1.25 (0.323 for log2) and an FDR of 5% were applied.

**Functional Classification of Differentially Expressed Proteins**

To ascertain the functional classes and interaction networks of differentially expressed proteins in Samson vs Giza132 during drought stress, we used the BarleyNet functional omics analysis server (Lee et al., 2020). In protein interaction networks, the dot (node) represents a protein and the link (line) represents the association.
CHAPTER 3:  
RESULTS  

Quality of the Protein Samples

We performed SDS-PAGE analysis to check the quality of our protein samples. Proteins separated nicely and consistently across lanes (Figure 4). There were no noticeable differences in band intensity, indicating no difference in the amount of protein loaded. This also confirms the spectrophotometric measurements of protein concentration using the 2D-quant kit. The protein samples were of high quality and ready for mass spectrometry analysis.

![SDS-PAGE Image](image)

Figure 4: SDS-PAGE showing clear separation of proteins extracted from penultimate internode samples.

Normalization of TMT-Labeled Data for Analysis of Differential Expression

The normalization method used in large proteomics data (as in our experiment where we analyzed multiple replicates in three separate TMT runs) is very critical because of its influence on the differential expression analysis. Therefore, we evaluated
and compared the performance of eight different normalization methods based on their ability to remove non-biological variations between technical replicates introduced during the MS analysis. We loaded the same amount of protein in each channel so any differences in the column total intensities are caused by non-biological variation. Normalization makes the totals similar as possible.

These variations are non-biological because we loaded an equal amount of protein from each sample into the mass spectrometer channels. As shown in Figures 5 and 6, the medians of the distributions and the centers of the smoothed density distributions of the log2-transformed data are not aligned indicating the presence of significant, but biologically irrelevant, variability. We can remove this variability by applying normalization. All of the normalization methods evaluated outperformed the log2-transformed data (Figures 5 and 6). All the methods, except log2 transformation, were equally effective in their abilities to remove bias and decrease variability between technical replicates (Figure 5). However, all of the normalization methods showed an extreme clustering by TMT run or technical replicates as the MDS plots shows (Figure 6). We tried quantile normalization, but it introduced negative values in our data making it incompatible with differential expression analysis. We chose CyclicLoess normalization because it produced the lowest variability between replicates (Figures 5 and 6).
Figure 5. Evaluation of eight normalization methods in NormalyzerDE for their ability to remove non-biological variation between technical replicates. The normalization methods used: CycLoess (cyclic Loess normalization), GI (global intensity), log2, mean, median, quantile, RLR (robust linear regression), and VSN (Variance stabilization normalization). (A) Boxplots showing the distribution of raw log2 data and normalized data. (B) The effect of normalization methods on intragroup variation between biological replicates was assessed using PCV (pooled coefficient of variation), PEV (pooled estimate of variance), and PMAD (mean of intragroup median absolute deviation).
Figure 6. MDS (multidimensional scaling) and correlation plots to evaluate the eight normalization methods for their ability to remove non-biological variation between technical replicates. (A) MDS plots and (B) correlation plots.

Since an equal amount of protein was loaded into each channel, we should have similar total protein intensity across all channels. On average, 14% of data had missing values which is relatively small for MS, but we might lose important information by excluding missing values. Therefore, it is better to use data imputation (a technique used
for replacing missing data with estimated values based on available information in a dataset) to predict the missing values before protein differential analysis.

Cyclic Loess normalization aligned the distribution centers of both the boxplot and density plot (Figure 7). However, there is a strong clustering by TMT experiment that can be seen from the MDS plot of Cyclic Loess normalized data (Figure 7). The bottom right boxplot shows the distributions of protein intensities after LSA imputation which are similar to the Cyclic Normalized data (Figure 7).

The density plot looks a little better after LSA imputation. Both the boxplot and density plot after sample loading normalization do not look different from the Cyclic Loess normalized plot (Figure 6). Trimmed Mean of M-values (TMM) normalization further normalized the data making the medians perfectly aligned and decreasing the width of the distribution boxes (Figure 7).

Despite applying two sensible normalizations that greatly improved the boxplot and density distributions, there is still an unexpected strong clustering by TMT experiment (Figure 7). Therefore, we needed to apply another normalization step (Internal Reference Scaling) to remove TMT experiment clustering. Boxplot and density plot of IRS normalized data show further improvement in the alignment of distribution centers (Figure 7). The cluster plot shows that the samples were grouped by biological relevance (cultivar and treatment) rather than some arbitrary external factor (the TMT experiment). IRS succeeded in removing the clustering effect because it uses a common reference (the pooled samples), that is constant in all of our three TMT runs, to scale all samples to a common average value (Figure 7).
Figure 7. Cluster plots, density plots, and boxplots revealed the efficacy of the normalization through SL, TMM, and IRS methods.
In addition to investigating the reproducibility of technical replicates, cross-correlation examinations were also performed where technical replicates of the same samples showed outstanding consistency with an average $R^2$ value of 0.997 (Figures 8 and 9). The scatter between replicates is dramatically reduced in all cases after the IRS corrections.

Figure 8. Multi-panel scatter plots of CyclicLoess-normalized data for Giza 132 and Samson.
Figure 9. Multi-panel scatter plots of IRS-normalized data for Giza 132 and Samson.
The multiple normalizations applied showed improvement of the coefficient of variation (CV) values of each channel reporter ion intensities from 38.6% to 0.55% in raw data and IRS and TMM normalized data, respectively (Figure 10). This step is critical because CV distributions are computed per protein which gives us more insight into how individual proteins are behaving.

Figure 10. Coefficient of variation distributions for raw filtered data, Cyclic Loess normalized data, and IRS normalized data.
TMT-Based Quantitative Proteomic Analysis of Barley Penultimate Internodes at the Grain Filling Stage

For TMT-based proteome analysis, extracted proteins from 12 samples in three replicates were subjected to an in-solution trypsin\Lys-C digestion based on the manufacturer’s protocol (Promega, V5073). Digested peptides from 3 biological replicates of 12 samples were labeled with TMTsixplex kit (cat # 90061, ThermoFisher Scientific, Waltham, MA) as shown in Table 1.

Table 1: TMT labeling of barley stem samples. Proteome coverage statistics from the LC-MS/MS analysis is also shown.

<table>
<thead>
<tr>
<th>Replicate/ Run</th>
<th>Sample</th>
<th>Group</th>
<th>TMT tag</th>
<th>Number of Identified Peptides</th>
<th>Number of Identified Unique Peptides</th>
<th>Sequence Coverage %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Samson Control</td>
<td>TMT6-126</td>
<td>14,293</td>
<td>10,963</td>
<td>19.3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Samson Drought</td>
<td>TMT6-127</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Giza132 Control</td>
<td>TMT6-128</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>Giza132 Drought</td>
<td>TMT6-129</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pooled1</td>
<td>Reference</td>
<td>TMT6-130</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pooled2</td>
<td>Reference</td>
<td>TMT6-131</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Samson Control</td>
<td>TMT6-126</td>
<td>14,606</td>
<td>11,197</td>
<td>19.5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Samson Drought</td>
<td>TMT6-127</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Giza132 Control</td>
<td>TMT6-128</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>Giza132 Drought</td>
<td>TMT6-129</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pooled1</td>
<td>Reference</td>
<td>TMT6-130</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pooled2</td>
<td>Reference</td>
<td>TMT6-131</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Samson Control</td>
<td>TMT6-126</td>
<td>14,504</td>
<td>11,140</td>
<td>19.6</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Samson Drought</td>
<td>TMT6-127</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Giza132 Control</td>
<td>TMT6-128</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>Giza132 Drought</td>
<td>TMT6-129</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pooled1</td>
<td>Reference</td>
<td>TMT6-130</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pooled2</td>
<td>Reference</td>
<td>TMT6-131</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The LC-MS/MS analysis led to the identification of 15,967 peptides and 12,423 unique peptides, matching to 2,664 protein groups. The average score for these identified proteins was 39.5 with an average sequence coverage of 21.3% and a unique sequence
coverage of 17.3%. Out of the 2,664 protein groups, 1,839 showed more than one unique peptide. The number of identified peptides (14,293, 14,606, 14,504, respectively), number of identified unique peptides (10,963, 11,197, 11,140, respectively), and average sequence coverage (19.3, 19.5, 19.6, respectively) showed great reproducibility across the three replicates (Table 1).

Differences in the Response of the Proteome between Samson and Giza 132

To reduce the impact of possible false positive identifications, we only retained the 2,467 proteins with 2 or more peptides quantified for the downstream analysis (Supplemental Table). Using the Bioconductor package limma (3.46.0) (Ritchie et al., 2015). We used a student’s t-test with a false discovery rate (FDR, Benjamini–Hochberg, 1995) of 0.05 (to control familywise error rate during multiple tests), and a fold change of more than 1.25 was applied to identify statistically significant proteins. This approach identified 44 differentially expressed proteins in Giza 132 during drought at the grain filling stage (Figure 11). Of these 44 proteins, 38 were up-regulated while 6 were down-regulated. No significant proteins were differentially expressed in the drought susceptible Samson (Figure 11).
Figure 11. Venn diagram (A) and volcano plots (right) of differentially expressed proteins between the drought susceptible Samson (B) and the drought tolerant Giza 132 (C).
Functional Classification of Differentially Expressed Proteins

Omics studies (genomics, transcriptomics, proteomics, etc.) generate large data. Analysis of this data often produces a long list of genes and proteins, which poses challenges to interpretation. A standard approach is to reduce the long list into smaller, more familiar concepts that are easy to interpret. This often involves pathway and network analysis. In addition to reducing data complexity, pathway and network analysis identifies the possible mechanisms of certain processes (such as disease, signal transduction, transport), predicts new roles for genes/proteins, improves statistical power because fewer tests are performed by combining data from multiple genes/proteins into one pathway or network, and facilitates integration of multiple data types (Creixell et al., 2015).

Proteins rarely carry out their function individually. Rather, they work together in a coordinated manner to accomplish specific functions. This interaction can be represented as a network, a simplified abstraction of complex biological processes as sets of molecular interactions or functional relationships. In protein-protein interaction networks, nodes denote proteins while edges (links) represent the predicted physical or functional interactions. To gain mechanistic insights into how drought influences protein-protein interactions, we analyzed the list of differentially expressed proteins (Figure 12) using the functional omics analysis server BarleyNet (Lee et al., 2020). This produced a network with one large cluster consisting of closely related up-regulated proteins (nodes) in the stem of Giza 132 (Figure 12 A). Down-regulated proteins form a network with five large clusters (Figure 12 B).
A

- **Cathepsin B-like cysteine proteinase**
- **Aspartic proteinase A**
- **Histone H2A 6**
- **Histone H2B.1**
- **HMG-Y-related protein**
- **Non-specific lipid transfer protein 4.2**
- **SNARE-interacting protein KEULE**
- **Calcium-dependent protein kinase 6**
- **Splicing factor U2af small subunit B**
- **4-coumarate:CoA ligase 1**
- **Chalcone-flavonone isomerase**
- **Ornithine decarboxylase**
- **Ubiquitin 6**
- **Phospho-2-dehydro-3-deoxyheptonate aldolase 2, chloroplastic**
- **Caleosin-related family protein**
Figure 12. Protein-protein interaction networks for proteins differentially regulated during drought in the stem of Giza 132. (A) Network for proteins up-regulated. (B) Network for proteins down-regulated. Interaction networks were constructed using the BarleyNet functional omics analysis server (Lee et al., 2020).
Gene ontology (GO) enrichment analysis, based on the interactions of differentially expressed proteins with other proteins that have known roles in various cellular functions, revealed the involvement of differentially expressed proteins in various cellular processes. Particularly, in drought tolerant Giza 132 proteins up-regulated during drought are associated with proteolysis, polyamine biosynthesis, flavonoid biosynthesis, and aromatic amino acid biosynthesis (Table 2), whereas down-regulated proteins are mainly involved in carbohydrate metabolism (Table 3).

Table 2: GO terms enriched in proteins up-regulated in Giza 132 during drought at the reproductive stage.

<table>
<thead>
<tr>
<th>Rank</th>
<th>TermID</th>
<th>TermDescription</th>
<th>p-value</th>
<th>Adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GO:0006508</td>
<td>Proteolysis</td>
<td>7.60E-04</td>
<td>1.44E-02</td>
</tr>
<tr>
<td>2</td>
<td>GO:0006596</td>
<td>Polyamine biosynthetic process</td>
<td>2.72E-03</td>
<td>2.29E-02</td>
</tr>
<tr>
<td>2</td>
<td>GO:0009813</td>
<td>Flavonoid biosynthetic process</td>
<td>3.62E-03</td>
<td>2.29E-02</td>
</tr>
<tr>
<td>4</td>
<td>GO:0009073</td>
<td>Aromatic amino acid family biosynthetic process</td>
<td>8.13E-03</td>
<td>3.86E-02</td>
</tr>
</tbody>
</table>

Table 3: GO terms enriched in proteins down-regulated in Giza 132 during drought at the reproductive stage.

<table>
<thead>
<tr>
<th>Rank</th>
<th>TermID</th>
<th>TermDescription</th>
<th>p-value</th>
<th>Adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GO:0005975</td>
<td>Carbohydrate metabolic process</td>
<td>1.54E-25</td>
<td>4.16E-24</td>
</tr>
</tbody>
</table>

The small number of proteins, with significantly altered levels of abundance, limited the usefulness of GO enrichment analysis for identifying potential functional roles of these proteins. Therefore, further functional annotation of each individual protein was performed by thorough searching of available literature and the public databases UniProtKB, Ensembl, and QuickGO (Tables 4 & 5).
Table 4: Functional classification of proteins up-regulated during drought in Giza 132.

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Molecular Function</th>
<th>Biological Process</th>
<th>Cellular Component</th>
<th>logFC</th>
<th>P.Value</th>
<th>adj.P.Val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histone H2A</td>
<td>DNA binding, protein heterodimerization activity</td>
<td>Transcription regulation, chromatin silencing</td>
<td>Nucleosome, nucleus, chromosome</td>
<td>2.131593</td>
<td>7.26E-07</td>
<td>0.000895</td>
</tr>
<tr>
<td>SapB_1 domain-containing protein</td>
<td>Aspartic-type endopeptidase activity, peptidase activity, hydrolase activity</td>
<td>Proteolysis, lipid metabolic process,</td>
<td>Vacuole, membrane, integral component of membrane</td>
<td>1.646411</td>
<td>1.10E-05</td>
<td>0.007535</td>
</tr>
<tr>
<td>Papain-like cysteine proteinase</td>
<td>Cysteine-type endopeptidase activity, peptidase activity, cysteine-type peptidase activity, hydrolase activity</td>
<td>Proteolysis, proteolysis involved in cellular protein catabolic process</td>
<td>Extracellular space, lysosome</td>
<td>1.055141</td>
<td>1.99E-05</td>
<td>0.007535</td>
</tr>
<tr>
<td>Caleosin-related (family)</td>
<td>Monoxygenase activity, calcium ion binding</td>
<td>Lipid transport, membrane expansion</td>
<td>Membrane, integral component of membrane</td>
<td>0.808157</td>
<td>2.14E-05</td>
<td>0.007535</td>
</tr>
<tr>
<td>Abhydrolase_3 domain-containing protein</td>
<td>Hydrolase activity</td>
<td>Proteolysis</td>
<td>Integral component of membrane</td>
<td>0.631863</td>
<td>2.15E-05</td>
<td>0.007535</td>
</tr>
<tr>
<td>Serine protease</td>
<td>Serine-type endopeptidase activity, serine-type peptidase activity</td>
<td>Proteolysis</td>
<td>Membrane</td>
<td>1.311906</td>
<td>2.44E-05</td>
<td>0.007535</td>
</tr>
<tr>
<td>Conserved oligomeric Golgi complex, subunit 4</td>
<td>Structural constituent of oligomeric Golgi complex</td>
<td>Golgi-to-ER retrograde transport, protein transport</td>
<td>Golgi complex</td>
<td>0.520235</td>
<td>4.48E-05</td>
<td>0.011557</td>
</tr>
<tr>
<td>Abhydrolase_3 domain-containing protein</td>
<td>Hydrolase activity</td>
<td>Proteolysis</td>
<td>Integral component of membrane</td>
<td>0.485763</td>
<td>7.21E-05</td>
<td>0.014986</td>
</tr>
<tr>
<td>Hydroxycinnamyloyl-CoA shikimate/quinate hydroxycinnamyloyl transferase</td>
<td>Acyltransferase, transferase</td>
<td>Cell wall organization</td>
<td>Cytosol, membrane</td>
<td>0.922338</td>
<td>7.29E-05</td>
<td>0.014986</td>
</tr>
<tr>
<td>COPI-interacting protein 1</td>
<td>Activator</td>
<td>Transcription regulation</td>
<td>Nucleus</td>
<td>0.475839</td>
<td>9.21E-05</td>
<td>0.017474</td>
</tr>
<tr>
<td>Pectin lyase-like superfamily protein</td>
<td>Polygalacturonase activity, hydrolase activity, acting on glycosyl bonds</td>
<td>Cell wall organization</td>
<td>Extracellular region, cell wall, vacuole</td>
<td>1.094541</td>
<td>0.000111</td>
<td>0.019479</td>
</tr>
</tbody>
</table>

(table continues)
<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Molecular Function</th>
<th>Biological Process</th>
<th>Cellular Component</th>
<th>logFC</th>
<th>P.Value</th>
<th>adj.P.Val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peroxidase</td>
<td>Peroxidase activity, oxidoreductase activity, heme binding, metal ion binding</td>
<td>ROS</td>
<td>Extracellular region</td>
<td>0.758823</td>
<td>0.000122</td>
<td>0.019479</td>
</tr>
<tr>
<td>Chalcone--flavonone isomerase</td>
<td>Intramolecular lyase activity</td>
<td>Biosynthesis of secondary metabolites, flavonoids biosynthetic process</td>
<td>Cytoplasm</td>
<td>0.348214</td>
<td>0.000465</td>
<td>0.037025</td>
</tr>
<tr>
<td>Aspartic protease A1</td>
<td>Aspartic-type endopeptidase activity</td>
<td>Proteolysis</td>
<td>Lysosome</td>
<td>0.933369</td>
<td>0.000492</td>
<td>0.037942</td>
</tr>
<tr>
<td>VHS domain-containing protein</td>
<td>Protein binding</td>
<td>Protein transport</td>
<td>Intracellular anatomical structure, membrane, integral component of membrane</td>
<td>0.423817</td>
<td>0.000565</td>
<td>0.038397</td>
</tr>
<tr>
<td>Cysteine peptidase</td>
<td>Cysteine-type endopeptidase activity</td>
<td>Proteolysis, proteolysis involved in cellular protein catabolic process</td>
<td>Extracellular space, lysosome</td>
<td>0.735194</td>
<td>0.000576</td>
<td>0.038397</td>
</tr>
<tr>
<td>Sec1-like protein (SNARE-interacting)</td>
<td>Syntaxin binding, metal ion binding</td>
<td>Protein transport, vesicle docking involved in exocytosis, vesicle-mediated transport</td>
<td>Retrotransposon nucleocapsid integral component of membrane, secretory granule</td>
<td>0.442467</td>
<td>0.000594</td>
<td>0.038397</td>
</tr>
<tr>
<td>Copper-transporting ATPase</td>
<td>Translocase</td>
<td>Ion transport</td>
<td>Membrane</td>
<td>0.602872</td>
<td>0.000607</td>
<td>0.038397</td>
</tr>
<tr>
<td>Ornithine decarboxylase</td>
<td>Catalytic activity, ornithine decarboxylase activity</td>
<td>Biosynthesis of secondary metabolites, polyamine biosynthetic process, putrescine biosynthetic process from ornithine</td>
<td>Cytoplasm</td>
<td>0.851269</td>
<td>0.00063</td>
<td>0.038841</td>
</tr>
<tr>
<td>Ubiquitin</td>
<td>Structural constituent of ribosome, protein binding, protein tag, ubiquitin protein ligase binding</td>
<td>Translation regulation, protein ubiquitination, modification-dependent protein catabolic process</td>
<td>Nucleus, cytoplasm, ribosome</td>
<td>1.034201</td>
<td>0.000671</td>
<td>0.040404</td>
</tr>
<tr>
<td>VQ domain-containing protein</td>
<td>DNA binding</td>
<td>Transcription regulation</td>
<td>Nucleus</td>
<td>0.383852</td>
<td>0.0007</td>
<td>0.041107</td>
</tr>
</tbody>
</table>

(table continues)
<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Molecular Function</th>
<th>Biological Process</th>
<th>Cellular Component</th>
<th>logFC</th>
<th>P.Value</th>
<th>adj.P.Val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histone H2B</td>
<td>DNA binding, protein heterodimerization activity,</td>
<td>Transcription regulation, nucleosome assembly</td>
<td>Nucleosome, nucleus, chromosome</td>
<td>0.681341</td>
<td>0.000756</td>
<td>0.04337</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peptidyl-prolyl cis-</td>
<td>mRNA binding, peptidyl-prolyl cis-trans isomerase activity, cystersporin A binding,</td>
<td>Protein folding, chaperone</td>
<td>Cytoplasm, chloroplast, apoplast</td>
<td>0.97472</td>
<td>0.000788</td>
<td>0.04419</td>
</tr>
<tr>
<td>trans isomerase</td>
<td>trans isomerase activity, cyclosporin A binding, isomerase activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSP20-like chaperone</td>
<td>Protein self-association, unfolded protein binding</td>
<td>Protein folding, chaperone</td>
<td>Cytoplasm</td>
<td>0.326685</td>
<td>0.000869</td>
<td>0.04662</td>
</tr>
<tr>
<td>(domain)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histone H2B</td>
<td>DNA binding, protein heterodimerization activity</td>
<td>Transcription regulation, nucleosome assembly</td>
<td>Nucleosome, nucleus, chromosome</td>
<td>0.418415</td>
<td>0.000896</td>
<td>0.047055</td>
</tr>
<tr>
<td>Histone H2B</td>
<td>DNA binding, protein heterodimerization activity</td>
<td>Transcription regulation, nucleosome assembly</td>
<td>Nucleosome, nucleus, chromosome</td>
<td>0.486447</td>
<td>0.000955</td>
<td>0.048823</td>
</tr>
</tbody>
</table>
Table 5: Functional classification of proteins down-regulated during drought in Giza 132.

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Molecular Function</th>
<th>Biological Process</th>
<th>Cellular Component</th>
<th>logFC</th>
<th>P.Value</th>
<th>adj.P.Val</th>
</tr>
</thead>
<tbody>
<tr>
<td>PfkB domain-containing protein</td>
<td>Fructokinase activity, kinase activity, transferase activity, phosphotransferase activity, alcohol group as acceptor.</td>
<td>Carbohydrate metabolism</td>
<td>Cytosol</td>
<td>-0.71958</td>
<td>5.20E-07</td>
<td>0.000895</td>
</tr>
<tr>
<td>Sucrose:fructan 6-fructosyltransferase</td>
<td>Hydrolase activity, hydrolyzing O-glycosyl compounds, transferase activity, hydrolase activity, transferring glycosyl groups, levansucrase activity, hydrolase activity, acting on glycosyl bonds</td>
<td>Carbohydrate metabolism</td>
<td>Membrane, integral component of membrane</td>
<td>-0.62111</td>
<td>2.38E-05</td>
<td>0.007535</td>
</tr>
<tr>
<td>23 kDa jasmonate-induced protein</td>
<td>RNA binding</td>
<td>Translation regulation</td>
<td>Cytosol</td>
<td>-0.679</td>
<td>4.68E-05</td>
<td>0.011557</td>
</tr>
<tr>
<td>Thiamine thiazole synthase</td>
<td>Synthase</td>
<td>Thiamine biosynthetic process</td>
<td>Chloroplast, mitochondrion</td>
<td>-0.4786</td>
<td>0.000548</td>
<td>0.038397</td>
</tr>
<tr>
<td>Jacalin-type lectin domain-containing protein</td>
<td>Carbohydrate binding</td>
<td>Plant defense</td>
<td>Apoplast, cytoplasm</td>
<td>-0.33922</td>
<td>0.000571</td>
<td>0.038397</td>
</tr>
<tr>
<td>Protein p23k-4</td>
<td>Carbohydrate binding</td>
<td>Carbohydrate transport</td>
<td>Membrane, integral component of membrane, ER membrane protein complex</td>
<td>-1.03138</td>
<td>0.000847</td>
<td>0.04643</td>
</tr>
</tbody>
</table>
The proportion of functional classes that are associated with differentially regulated proteins is presented in Figure 13. Drought induced proteins were associated with proteolysis (24%), transcription regulation (21%), lipid transport (8%), protein transport (5%), cell wall organization (5%), protein folding and chaperone (5%), aromatic amino acid biosynthesis (5%), translation regulation (10%), secondary metabolites biosynthesis (8%), protein phosphorylation (3%), protein phosphatase (3%), and copper ion transport (3%). We found down-regulated proteins to be associated with carbohydrate metabolism (33%), translation regulation (16%), thiamine biosynthesis (17%), plant defense (17%), and carbohydrate transport (17%).

Figure 13. Pie charts showing the percentage of differentially expressed proteins involved in biological processes, during the grain filling stage in the penultimate internode of drought-stressed Giza 132. The characterization of biological processes is based on the 38 up-regulated proteins (left) and the 6 down-regulated proteins (right).
Drought causes cellular damage and degrades crop productivity. Plants deploy various, finely tuned strategies, to protect against stress-induced cellular damages. At the proteome and the metabolite levels, many molecules are differentially regulated to provide protection against drought (Ghatak et al., 2021; Hein et al., 2016).

In our study, we identified 2,467 proteins. However, only proteins in drought tolerant barley Giza 132 were differentially regulated. We used these differentially regulated proteins to reconstruct networks against the 1,272,200 co-functional links among the 26,145 high-confidence barley genes available in the BarleyNet server (Lee et al., 2020). We also obtained useful information (Tables 4 & 5) about molecular function, biological process, and cellular components of differentially expressed proteins by searching against online databases (UniProtKB, QuickGO, and Ensembl Plants) and carefully examining the relevant literature.

The fact that only the proteome of drought tolerant Giza 132 changed significantly (Figure 11) in response to drought suggests that the DEPs proteins are related to the drought-tolerance mechanisms that are specific to this cultivar. It's worth mentioning that we applied gradual non-lethal drought in order to mimic natural drought and identify proteins relevant for stress tolerance and reduce cellular damage.

GO functional analysis revealed that the 38 up-regulated proteins were primarily involved in proteolysis and the biosynthesis of flavonoids, polyamines, and aromatic amino acids (Table 2) while the 6 down-regulated proteins were involved in carbohydrate
metabolism (Tables 3). Aromatic (phenylalanine) and nonaromatic (arginine) amino acids are necessary for the biosynthesis of flavonoids (Figure 18) and polyamines (Figure 19), respectively. Next, we will discuss the role of these processes in drought tolerance.

**Up-Regulated Proteins**

**A. Proteolysis**

Proteolysis plays an essential role in plant physiology and development. Plant cells use protein synthesis and degradation to constantly update their protein content during different developmental stages and environmental conditions. Plants use post-translational modifications such as protein phosphorylation (protein kinases) and ubiquitination (E3 ligases) to control the rate of protein turnover and balance between protein degradation and synthesis (Nelson & Millar, 2015) (Figure 14). It is used by plant cells for housekeeping by targeted degradation of abnormal and unnecessary proteins to supply amino acids for the de novo synthesis of drought response proteins. In addition to the common lysosomal proteolysis, our data shows up-regulation of proteolysis specific proteins in different organelles such as chloroplast, apoplast, thylakoid, vacuole, and membrane as can be seen under the header cellular component (Table 4).

This further emphasizes the importance of protein degradation as it can occur in different compartments of the plant cell hence controlling the diverse cellular activities to mitigate drought. Proteolysis regulates a wide range of cellular activities through controlling the amino acid supply within the cell and by reducing the abundance of key enzymes and regulatory proteins. Therefore, it controls the plant's metabolism, homeostasis, and development. For example, Wang et al. (2006) compared wild type
Arabidopsis leaves to those of a mutant Arabidopsis with a dysfunctional ubiquitin ligase system and found differential expression of proteins that were involved in photomorphogenesis, circadian oscillation, post-translation process, stress-responses and cell expansion in the mutant leaves. They also compared transcript levels of those genes but found no differences, which suggests that those differentially expressed proteins were regulated by the ubiquitin proteolysis.

The recent review by D'Ippólito et al. (2021) summarizes our current knowledge about the role of proteinases in the signaling pathways of plants in response to drought. Multiple studies have shown the link between proteolysis and other processes that are critical for plant adaptation to drought stress. For example, Dou et al. (2021) found that in Arabidopsis thaliana E3 ubiquitin ligase MREL57 attaches to the microtubule stabilizing protein WDL7, tagging it for degradation to mediate stomatal closure, through regulation of microtubule disassembly during drought and abscisic acid treatment. The effectiveness of the proteolysis pathway for drought tolerance stems from its ability to target specific proteins for degradation which in turn enables it to regulate various cellular processes. Additionally, protein degradation affects the nutritional content of plants through regulating the biosynthesis of secondary metabolites and free amino acid content.
Figure 14. Protein biosynthesis and degradation in plants and its regulation. The ubiquitin proteasome system (UPS) uses a group of ligases (E1, E2, E3) to tag proteins for degradation. The SP1 protein in the outer plastid membrane and facing the cytosol is an E3 ligase that interacts with the translocon of the outer chloroplast (TOC) to import necessary proteins and UPS. Ubiquitin protease 27 (UBP27) is located on the outer mitochondrial membrane and is involved in mitochondrial morphology. Specific proteases degrade plastid and mitochondrial proteomes, notably the AAA-class Lon, FtsH and Clp proteases. Post-translational modification processes can influence degradation. For example, DELLA proteins regulate gibberellin-dependent growth in plants and their degradation is controlled by phosphorylation, with catabolism ultimately mediated by another E3 ligase, SCF. Autophagy is another pathway for protein degradation which is localized in the vacuole and plays an important role in plant response to various physiological conditions. The mitogenic target of rapamycin (mTOR) is an important regulator of these processes that inhibits autophagy and stimulates protein synthesis via its effects on the small ribosomal subunit 6 kinase (S6K) and E2Fa transcription factor. The figure is based on Nelson and Millar (2015).
B. Biosynthesis of aromatic amino acids and secondary metabolites

The seven-step synthesis of chorismate in the shikimate pathway begins with the condensation of two intermediates of carbohydrate metabolism, phosphoenolpyruvate (PEP) and erythrose 4-phosphate (E4) which are synthesized in the glycolysis and pentose phosphate pathways, respectively, in a reaction catalyzed by 3-deoxy-D-arabino-heptulosonate-7-phosphosphate (DAHP) synthase (Figure 16). This enzyme catalyzes the step committed to the shikimate pathway, hence it controls the amount of carbon entering the pathway. We found a significant increase in the abundance of DAHP synthase in Giza 132 during drought (Table 4). Furthermore, this enzyme is a metal-dependent enzyme (including copper) which could explain the increase in abundance of copper ion transporter protein (Table 4).
Figure 15. Schematic representation of shikimate, phenylpropanoid, and flavonoid biosynthesis pathways (Pott et al., 2020).
Figure 16. Reactions of the Shikimate pathway. Reactions denoted as numbers are catalyzed by following enzymes: (1) 3-deoxy-D-arabino-heptulosonate-7-phos-phate (DAHP) synthase, (2) dehydroquinate synthase, (3) 3-dehydroquinatde dehydratase, (4) shikimate-NADP oxidoreductase, (5) shikimate kinase, (6) 5-enolpyruvylshikimate-3-phosphate synthase, (7) chorismate synthase, (8) chorismate mutase, (9) prephenate dehydrogenase, (10) phenylalanine-2-oxoglutarate transaminases, (11) phenylalanine 4-monooxygenase, (12) anthranilate synthase, (13) anthranilate phosphoribosyl transferase, (14) phosphoribosyl anthranilate isomerase, (15) indole-3-glycerolphosphate synthase, (16) tryptophan synthase. The star indicates up-regulation. Image based on Bromke (2013).

The three aromatic amino acids, tryptophan, tyrosine, and phenylalanine, are derived from the common precursor chorismate, which is the end product of the shikimate pathway (Figure 16). These amino acids and their metabolism are linked to the synthesis
of secondary metabolites including flavonoids, plant hormones, and a variety of different biological polymers (Parthasarathy et al., 2018) (Figure 17).

Polyphenol synthesis occurs through the general phenylpropanoid pathway, which uses aromatic amino acids (phenylalanine and tyrosine) to make many molecules with a phenol backbone by the action of multiple enzyme super families, including but not limited to ligases, oxygenases, oxidoreductases, and transferases. The initial steps of the phenylpropanoid pathway are catalyzed by phenylalanine ammonia lyase (PAL), cinnamate-4-hydroxylase (C4H), and 4-coumaroyl CoA-ligase (4CL) (Figure 14). We found a significant increase in the abundance of 4CL in Giza 132 in response to drought (Table 4). This up-regulation of the phenylpropanoid pathway was expected because we found an increase in the expression of major enzymes involved in both the aromatic amino acid biosynthesis pathway (provides the aromatic amino acid phenylalanine as a precursor for the phenylpropanoid pathway) and the flavonoids biosynthesis pathway (uses 4-Coumaroyl CoA, which is produced from the phenylpropanoid pathway). Hein et al. (2016) found a significant accumulation of five different families of amino acids, including phenylalanine in the spike organs of barley during drought.
C. Flavonoids

Plants use their biosynthetic machinery to synthesize a diverse array of metabolites needed for maintenance of internal homeostasis under different developmental stages and changing environments. These metabolites include flavonoids, which are carbon-based, polyphenolic secondary metabolites synthesized from amino acids. Flavonoids are classified into six major subgroups: chalcones, flavonols, flavones, flavandiols, anthocyanins, and proanthocyanidins (Ferreyra et al., 2012). Flavonoids have various functions in plant growth, development, reproduction, and protection from stress (Shah & Smith, 2020). In addition to the established functions as ROS scavenging and
UV radiation protection, plant flavonoids are up-regulated in response to a wide range of abiotic stresses, such as cold, salinity, and drought (Dias et al., 2020, 2012; Ma et al., 2014; Panche et al., 2016). However, little is known about the specific role of the flavonoid biosynthesis pathway and its key enzymes in plant response to drought (Ma et al., 2014).

Flavonoid-based compounds are produced by the condensation of one molecule of 4-coumaroyl-CoA and three molecules of malonyl-CoA by the enzyme chalcone synthase (CHS). Phenylalanine and malonyl-CoA are essential precursors for the synthesis of flavonoids and they originate from the shikimate pathway and acetate pathway, respectively. Phenylalanine, once produced from the shikimate pathway, undergoes three enzymatic reactions in the phenylpropanoid pathway to make Coumaroyl-CoA (Figure 18). We detected an increase in the expression of both chalcone isomerase (CHI) and 4-coumaroyl CoA-ligase (4CL), respectively (Table 4). Our results agree with Ma et al. (2014) who showed that flavonoids accumulate in wheat leaves in response to drought. Increased production of flavonoids and improved drought tolerance has also been observed in *Arabidopsis thaliana* (Nakabayashi et al., 2014).

In barley, Yuan et al. (2018) compared changes in the transcriptome and metabolome of drought tolerant and sensitive Tibetan hulless barley using polyethylene glycol. They found that 118 metabolites accumulated in 1 h of drought treatment (early drought), 117 accumulated after 4 h of stress (intermediate drought) and 92 after 48 h (severe drought). These metabolites were enriched with flavonoids and the products of phenylpropanoid biosynthesis pathway. They also observed down-regulation of the lignin
biosynthesis pathway. Reprogramming of the phenylpropanoid pathway and down-regulation of lignin biosynthesis improves tolerance to drought stress by modifying the architecture of the cell wall (Xu et al., 2021).

Figure 18. Diagram of the flavonoid biosynthetic pathway in plants based on Ravaglia et al. (2013). Genes encoding enzymes for each step are indicated as follows: PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate-CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; DFR, dihydroflavonol 4-reductase; LDOX, leucoanthocyanidin dioxygenase; FLS, flavonol synthase; LAR, leucoanthocyanidin reductase; ANR, anthocyanidin reductase; UFGT, UDP-glucose: flavonoid-3-O-glycosyltransferase. The enzymes next to the star are the up-regulated enzymes found in Giza132.
Additionally, there is evidence linking flavonoids with the control of the polar transport of the plant hormone auxin, which plays a role in stress response by regulating stomatal opening and by allocating resources (Ferreyra et al., 2012).

D. Polyamines

Polyamines are low molecular weight amine-containing molecules that are ubiquitous in plants and are present in their free forms. Putrescine, spermidine, and spermine are the main polyamines in plants and they have been considered as a new kind of plant biostimulant that regulates diverse physiological processes such as flower development, embryogenesis, organogenesis, senescence, fruit maturation and development, and environmental stress responses (Chen et al., 2019).

Putrescine is the precursor of spermidine and spermine. There are two different routes of putrescine biosynthesis in plants. In the main putrescine synthesis pathway, arginine is converted into agmatine by arginine decarboxylase (ADC). Then, the enzyme agmatine iminohydrolase removes a nitrogen atom from agmatine to make N-carbamoyl putrescine (NCPA). Finally, NCPA is hydrolyzed by N-carbamoylputrescine amidohydrolase (NCPAH) to form putrescine (Chen et al., 2019) (Figure 19). In the second route, ornithine is made from arginine by the enzyme arginase. Then, ornithine decarboxylase (ODC) removes the carboxyl group from ornithine to form putrescine. The ODC gene has been lost from Arabidopsis thaliana and many members of the Brassicaceae indicating that the ornithine pathway is not essential for normal growth (Chen et al., 2019).
The primary polyamine biosynthesis pathway involves three enzymatic steps that go from arginine to putrescine while the second less common pathway involves two steps making it more efficient in producing putrescine (Figure 19). We found that ODC is up-regulated in the drought tolerant Giza 132 (Table 4). Giza 132 may have conserved the pathway that uses ODC as an adaptation giving it the ability to produce polyamines with a higher efficiency during drought. There is a third pathway that converts arginine into citrulline but has only been found in sesame (Chen et al., 2019). Studies using exogenous polyamines, polyamine synthesis inhibitors, and transgenics have revealed that polyamines are important in plant growth, stability of nucleic acids and membranes, and stress resistance and survival (Agudelo-Romero et al., 2013; Pál et al., 2015; Sequera-Mutiozabal et al., 2016).
D. Protein kinases and phosphatases

The plant cell, like any other living cell, uses calcium as a secondary messenger to communicate and drive intracellular processes in response to drought or other stimuli. Calcium carries its role in cellular signaling by interacting with calcium binding proteins such as calcium-dependent protein kinases (CDPKs). In our study, we showed that calcium-dependent protein kinase 6 (CPK6) is up-regulated in response to drought in Giza 132 (Table 4). This protein has been suggested by Mori et al., 2006 to function as a calcium sensor and positive transducer of stomatal ABA signaling in Arabidopsis.

Moreover, CDPKs control ABA-induced stomatal closure through regulation of different ion channels in guard cells and regulation of gene expression of different components of

---

Figure 19. Polyamine biosynthetic pathway in plants based on Wojtasik et al. (2015). Putrescine is synthesized directly from ornithine by ornithine decarboxylase and indirectly from arginine by arginine decarboxylase, agmatine iminohydrolase and N-carbamoylputrescine amidohydrolase. Spermidine and spermine are, respectively, formed by the subsequent addition of an aminopropyl moiety to putrescine and spermidine by spermidine synthase and spermine synthase, respectively. Two enzymes, diamine oxidase, and polyamine oxidase, participate in degradation of polyamine. Up-regulated proteins are labeled with a star.
such channels (Zhu et al., 2007; Ronzier et al., 2014; Geiger et al., 2010; Franz et al., 2011).

E. Regulation of transcription and translation

Up-regulation of histones H2A and H2B in Giza 132 during drought (Table 4) suggests that histone assembly and disassembly is important for transcription of certain genes. Other studies have also shown that histone H2A variants play a major role in chromatin dynamics impacting chromatin condensation and accessibility for transcription (Osakabe et al., 2018).

Aubert et al. (2010) tested the role of stress-inducible caleosin family gene (RD20), which encodes for a peroxigenase and acts as a fatty acid hydroperoxide reductase, in Arabidopsis thaliana. They showed that rd20 knock-out had a higher transpiration rate that correlated with enhanced stomatal opening and reduced tolerance to drought compared with the wild type. We also observed up-regulation of RD20 in Giza 132, suggesting that this protein plays a role in drought tolerance in barley as well (Table 4).

Down-Regulated Proteins

A. Carbohydrate metabolism and transport

The stability of central carbohydrate metabolism plays a crucial role in plant stress response. Carbohydrates are involved in numerous metabolic and stress-responsive reactions and are well known for their diverse roles in signaling processes (Weiszmann et al., 2018). Phosphofructokinase (PFK), a key regulatory enzyme in plant glycolysis
(Figure 20), catalyzes the phosphorylation of fructose-6-phosphate to fructose-1,6-bisphosphate. In the primary plant glycolytic pathway, PFK is the main rate-limiting enzyme and regulatory point (Wang et al., 2021). PFK is an ATP-dependent enzyme that is controlled by feedback inhibition by phosphoenolpyruvate. When PFK is inhibited, the plant cell stops providing the mitochondria with respiratory substrates (pyruvate and malate). Therefore, the down-regulation of PFK (Table 5) would slow the consumption of glucose by glycolysis leading to an increase in the pool of available glucose in the stem cells which can be mobilized to the grain during grain filling under drought conditions.

Additionally, regulation of sucrose synthesis and degradation is critical to many stress-related processes because it is the primary product of photosynthetic tissues and the primary sugar transported in the phloem of most plants making it important for stress tolerance. In non-photosynthetic tissues, the transported sucrose serves as the primary precursor for energy as well as carbon skeleton for the production of other organic molecules such as amino acids, nucleotides, and structural carbohydrates (Stein & Granot, 2019). Additionally, sucrose can be transported to the vacuole (Figure 20), where it can be stored, transformed into fructans by the enzyme fructosyltransferases (which we found to be down-regulated), or hydrolyzed by vacuolar invertase to be stored as hexoses (Figure 20). Down regulation of fructosyltransferases would increase the pool of sucrose which plants mobilize to the grain under drought conditions.
Figure 20. Key enzymes in the primary carbohydrate metabolism. The 13 selected enzymes (underlined) are key enzymes of primary carbohydrate metabolism, among which cell wall invertase (cwInv), vacuolar invertase (vacInv), cytoplasmic invertase (cytInv), and sucrose synthase (Susy) possess sucrolytic activity, fructokinase (FK) is important for sucrose biosynthesis; hexokinase (HXK), phosphoglucoisomerase (PGI), phosphofructokinase (PFK), and aldolase (Ald) are important for glycolysis; ADP-glucose pyrophosphorylase (AGPase) and phosphoglucomutase (PGM) are important for starch biosynthesis; UDP-glucose pyrophosphorylase (UGPase) is important for cell wall biosynthesis; and glucose-6-phosphate dehydrogenase (G6PDH) is important for the oxidative pentose phosphate pathway. Down-regulated proteins are labeled with a star. Image based on Jammer et al. (2015).
B. Thiamine biosynthesis (Thiamine synthase)

Thiamine plays important roles in both plant growth and development (Figures 21 & 22). Thiamine pyrophosphate (TPP) is the active form of thiamine and functions as a cofactor in plant central energy metabolizing enzymes which are pyruvate dehydrogenase (PDH) of glycolysis, α-ketoglutarate dehydrogenase (α-KGDH) of the tricarboxylic acid cycle, and transketolase (TK) of the Calvin-Benson cycle. It is also a cofactor in aerobic energy metabolism, carbohydrate catabolism, pentose phosphate pathway, and branched-chain amino acid biosynthesis (Feng et al., 2019). Feng et al., 2019 observed a significant decrease in the activities of pyruvate dehydrogenase (PDH) and pyruvate decarboxylase (PDC) in thiamine deficient mutant soybean cultivars compared to wild type plants.

Plants are known to respond to drought by synthesizing abscisic acid (ABA) which function as mobile signal or a hormone that has the effect of inducing stomatal closure, through activation of guard cells anion channels, in order to reduce transpirational water loss (Li et al., 2016). Thiamine thiazole synthase, which we found to be downregulated (1.4-fold), has been found to be involved in guard cell abscisic acid (ABA) signaling and drought response in Arabidopsis (Arabidopsis thaliana) where plants overexpressing thiamine thiazole synthase proved to be more sensitive to ABA than the wild type (Li et al., 2016). These studies suggest additional roles of thiamine thiazole synthase in the regulation of plant abiotic stress responses in addition to its known functions in thiamine biosynthesis and mitochondrial DNA damage tolerance. Moreover, thiamine is a very important nutrient because it functions as an essential cofactor in several enzymatic reactions, in all living organisms. Our finding that drought
not only reduces crops yield but also their nutritional value. Interestingly, Li et al. (2016) discovered a novel regulatory role of thiamine thiazole synthase where it regulates kinase activity in guard cell signaling by interacting with and suppressing the calcium-dependent protein kinase, CDPK23. The calcium-dependent protein kinase (CDPK) family is critical in plant signaling. We found CDPK6 (Table 4), which belongs to the subclass of stress-inducible CDPKs and is stimulated by salt and osmotic stress (Xu et al., 2010), to be up-regulated in barley stem in response to drought, which might confer tolerance by interacting with different proteins as shown by Li et al. (2016).
Figure 21. Cellular location of thiamine-dependent enzymes. Thiamine pyrophosphate (TPP), the active form of thiamine, works as an essential coenzyme for the enzymes involved in photosynthesis in the chloroplast, pentose phosphate pathway, alcoholic fermentation, and ATP synthesis in oxidative decarboxylation of pyruvate and tricarboxylic acid cycle. Thiamine has also been shown to be directly involved in abiotic stress response, photoperiod, and works as an antioxidant scavenging ROS. Indirectly, thiamine contributes to the cell’s energy pool, conferring the necessary metabolic flexibility to adapt to new conditions. Thiamine-dependent enzymes shown are α-ketose transketolase (TK); 1-deoxy-D-xylulose-5-phosphate synthase (DXPS); acetohydroxyacid synthase (AHAS); pyruvate dehydrogenase (PDH); 2-oxoglutarate dehydrogenase (OGDH); and branched chain 2-oxoacid dehydrogenase (BCOADH). MEP, methyerythritol pathway. Image based on Rosado-Souza et al. (2020).
Figure 22. Illustration of plant intracellular communication. Anterograde (nucleus to organelle) and retrograde (organelle to nucleus) signaling pathways, as well as the main active site of ascorbate (Asc) and thiamine (B1) as signaling molecules, are shown. The ubiquitous existence of ascorbate and thiamine in cellular organelles, as well as the tight interconnection of the two vitamins between chloroplast and mitochondria, points to their important roles in the crosstalk between the two organelles. Image from Rosado-Souza et al. (2020).
CONCLUSION

Our findings suggest that Giza 132 adjusts its proteome dynamically to cope with drought. We did not observe changes in the proteome of the drought susceptible cultivar, Samson. This intrinsic proteome adjustment helped Giza 132 to efficiently overcome the effect of drought at the grain filling stage. A total of 44 proteins were exclusively regulated in Giza 132 during drought. These proteins are involved in proteolysis, flavonoids biosynthesis, polyamines biosynthesis, aromatic amino acids biosynthesis, and carbohydrate metabolism. They could potentially serve as novel stress tolerance biomarkers to improve drought tolerance in barley and other cereal crops. Also, the down-regulation of carbohydrate metabolism enzymes (PFK, fructosyltransferase, and thiamine synthase) would increase the pool of available soluble carbohydrates (glucose and sucrose) in the stem that could be mobilized to the grain under drought.
REFERENCES


APPENDIX A

DROUGHT TREATMENT SCHEDULE
### Samson

Experiment date: **3/22/2020** to **5/15/2020**

Post size: **2-gal**

Experiment name: **Shotgun proteomics of barley stem**

Soil saturated on: **3/17/2020**

Total soil: **15 kg** (1.5 kg/pot)

Vol. of water: **30 L** (2 L/kg soil)

Date pots were filled with 4.8 kg saturated soil: **3/21/2020**

Date weight at field capacity (4.8 kg/pot) was adjusted: **3/22/2020**

Date seeds were planted (8 seeds/pot): **3/22/2020**

Planting depth: **1.5 cm**

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
<th>Day 6</th>
<th>Day 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3/22/2020</td>
<td>5 g Osmocot</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Plant seeds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3/29/2020</td>
<td>† 100ml Jack's fertilizer</td>
<td>2nd leaf emerged</td>
<td>3rd leaf emerged</td>
<td>1st tiller appeared</td>
<td>Water to 4,000g</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Water to 4,000g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4/5/2020</td>
<td>† 1st tiller appeared</td>
<td>Water to 4,000g</td>
<td>3rd tiller appeared</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thin to 6 plants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4/12/2020</td>
<td>5 g Osmocot</td>
<td>Remove tillers</td>
<td>Remove tillers</td>
<td>4th tiller appeared</td>
<td>Water to 4,300g</td>
<td>Remove tillers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Water to 4,300g</td>
<td>Water to 4,300g</td>
<td>Water to 4,300g</td>
<td></td>
<td></td>
<td>Water to 4,300g</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4/19/2020</td>
<td>Remove tillers</td>
<td>Remove tillers</td>
<td></td>
<td></td>
<td></td>
<td>Remove tillers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Water to 4,300g</td>
<td>Water to 4,300g</td>
<td></td>
<td></td>
<td></td>
<td>Water to 4,300g</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4/26/2020</td>
<td>100ml Jack's fertilizer</td>
<td>Remove tillers</td>
<td>Remove tillers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Water to 4,300g</td>
<td>Water to 4,300g</td>
<td>Water to 4,300g</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5/3/2020</td>
<td>G Controls-4,300g</td>
<td></td>
<td>3rd leaf emerged</td>
<td>controls</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Drought-3,870g (90%)</td>
<td></td>
<td></td>
<td>Drought-3,440g (80%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5/10/2020</td>
<td>Controls-4,300g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Drought-3,400g (80%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† Jack’s Professional Petunia FeED Plus Mg is NPK 20-3-19, product no. 77770.

On day 14 from planting, thin plants to six plants/pot.**++)**.

Start removing tillers every 2 - 3 days until the end of the 6th week.

A Flag leaf emerges.

B Tip of awn emerges in 50% of plants.

C Flag leaf sheath opened.

D 50% of the awn is visible (2nd day after tip of awn emerges).

E 50% of the head emerges.

F Head fully emerges.

G First day of drought begins on day 48 from planting.

* Harvest on the 6th – 7th day of stress (55 – 56 days from planting) between 1200 and 1300 (12:00 – 1:00 p.m.). This corresponds to Zadoks stage 83 (early dough).
## Giza 132

**Important**: Plant two weeks after Samson (on 4/5/2020).

- **Experiment date**: 3/22/2020 to 5/15/2020
- **Post size**: 2-gal
- **Experiment name**: Shotgun proteomics of barley stem
- **Soil saturated on**: 3/17/2020
- **Total soil**: 15 kg (1.5 kg/pot)
- **Vol. of water**: 30 L (2 L/kg soil)

**Date pots were filled with 4.8 kg saturated soil**: 4/4/2020

**Date weight at field capacity (4.8 kg/pot) was adjusted**: 4/5/2020

**Date seeds were planted (8 seeds/pot)**: 3/22/2020

### Planting depth: 1.5 cm

### Week | Date | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Day 6 | Day 7 | Day of the week
---|---|---|---|---|---|---|---|---|---
1 | 4/5/2020 | 5 g Osmocot | | | 2nd tiller appeared | 1st tiller appeared | | | Seedlings emerge
2 | 4/12/2020 | 100ml Jack’s fertilizer | 2nd leaf emerged | 3rd leaf emerged | | | | Water to 4,000g
3 | 4/19/2020 | Water to 4,000g | 3rd tiller appeared | | | | | Water to 4,000g
4 | 4/26/2020 | 5 g Osmocot | Remove tillers | | | 4 Remove tillers | | Water to 4,300g
5 | 5/3/2020 | 100ml Jack’s fertilizer | Water to 4,300g | | | 5 Remove tillers | | Water to 4,300g
6 | 5/10/2020 | Controls- 4,300g Drought- 3,870g (90%) | Controls- 4,300g Drought- 3,440g (80%) | Controls- 4,300g Drought- 3,010g (70%) | Controls- 4,300g Drought- 2,795g (65%) | Controls- 4,300g Drought- 2,580g (60%) | Controls- 4,300g Drought- 2,365g (55%) | Controls- 4,300g Drought- 2,150g (50%) |

† Jack’s Professional Petunia FeED Plus Mg is NPK 20-3-19, product no. 77770.

a On day 14 from planting, thin plants to six plants/pot

b Start removing tillers every 2 - 3 days until the end of the 4th week.

A Flag leaf emerges.

b Tip of awn emerges in 50% of plants.

C Flag leaf sheath opened.

D 50% of the awn is visible (2nd day after tip emergence).

E 50% of the head emerges.

F Head fully emerges.

G First day of drought begins on day 36 from planting.

H Harvest on 6th – 7th day of stress (35 – 41 days from planting) between 1200 and 1300 (12:00 – 1:00 p.m.). This corresponds to Zadoks stage 83 (early dough).
APPENDIX B

R CODE

# Load packages
library("edgeR") # TMM normalization
library("limma") # DE analysis
library(tidyverse) # ggplot2 and much more
library(psych) # scatter plot panels
library(NormalyzerDE)
library(missMethods)

start_data<-read.delim(file="TMT_Morex V2/proteinGroups_Morex V2.txt", header=TRUE, row.names = 1, sep="\t")
# Remove contaminants
filtered = start_data$Only.identified.by.site=="+",] # the $ operator is used to extract or subset a specific part of a data object in R. Here we are pulling elements from the data object when Only.identified.by.site is +.
filtered = filtered$Reverse=="+",] # ! is the logical NOT operator in R.
filtered = filtered$Potential.contaminant=="+",]
# Extract reporter.intensity.corrected values
filtered = filtered[,40:57]
filtered==0 <- NA
# Barplot of a dataframe
total<-(colSums(filtered[,1:18],na.rm=TRUE)) # to find column totals using the colSums() function and na.rm to ignore missing values
barplot(t(as.matrix(total))) # vertical barplot of column totals
# Barplot of percent missing values
col_nas= colSums(is.na(filtered)) # Total number of NAs per column in a data frame
col_total=colSums (filtered, na.rm = T, dims = 1) # Column totals
col_percent = (col_nas * 100 )/(nrow(filtered))# Percentage of missing values per column
barplot(t(as.matrix(col_percent)))
# Return rows that have at least TWO non-NA values across all samples.
filtered=filtered[rowSums(is.na(filtered))< (length(filtered)-1),]
write.table(filtered, file="TMT_Morex V2/Reported intensity corrected data min 2 values across all samples_Morex V2.txt",sep="\t",row.names = T, quote = F)
# Normalize data using normalyzerDE
normalyzer(jobName="TMT_now", designPath="TMT_Morex V2/Normalyzer_design_matrix_TMT data.txt", dataPath="TMT_Morex V2/Reported intensity corrected data_min 2 values across all samples_Morex V2.txt")
# All normalization methods are better than log2 transformation. CycLoess normalization is slightly better (low replicate variability) and we used this data for downstream analysis.
# Import the CycLoess-normalized data into R Studio for imputation.
CycLoess<-read.delim(file="TMT_Morex V2/CycLoess-normalized.txt", header=TRUE, sep="\t", row.names=1)
# Plot multiple histograms with density and normal fits on one page using psych package.
# Increase the size of the plot pane. Otherwise R Studio will give "Error in plot.new(): figure margins too large".
multi.hist(CycLoess, dcol= c("blue","red"),dlyt=c("dotted", "solid"), main=colnames(CycLoess))
# Boxplot of the normalized data. If the plot is small, enter "dev.off()" to turn on/off multiple graphics devices
boxplot(CycLoess, main="Boxplot of CycLoess-normalized TMT data", col=" light blue")
plotDensities(CycLoess,legend=FALSE,main="Density plots of CycLoess-normalized TMT data")
# Make MDS plot
plotMDS(CycLoess, main="MDS plot of CycLoess-normalized TMT data") # samples clustered by TMT experiment
# We tried to impute the data using 'llsimpute' function of the pcaMethods. However, llsimpute produced negative numbers.
llsImpute_CycLoess <-llsImpute(CycLoess, k = 10, center = TRUE, correlation = "pearson", allVariables = TRUE)
llsImpute_CycLoess <- completeObs(llsImpute_CycLoess) # to get the estimated complete observation (cObs)

# We used the impute_LS_adaptive() function (LSA) in the missMethods package.
lsaImpute_CycLoess <- impute_LS_adaptive(CycLoess)
head(lsaImpute_CycLoess, n=1)

# Explore the imputed data using boxplots and density plots
boxplot(lsaImpute_CycLoess,col=rep(c('red','green','blue'), each=6), notch=TRUE, main='LSA-imputed data: run1 (red), run2 (green), run3 (blue)', xlab='TMT Samples', ylab='log2 of Intensity')

# We used the impute_LS_adaptive() function (LSA) in the missMethods package.
lsaImpute_CycLoess <- impute_LS_adaptive(CycLoess)
head(lsaImpute_CycLoess, n=1)

# The above box plots and density distributions are not well aligned. There was supposed to be the same amount of protein labeled in each sample. We have significant differences that we should correct with some basic normalizing. We can average the numbers below and compute normalization factors to make the sums end up the same.

# In [143]: separate the data by TMT experiment (run)
run1_raw <- data_raw[c(1:6)]
run2_raw <- data_raw[c(7:12)]
run3_raw <- data_raw[c(13:18)]

# Adjust each TMT experiment to equal signal per channel (sample loading normalization)
target <- mean(c(colSums(run1_raw), colSums(run2_raw), colSums(run3_raw)))
norm_facs <- target / colSums(run1_raw)
run1_sl <- sweep(run1_raw, 2, norm_facs, FUN = '*') # sweep() is useful for systematic manipulation of a large matrix data (column by column). Here, we want to produce a summary statistic on run1 data, beginning on the second column and multiplying (Fun=**) by norm_facs.
norm_facs <- target / colSums(run2_raw)
run2_sl <- sweep(run2_raw, 2, norm_facs, FUN = '*')
norm_facs <- target / colSums(run3_raw)
run3_sl <- sweep(run3_raw, 2, norm_facs, FUN = '*')

# make a pre-IRS data frame after sample loading (SL) normalizations
data_sl <- cbind(run1_sl, run2_sl, run3_sl)

# Look at the SL-normalized data with box and density plots (like a distribution histogram).
boxplot(data_sl, col = rep(c('red', 'green', 'blue'), each = 6), notch = TRUE, main = 'Intensities after normalization for sample loading (SL): run1 (red), run2 (green), run3 (blue)', xlab = 'TMT Sample', ylab = 'log2 of Intensity', las=3, par(mar=c(12,5,5,3)))

# In [144]: check the column totals. If the totals are the same, the sample loading (SL) normalization has worked.
format(round(colSums(data_sl), digits = 0), big.mark = ',')

# In [145]: Do TMM on raw data - we use it later for CV analyses
raw_tmm <- calcNormFactors(data_raw)
data_raw_tmm <- sweep(data_raw, 2, raw_tmm, FUN = '/') # raw data after TMM on original scale

# perform TMM on the SL-normalized data and visualize the resulting distributions.
sl_tmm <- calcNormFactors(data_sl)
data_sl_tmm <- sweep(data_sl, 2, sl_tmm, FUN = '/') # data after SL and TMM on original scale.

# Visualize the distribution of TMM-normalized.
boxplot(log2(data_sl_tmm), notch = TRUE, col = rep(c('red', 'green', 'blue'), each = 6), main = 'TMM normalization of SL data: run1 (red), run2 (green), run3 (blue)', xlab = 'TMT Sample', ylab = 'log2 of Intensity', las=3, par(mar=c(12,5,5,3)))

# In [146]: check column totals. Should be different
format(round(colSums(data_sl_tmm), digits = 0), big.mark = ',')

# In[147]: See how things cluster
plotMDS(log2(data_sl_tmm), col = rep(c("red", "green", "blue"), each = 6), main = "SL/TMM clusters group by TMT experiment")

# Perform internal reference scaling (IRS) normalization
irs <- tibble(rowSums(run1_raw), rowSums(run2_raw), rowSums(run3_raw)) # tibble()
constructs a data frame:
colnames(irs) <- c("sum1", "sum2", "sum3")
irs1 <- tibble(rowMeans(run1_raw, c(5:6)))
irs2 <- tibble(rowMeans(run2_raw, c(11:12)))
irs3 <- tibble(rowMeans(run3_raw, c(17:18)))
# Combine
irs <- cbind(irs1, irs2, irs3)
colnames(irs) <- c("sum1", "sum2", "sum3")
# get the geometric average intensity for each protein in the reference channel from the
irs data.
irs$average <- apply(irs, 1, function(x) exp(mean(log(x))))
# compute the scaling factor vectors
irs$fac1 <- irs$average / irs$sum1
irs$fac2 <- irs$average / irs$sum2
irs$fac3 <- irs$average / irs$sum3
# Extract all column in the imputed data excluding the reference channels (columns 1 to 4)
run1_raw <- run1_raw[, 1:4]
run2_raw <- run2_raw[, 1:4]
run3_raw <- run3_raw[, 1:4]
# make new IRS-normalized data frame
data_irs <- run1_raw * irs$fac1
data_irs <- cbind(data_irs, run2_raw * irs$fac2)
data_irs <- cbind(data_irs, run3_raw * irs$fac3)
# see how the IRS data look like using boxplots and density plots
boxplot(log2(data_irs), col = rep(c("red", "green", "blue"), each = 4), main = "Internal Reference Scaling (IRS) normalized data: run1 (red), run2 (green), run3 (blue)", xlab = 'TMT Sample', ylab = 'log2 of Intensity', notch = TRUE, las=3, par(mar=c(12,5,5,3)))
plotDensities(log2(data_irs), col = rep(c("red", "green", "blue"), 4), main = "IRS data", legend = FALSE)
# In [149]: check column totals
format(round(colSums(data_irs), digits = 0), big.mark = ",")
# In [150]: data_irs is the data after SL and IRS normalization on the original scale
# see if box plots are aligned and density plots are smooth
boxplot(log2(data_irs), notch = TRUE, col = rep(c("red", "green", "blue"), each = 4), main = "TMM normalization of IRS data: run1 (red), run2 (green), run3 (blue)", xlab = 'TMT Sample', ylab = 'log2 of Intensity', notch = TRUE, las=3, par(mar=c(12,5,5,3)))
plotDensities(log2(data_irs), col = rep(c("red", "green", "blue"), 4), main = "IRS/TMM data", legend = FALSE)
# In [151]: check the column totals
format(round(colSums(data_irs), digits = 0), big.mark = ",")
# In [152]: see how things cluster after IRS plus TMM
col_vec <- c("red", "orange", "green", "blue")
plotMDS(log2(data_irs), col = col_vec, main = "IRS/TMM clusters group by cultivar x treatment")
# In [164]: lets compare the combination of SL and TMM normalizations to SL/IRS/TMM
# again using the idea that replicates of the same treatment should be similar
# Before IRS: Samson
pairs.panels(log2(data_raw[c(1, 7, 13)]), lm = TRUE, main = "CycLoess-normalized Samson control")
pairs.panels(log2(data_raw[c(2, 8, 14)]), lm = TRUE, main = "CycLoess-normalized Samson drought")
# After IRS: Samson
pairs.panels(log2(data_irs[c(1, 5, 9)]), lm = TRUE, main = "IRS-normalized Samson control")
pairs.panels(log2(data_irs[c(2, 6, 10)]), lm = TRUE, main = "IRS-normalized Samson drought")
# Before IRS: Gizza132
pairs.panels(log2(data[c(3, 9, 15)]), lm = TRUE, main = "CycLoess-normalized Gizza132 control")
pairs.panels(log2(data[c(4, 10, 16)]), lm = TRUE, main = "CycLoess-normalized Gizza132 drought")
# After IRS: Gizza132
pairs.panels(log2(data_irs[c(3, 7, 11)]), lm = TRUE, main = "IRS-normalized Giza132 control")
pairs.panels(log2(data_irs[c(4, 8, 12)]), lm = TRUE, main = "IRS-normalized Giza132 drought")

# Compare the CV distributions for the different normalized data
# The values for the UNNORMALIZED data are extracted from the filtered data by excluding reference channels.
filtered_no_ref <- subset(filtered, select=c(1:4, 7:10, 13:16))
make_cvs <- function(df){
  Samson_control <- df[c(1, 5, 9)]
  Samson_drought <- df[c(2, 6, 10)]
  Giza132_control <- df[c(3, 7, 11)]
  Giza132_drought <- df[c(4, 8, 12)]
  Samson_control$ave <- rowMeans(Samson_control)
  Samson_control$sd <- apply(Samson_control[1:3], 1, sd)
  Samson_control$cv <- 100 * Samson_control$sd / Samson_control$ave
  Samson_drought$ave <- rowMeans(Samson_drought)
  Samson_drought$sd <- apply(Samson_drought[1:3], 1, sd)
  Samson_drought$cv <- 100 * Samson_drought$sd / Samson_drought$ave
  Giza132_control$ave <- rowMeans(Giza132_control)
  Giza132_control$sd <- apply(Giza132_control[1:3], 1, sd)
  Giza132_control$cv <- 100 * Giza132_control$sd / Giza132_control$ave
  Giza132_drought$ave <- rowMeans(Giza132_drought)
  Giza132_drought$sd <- apply(Giza132_drought[1:3], 1, sd)
  Giza132_drought$cv <- 100 * Giza132_drought$sd / Giza132_drought$ave
  ave_df <- data.frame(Samson_control$ave, Samson_drought$ave, Giza132_control$ave, Giza132_drought$ave)
  sd_df <- data.frame(Samson_control$sd, Samson_drought$sd, Giza132_control$sd, Giza132_drought$sd)
  cv_df <- data.frame(Samson_control$cv, Samson_drought$cv, Giza132_control$cv, Giza132_drought$cv)
  return(list(ave_df, sd_df, cv_df))
}
list_filtered_no_ref <- make_cvs(filtered_no_ref)
list_CycLoess <- make_cvs(data_raw)
list_irs <- make_cvs(data_irs)
par(mfrow = c(1, 3))
boxplot(list_filtered_no_ref[c(3)], notch = TRUE, main = "filtered CVs (Ave = 38.6%)", ylab = "CV (%)", las=3, par(mar = c(10, 5, 4, 5))
boxplot(list_CycLoess[c(3)], notch = TRUE, main = "CycLoess CVs (Ave = 1.6%)", ylim = c(0, 30), ylab = "CV (%)", las=3, par(mar = c(10, 5, 4, 5))
boxplot(list_irs[c(3)], notch = TRUE, main = "IRS CVs (Ave = 0.55%)", ylim = c(0, 10), ylab = "CV (%)", las=3, par(mar = c(10, 5, 4, 5)))
(filtered_med_cv <- round(mean(apply(list_filtered_no_ref[c(3)], na.rm=TRUE, median)), 2))
(CycLoess_med_cv <- round(mean(apply(list_CycLoess[[3]], 2, median)), 2))
(irs_med_cv <- round(mean(apply(list_irs[[3]], 2, median)), 2))
# Use CycLoess_irs_tmm for differential expression analysis.
write.table(data_irs, file = "TMT_Morex V2/CycLoess, SL and IRS-normalized data_Morex V2 assembly.txt", sep = "\t")

="/"
protein.
names(fit)
fit2 <- contrasts.fit(fit, contrast.matrix)
# Produce empirical Bayes statistics
fit2 <- eBayes(fit2)
topTable(fit2) # Makes a table of top differentially expressed proteins from the linear model. Default is 10.
# Produce a list of deferentially expressed proteins
results <- decideTests(fit2, lfc=0.321928095)
summary(results) # lists the number of up/down-regulated proteins
vennDiagram(results, include=c("up","down"), cex=c(1,1,1,1,1))
SD_SC<topTable(fit2, coef=1, n=Inf, sort.by="P")
GD_GC<topTable(fit2, coef=2, n=Inf, sort.by="P")
SC_GC<topTable(fit2, coef=3, n=Inf, sort.by="P")
SD_GD<topTable(fit2, coef=4, n=Inf, sort.by="P")
#Produce volcano plot of pairwise comparisions
EnhancedVolcano(GD_GC, lab= rownames(GD_GC),title = 'Differentially expresse proteins in Giza132',x= "logFC", y="adj.P.Val", FCcutoff = 0.321928095, pCutoff = 0.05, xlim = c(-1.8,1.8), ylim = c(0,-log10(10e-5)), labSize = 2, pointSize = 3, cutoffLineWidth = 1)
EnhancedVolcano(SD_SC, lab= rownames(SD_SC),title = 'Differentially expresssione proteins in Giza132',x= "logFC", y="adj.P.Val", FCcutoff = 0.321928095, pCutoff = 0.05, xlim = c(-1.2,1.2), ylim = c(0,-log10(10e-5)), labSize = 2, pointSize = 3, cutoffLineWidth = 1)
# Save the DE analysis result
write.table(SD_SC, file="TMT_Morex V2/Morex V2_LSA-imputed_SL_IRS-normalized_DE proteins during drought in Samson.txt", sep="\t")
write.table(GD_GC, file="TMT_Morex V2/Morex V2_LSA-imputed_SL_IRS-normalized_DE proteins during drought in Giza 132.txt", sep="\t")
write.table(SC_GC, file="TMT_Morex V2/Morex V2_LSA-imputed_SL_IRS-normalized_up-regulated proteins in control Samson Vs. control Giza 132.txt", sep="\t")
write.table(SD_GD, file="TMT_Morex V2/Morex V2_LSA-imputed_SL_IRS-normalized_up-regulated proteins in stressed Samson vs. stressed Giza 132.txt", sep="\t")