

Leveraging mass spectrometry-based proteomics and post-translational modifications for system-wide understanding of bioenergy plants and their environment

**A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville**

**Him K. Shrestha
December 2022**

Copyright © 2022 by Him K. Shrestha.
All rights reserved.

ACKNOWLEDGEMENTS

It has been a long road since I first decided to leave my home country and travel to the USA to continue my education. Along the road, I have met exceptional people that have widened my academic and cultural horizons. I would like to thank everyone who made my Ph.D. possible.

On the top of the list comes my educational father, advisor, and mentor Dr. Robert L. Hettich (Bob). I will always be thankful that I got to know Bob and learn from him. He provided me with the opportunities and resources needed for my successful graduate career. His fatherly advice has helped me through some of the roughest times of my graduate school. Throughout my time at graduate school, Bob has been a constant source of inspiration. His work ethics, wisdom, and ability to provide honest and constructive feedback are some of the qualities I would like to replicate in myself. Besides Bob, I would like to thank Dr. Paul E. Abraham and Dr. Richard J. Giannone, staff scientists in our group. Paul has been a great mentor who has helped to shape my graduate career. I had numerous opportunities to closely work with him and I love working with him. Similarly, Rich helped me a lot with valuable technical insights into mass spectrometry.

A special thanks to my respected committee members Dr. Albrecht von Arnim, Dr. Bode Olukolu, Dr. Jesse Labbe, and again Dr. Paul Abraham for their valuable time during committee meetings. Their critical feedback and suggestions have helped me improve my research on numerous occasions. I would like to acknowledge my graduate program Genome Science & Technology (GST) for providing me with opportunities and a positive environment to flourish during my graduate school.

I would like to thank Dr. Alexander Cope, my graduate school mentor, Dr. Alfredo Blakeley-Ruiz, my first officemate, and Dr. Suresh Poudel, my Nepali brother, who helped me acclimatize to graduate school during my early days. Similarly, I would like to thank Manasa R Appidi, my current officemate. I would also like to thank all my lab mates and friends from different parts of the world including India, Nepal, the U.S., Nigeria, Mexico, China, and other places. My interactions with them have always been pleasant and I will always cherish the time I got to spend with them.

Finally, I would like to thank my father, Mahendra Kumar Shrestha, and mother, Laxmi Shrestha. They have instilled the moral values and principles that make me the person I am today. They have been my biggest support system and have provided me with all the resources I needed to pursue higher education. Similarly, I thank my sisters Rekha Shrestha and Chadani Shrestha for always supporting and motivating me to do better. I would also like to thank my friend, Danielle. Since the beginning of our undergraduate, she has supported me through my highs and lows. She has always guided me on the right path and deserves lots of credit for success in my educational career. Lastly and importantly, I would like to thank my partner, Arati Shrestha. Without her support, this journey would be extremely hard.

ABSTRACT

Bioenergy research focuses on utilizing renewable biomass feedstocks to produce biofuels and bioproducts to address growing concerns about energy security and climate change. To better understand the details of bioenergy crop production and conversion to bioproducts, it is essential to characterize bioenergy plants and their environments at a molecular systems level. Mass spectrometry has emerged as a promising technique for detailed proteomic information, including post-translational modifications (PTMs), of molecular processes and cellular functions of biological systems. In this dissertation, proteomic approaches have been optimized and implemented to deepen our understanding of the interaction of plants and their environment in a bioenergy research context.

Growing studies support the concept of the plant holobiont, where plant growth and fitness are a result of the collective genome expression of both the plant and its associated microbes. This dissertation illustrates the optimization of proteomic workflows to understand how the poplar plant and its associated microbes respond to complex ecological interactions occurring in the rhizosphere, such as a key protein kinase that controls plant-fungal symbiosis. Similarly, abiotic factors such as the diurnal cycle play a key role in synchronizing plant behavior with external cues and are important for a plant's phenotypic plasticity. We have utilized an advanced bioinformatic workflow with *de novo*-assisted database searching to provide the first molecular evidence of how both proteins and their PTMs are regulated by the diurnal cycle in poplar plants.

For the goal of efficient conversion of biomass to biofuels, the plant's lignin content has been a key bottleneck because it limits the accessibility of cellulosic sugar. Genetic modification of the lignin biosynthesis pathway can help reduce lignin, but this often results in stunted growth phenotype. Multi-omics analysis of lignin-modified model grass, *Brachypodium distachyon* provided a deeper understanding of the undesired phenotypic effects caused by lignin modification and revealed key molecular signals that could be targeted for genetic manipulations.

Overall, this dissertation research is one of the first concerted proteomics investigations of the biotic and abiotic details of poplar and revealed key aspects of symbiotic microbial associations as well as the pleiotropic effects caused by lignin modification in *Brachypodium*.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION TO PROTEOMICS AND POST-TRANSLATIONAL MODIFICATIONS (PTMS) AND RESEARCH TOPICS COVERED IN THIS DISSERTATION	1
1.1 Overview and application of mass spectrometry-based proteomics to understand system-level biology.....	1
1.2 Need for post-translational modifications perspective in biological research	7
1.3 The challenges with PTM studies and recent advancements in the identification of PTMs	8
1.4 Introduction to the bioenergy plant research and its importance	10
1.4.1 Understanding bioenergy plants and their environment for improved fitness and growth of bioenergy plants	11
1.4.1.1 Plants' biotic environment	12
1.4.1.1.1 Plants' biotic environment: Microbe-microbe interactions	14
1.4.1.1.2 Plants' biotic environment: Plant-microbe interactions (Plant's response)...	15
1.4.1.1.3 Plants' biotic environment: Plant-microbe interactions (Microbe's response)	17
1.4.1.2 Plants' abiotic environment: Response to the diurnal cycle	18
1.4.2 Reducing the recalcitration in bioenergy plants for the efficient biomass conversion	19
1.4.2.1 Current hypotheses for lignin modification induced negative growth (dwarf phenotype).....	22
1.5 Dissertation summary	24
CHAPTER 2: EXPERIMENTAL, ANALYTICAL, AND COMPUTATIONAL METHODOLOGIES FOR PROTEOMICS AND PTM ANALYSIS	26
2.1 Overview of bottom-up proteomics and post-translational modifications workflow	26
2.2 Proteomics sample preparation	28
2.2.1 Considerations before proteomics sample preparation	28
2.2.2 Biological samples to protein content.....	29
2.2.3 Protein digestion and cleanup	31
2.2.4 Protein/Peptide enrichment for PTM analysis	32
2.3 Fundamentals of peptide separation and tandem MS/MS data acquisition	33
2.3.1 Peptide separation using reversed-phase (RP)	34
2.3.2 Multidimensional Protein Identification Technology (MudPIT).....	35
2.3.3 Peptide Ionization	38
2.3.4 Mass analyzer.....	41
2.3.5 Peptide fragmentation	47
2.3.6 Data acquisition Mode	49
2.4 Bioinformatics analysis of proteomics and PTMs data	49
2.4.1 Reference database search	50
2.4.2 Spectral library searching	51
	vii

2.4.3 <i>De-novo</i> peptide sequencing and <i>de-novo</i> assisted database searching	53
2.4.4 Peptide/Protein Quantification: Labeled and label-free quantification	54
2.4.5 PTM data analysis.....	54
2.4.6 From protein identification and quantification to biological interpretation.....	55
2.4.6.1 Data processing and quality control filtration.....	55
2.4.6.2 Statistical analysis	57
2.4.6.3 Data visualization and interpretation	58
2.5 Summary	59
CHAPTER 3: METAPROTEOMICS REVEALS INSIGHTS INTO MICROBIAL STRUCTURE, INTERACTIONS, AND DYNAMIC REGULATION IN DEFINED COMMUNITIES AS THEY RESPOND TO ENVIRONMENTAL DISTURBANCE	
3.1 Introduction.....	62
3.2 Materials and methods	65
3.2.1 Preparation of 10-member microbial communities	65
3.2.2 Cellular protein extraction	66
3.2.3 Extracellular protein extraction.....	67
3.2.4 Protein identification and quantification.....	68
3.2.5 PTM identification by PEAKS	69
3.3 Results.....	70
3.3.1 Metaproteomics reveals details of microbial community stabilization and population equilibration as a function of growth media	70
3.3.2 Quantifying microbial species reveals functional behaviors during community assembly.....	75
3.3.3 <i>Pseudomonas</i> sp. GM17 behaves antagonistically during the community selection process.....	75
3.3.4 <i>Pantoea</i> sp. YR343 is effective at adapting to the changes in the community in R2A media.....	79
3.3.5 <i>Bacillus</i> sp. BC15 sporulates as an adaptive response to nutritional competition.....	81
3.3.6 Post-translational modifications offer critical insights into how microbes are adapting to changing communities and environments.....	84
3.3 Discussion	86
3.4 Conclusions	90
CHAPTER 4: CROSS-SPECIES PHOSPHOPROTEOME PROFILING REVEALS A KINASE- INDUCED SIGNALING MECHANISM USED BY PLANTS TO PERCEIVE BENEFICIAL FUNGUS	
4.1 Introduction.....	93
4.2 Materials and Methods.....	97
4.2.1 Plant and fungal growth	97
4.2.2 Proteomics Sample Preparation	97

4.2.3 LC-MS/MS analysis	98
4.2.4 Proteome Database Search.....	98
4.2.5 Data Analysis	99
4.2.6 GO and Functional Pathway Analysis	99
4.2.7 Network analysis of switchgrass phosphoprotein interactions	100
4.2.8 Bimolecular fluorescence complementation assay (BiFC assay)	100
4.3 Results.....	100
4.3.1 Assessment of switchgrass phosphoproteomics results reveal distinct phosphorylation events between fungus inoculated transgenic and WT plant	100
4.3.2 Differentially phosphorylated proteins between fungus inoculated transgenic and WT switchgrass are regulated uniquely at the post-translational level.....	102
4.3.3 Functional analysis identified key processes and pathways regulated through phosphorylation by PtLecRLK1 in switchgrass	105
4.3.4 PtLecRLK1 regulates the phosphorylation status of defense-related transcription factors and kinases	107
4.3.5 PtLecRLK1 regulates nutritional uptake via phosphorylation of transporter proteins	108
4.3.6 Network analysis and molecular assay identified and validated cGMP kinase as a direct interacting partner of PtLecRLK1	110
4.3.7 Cross-species comparison of phosphorylated proteins showed similarity in the signaling mechanism.....	112
4.4 Discussions	112
CHAPTER 5: UNDERSTANDING THE PLANT-RESPONSIVE TRANSCRIPTION FACTOR PIPR REGULATED PATHWAYS IN THE POPULUS ENDOPHYTE <i>PSEUDOMONAS</i> GM79	
120	
5.1 Introduction to orphan LuxR (PipR) and its role in helping microbes perceive plant-derived metabolite.....	120
5.2 Materials and Methods.....	122
5.2.1 <i>Pseudomonas</i> sp. GM79 growth and bottom-up proteomics samples preparation....	122
5.2.2 LC-MS/MS and data analysis	122
5.3 Results and discussions.....	123
5.3.1 Proteomics results support the HEHEAA dependent PipR-signaling model of <i>Pseudomonas</i> GM79.....	123
5.3.2 Quantitative analysis reveals the differential regulation of pathways relevant to plant colonization.....	124
5.3.3 HEHEAA influences chemotaxis and motility in <i>Pseudomonas</i> GM79	127
5.3.4 HEHEAA increases the abundance of the Type VI secretion system in <i>Pseudomonas</i> GM79	129
5.3.4 HEHEAA increases the siderophore synthesis and transport in <i>Pseudomonas</i> GM79	130
5.4 Conclusions.....	130

CHAPTER 6: TEMPORAL DYNAMICS OF PROTEIN AND POST-TRANSLATIONAL MODIFICATION ABUNDANCES IN POPULUS LEAF ACROSS A DIURNAL PERIOD. 133

6.1 Introduction to the importance of the diurnal cycle.....	133
6.2 Proteomics and PTM approach to characterize this phenomenon at a molecular level....	134
6.3 Results.....	136
6.3.1 Proteome-wide changes across the diurnal period.....	136
6.3.2 PTM analysis provides the unique regulation throughout the diurnal cycle	139
6.3.3 An instance of dynamic regulation at both the protein and PTM level and its.....	142
biological implications.....	142
6.4 Conclusions.....	142

CHAPTER 7: MULTI-OMIC CHARACTERIZATION OF BIFUNCTIONAL PEROXIDASE 4-COUMARATE 3-HYDROXYLASE KNOCKDOWN IN *BRACHYPODIUM DISTACHYON* PROVIDES INSIGHTS INTO LIGNIN MODIFICATION-ASSOCIATED PLEIOTROPIC EFFECTS 144

7.1 Introduction.....	144
7.2 Materials and methods	147
7.2.1 Plant material and growth conditions	147
7.2.2 Proteomics sample preparation	147
7.2.3 Proteomics LC-MS/MS analysis	148
7.2.4 Protein identification and quantification.....	149
7.2.5 Functional enrichment analysis.....	149
7.2.6 Untargeted metabolic profiling.....	150
7.2.7 RNA-seq data analysis.....	151
7.2.8 Reactive oxygen species (ROS) and H ₂ O ₂ measurement	151
7.3 Results.....	152
7.3.1 C3H/APX knockdown leads to substantial alterations in the Brachypodium stem proteome	152
7.3.2 C3H/APX deficiency impacts enzymes associated with the phenylpropanoid biosynthesis pathway	155
7.3.3 C3H/APX knockdown altered abundance of transcriptional regulators involved in the phenylpropanoid biosynthesis pathway	156
7.3.4 C3H/APX suppression alters the abundance of the proteins associated with phytohormone signaling and senescence.....	159
7.3.5 C3H/APX deficient Brachypodium exhibit disrupted redox homeostasis, photosynthesis, and DNA repair	162
7.3.6 C3H/APX deficiency impacts cellular kinases implicated in plant defense.....	165
7.3.7 Transcriptomics, metabolomics, and ROS levels support proteomics findings	166
7.4 Discussion.....	170

CHAPTER 8: OUTLOOK AND CONCLUSIONS	178
8.1 Overview of this dissertation	178
8.2 Remaining challenges and future considerations.....	182
8.3 Final thoughts	185
LIST OF REFERENCES	187
VITA.....	230

LIST OF TABLES

Table 2.1: Typical figures of merit for five common mass analyzers.	45
Table 2.2: Some useful tools and software for proteomics data analysis and data interpretation.	61

LIST OF FIGURES

Figure 1.1: The increasing complexity from genome to proteome.....	3
Figure 1.2: Two fundamental strategies for protein identification using mass spectrometry.....	5
Figure 1.3: Plants' biotic environment and their complex interactions.....	13
Figure 1.4: Conversion of bioenergy plant biomass to bioproducts.....	20
Figure 1.5: Lignin biosynthesis pathway.....	21
Figure 2.1: Typical workflow of bottom-up proteomics and post-translational modification analysis experiment.....	27
Figure 2.2: Example of a typical Multidimensional Protein Identification Technology (MudPIT) setup.....	36
Figure 2.3: A schematic illustration of the electrospray ionization process.....	40
Figure 2.4: Formula and different methods of calculating mass resolving power.....	42
Figure 2.5: Detail schematic of the Thermo Scientific Q Exactive Plus mass spectrometer.....	46
Figure 2.6: A systematic alphabetic code and potential cleavage sites for a peptide sequence. ..	48
Figure 2.7: Three widely utilized peptide identification strategies.....	52
Figure 2.8: Tandem mass spectrum illustrating a specific mass shift caused by phosphorylation modification.....	56
Figure 3.1: Schematic diagram of study design.....	72
Figure 3.2: Assessment of microbial population size.....	73
Figure 3.3: Metaproteomics analysis of <i>Pseudomonas</i> sp. GM17 functional behavior during community assembly.....	77
Figure 3.4: Metaproteomics analysis of <i>Pantoea</i> sp. YR343 functional behavior during community assembly.....	80
Figure 3.5: Metaproteomics analysis of <i>Bacillus</i> sp. BC15 functional behavior during community assembly.....	83
Figure 3.6: Dynamic regulation of methylation modification in lysine (K) residue of EF-Tu in nutrient-deprived organisms.....	87
Figure 4.1: Experimental design and qualitative assessment of phosphoproteomics result from the root samples inoculated with <i>L. bicolor</i>	101
Figure 4.2: Differentially abundant phosphopeptide analysis of transgenic and WT root sample.	103
Figure 4.3: Pathway analysis of differentially abundant phosphoprotein isoforms.....	106
Figure 4.4: Adjustment of ammonium transporter via phosphorylation and phosphorylation profile of ammonium transporter.....	109
Figure 4.5: Network analysis of differentially abundant phosphoprotein isoforms.....	111
Figure 4.6: Bimolecular fluorescence complementation (BiFC) assay validates the interacting partner of PtLecRLK1.....	113
Figure 4.7: Phenotypic characterization of <i>L. bicolor</i> colonization in roots of wild-type and transgenic poplar and assessment of phosphoproteins from poplar samples.....	114
Figure 4.8: Proposed signaling cascade of PtLecRLK1.....	119

Figure 5.1: Current model of HEHEAA-PipR regulated proteins with the abundance profiles of few proteins identified in proteomics measurement.	125
Figure 5.2: Differentially abundant proteins and enriched KEGG pathways.	126
Figure 5.3: Model representing the types of motilities and abundance profile of chemotaxis-related proteins identified.	128
Figure 5.4: Impact of HEHEAA in <i>Pseudomonas</i> sp. GM79.	132
Figure 6.1: Qualitative assessment of poplar leaf proteome across a 12/12-hour light/dark period.	137
Figure 6.2: Quantitative assessment of poplar leaf proteome abundance profile changes across a diurnal period.	138
Figure 6.3: Overview of poplar leaf protein modifications observed across a diurnal period....	140
Figure 7.1: Phenotypic characterization and Brachypodium C3H/APX protein abundance profile across experimental conditions.	153
Figure 7.2: Global assessment of proteomics measurements.	154
Figure 7.3: Impact of C3H/APX knockdown in lignin biosynthesis pathway.	157
Figure 7.4: Protein abundance profile of transcriptional regulators.	160
Figure 7.5: C3H/APX-deficient Brachypodium exhibit altered redox homeostasis and DNA damage response.	163
Figure 7.6: C3H/APX-deficient Brachypodium exhibit elevated levels of disease resistance proteins.	167
Figure 7.7: Reactive oxygen species (ROS) and H ₂ O ₂ levels in Brachypodium C3H/APX-KD lines.	169
Figure 7.8: Simplified model linking C3H/APX and the lignin biosynthesis pathway to other biological processes.	176

CHAPTER 1: INTRODUCTION TO PROTEOMICS AND POST-TRANSLATIONAL MODIFICATIONS (PTMS) AND RESEARCH TOPICS COVERED IN THIS DISSERTATION

1.1 Overview and application of mass spectrometry-based proteomics to understand system-level biology

Since the invention by Sir Joseph John (J.J.) Thomson in the early 1900s, mass spectrometry (MS) has undergone numerous improvements (Griffiths, 2008). The development of gas chromatography coupled to mass spectrometry in the late 1950s revolutionized the analysis of volatile compounds (Gohlke, 1959). Similarly, discoveries of electrospray and matrix-assisted laser desorption ionization in the late 1980s have revolutionized the ability to study non-volatile compounds such as peptides, lipids, oligosaccharide, etc. (Karas et al., 1987, Karas and Hillenkamp, 1988, Fenn et al., 1989). The improvements in mass spectrometry over time have allowed us to analyze compounds with molecular masses exceeding several hundred thousand Daltons, such as synthetic polymers, proteins, glycans, and polynucleotides (Griffiths, 2008). Another revolution occurred more recently with the invention of a new mass analyzer called orbitrap which improved the accuracy, resolution, sensitivity, and detection limits (Hardman and Makarov, 2003, Hu et al., 2005, Scigelova and Makarov, 2006). Following this, several developments have already been made and will continue to expand the capabilities and applications. The most impressive ones have occurred in the field of systems biology, particularly in proteomics.

Systems biology represents a paradigm shift from traditional biological studies. It aims to understand the biological system holistically by studying the combined behavior of the different biological molecules that create networks of interactions. Systems biology requires the development of new tools and protocols that can enable the ability to address fundamental questions that would have otherwise been out of reach such as how change in one or small set of biological molecules causes the system-wide changes. Like next-generation genome sequencing,

MS is now at the heart of countless applications allowing scientists to investigate a very broad range of biological questions. Together with massively parallel genome sequencing tools, MS has fostered a high-throughput data generation that has been a foundation for the field of -omics. Although all available omics-related tools and analyses are important for the study of systems biology, this dissertation will mainly focus on mass spectrometry-based proteomics to understand systems biology.

The emergence of genome sequencing and DNA microarray technologies in the 1990s gave rise to the era of -omics research. Genomics studies the set of genes contained in chromosomes. Transcriptomics studies the set of genes that are expressed in a system at a given time under specific conditions. Proteomics and metabolomics studies the set of proteins and metabolites expressed in a system at a given time under specific conditions. Following the advancement in genome sequencing and transcriptional profiling, proteomics was the logical continuation (Wilkins et al., 1996, Karpievitch et al., 2010). This field of proteomics is concerned with studying an organism's proteins encoded by its genome and understanding their roles as a part of the larger networked system (**Figure 1.1**) (Karpievitch et al., 2010). Today, the field looks to understand protein's roles not just within the context of the organism from which they are derived, but how they impact other biological systems as well as their surroundings, and the larger ecosystem. This represents a vital piece of systems biology, which aims to characterize the system behavior rather than the behavior of a single component. Proteins are the cellular machineries that contribute to the molecular functions and phenotype of the biological system (Bunnik and Le Roch, 2013). According to the central dogma of molecular biology, DNA is transcribed to RNA and RNA is translated to protein (**Figure 1.1**). However, not all genes from the genome or DNA are expressed at once. Therefore, genomics study can provide the functional potential of an organism but not the current functional state. RNA or transcriptome can be used to assess the functional state of an organism however, levels of messenger RNA (mRNA) involved in protein synthesis don't always correspond to the protein level in a cell (Bunnik and Le Roch, 2013). This is due to the fact that not all mRNA results in a protein product (Shyu et al., 2008). Some mRNAs are degraded before translation and for the mRNAs that get translated, proteins produced can still undergo many levels of regulation such as post-translational modifications (PTMs) (**Figure 1.1**). Therefore, proteome

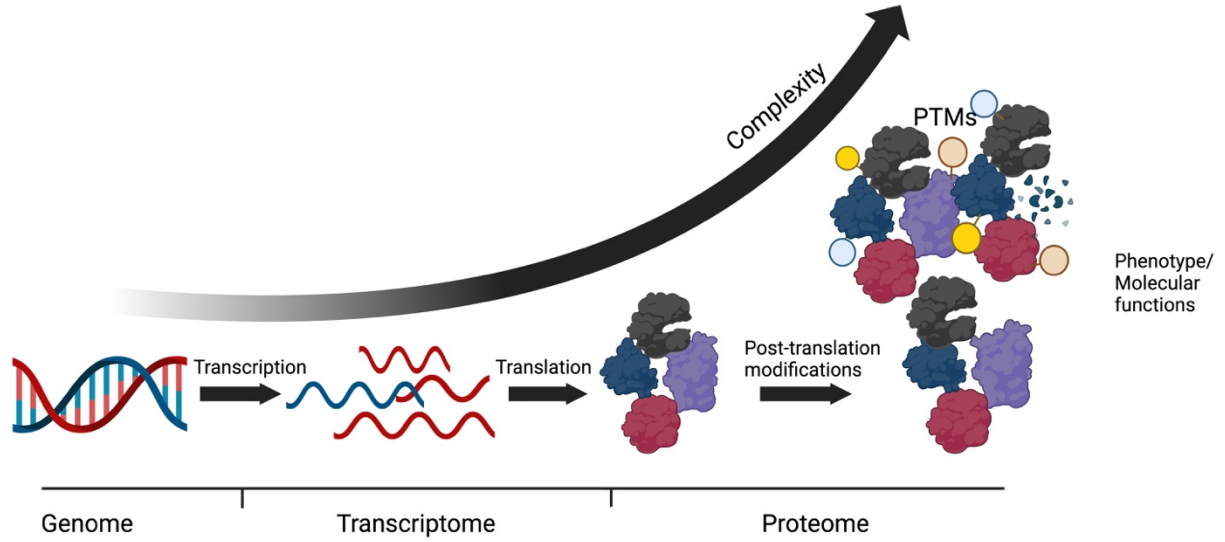


Figure 1.1: The increasing complexity from genome to proteome.

A genome can be transcribed, the resulting RNA can undergo alternative splicing and mRNA editing that increases the complexity of the transcriptome. Transcripts can be translated to proteins which can further be modified increasing the overall complexity at the exponential level. These PTMs can be due to the covalent linkage of a certain functional group or due to the cleavage of proteins into a smaller mature form.

characterization provides information more closely related to biological functions and phenotype than the genome or the transcriptome.

As in any other field, the field of proteomics contains a range of methodologies that are driven by technology advancements. Although the concept of proteomics was proposed over half a century ago, proteomics-related research did not initiate till the mid-1990s (Anderson et al., 2001). The initial advancement in proteomics research was due to advancement in a few key areas i.e. i) robust separation of proteins using two-dimensional gel electrophoresis technique (Sperling, 2001), ii) advancement of MS methods for the analysis of accurate mass and chemical structure (Loo et al., 1999, Aicher et al., 1998), iii) substantial effort on large-scale genome research that helped to generate high-quality databases (Zhang et al., 2002), and iv) progress in developing bioinformatics tools that helped to analyze large volumes of MS data (Nefedov et al., 2011). Protein analysis using mass-spectrometry focuses mainly on two fundamental strategies of protein identification and characterization: the top-down approach and the bottom-up approach (**Figure 1.2**) (Chait, 2006). In the top-down approach, the intact protein ions are subjected to direct analysis by MS (**Figure 1.2**) (Chait, 2006). Mass and fragment ions information that are generated from MS analysis is then used for protein characterization. On the other hand, bottom-up protein analysis (also commonly referred to as shotgun proteomics) starts with protein digestion using chemicals or enzymes and the digested proteins (peptides) are then analyzed by MS (**Figure 1.2**) (Chait, 2006). The information obtained on the peptides are matched to the proteins or protein mixture to deduce information about the proteins.

In general, typical bottom-up proteomics can identify anywhere from thousands to tens of thousands of proteins from an organism or biological system. When comparing bottom-up results to the top-down, top-down proteomics can only identify a fraction of that with a similar type of sample but requires substantial effort to simplify the protein mixtures (Tao et al., 2019). This is because bottom-up proteomics consists of peptide samples with relatively uniform length, chemistry, and a relatively smaller range of charge state and molecular weight whereas the intact protein analyzed by the top-down method can have large variation in terms of charge state and molecular weight, molecular structure, and chemical properties like hydrophobicity (Tao et al., 2019). The inherent advantages of bottom-up proteomics include easier and better separation with

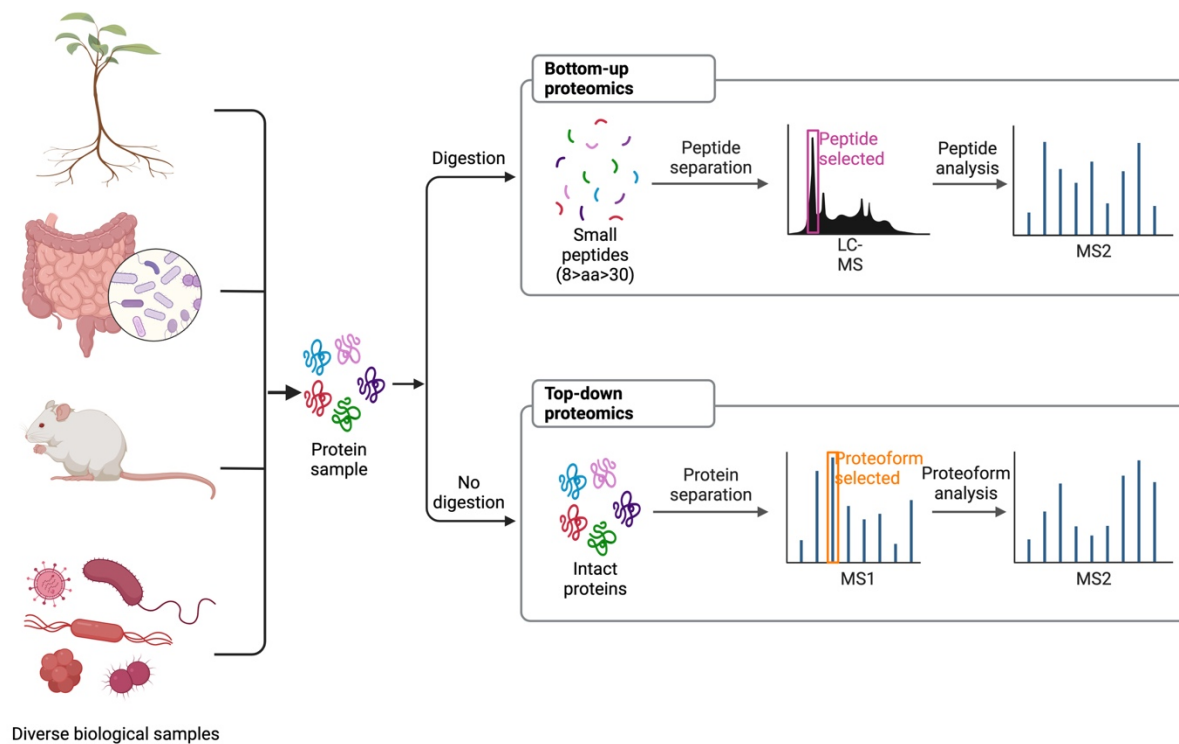


Figure 1.2: Two fundamental strategies for protein identification using mass spectrometry.

In the bottom-up approach (top), proteins extracted from diverse biological samples are first digested using a protease like trypsin. The peptides generated from the protein digestion are separated by liquid chromatography and analyzed using mass spectrometry. In the top-down approach (bottom), intact proteins extracted from diverse biological samples can also be separated and then analyzed by the mass spectrometer. In recent years, another approach called the middle-down approach has also been introduced which is similar to bottom-up proteomics, but it aims to digest proteins using a special protease to generate longer peptides (not shown in the figure).

liquid-chromatography (LC), relatively more uniform ionization, better fragmentation in MS, and easier interpretation of fragmentation patterns compared to top-down proteomics making bottom-up proteomics a preferred choice for mass spectrometry-based protein analysis. Nonetheless, top-down proteomics can still be valuable especially when the goal of the research is to gain insight into the structural characteristic of the proteins or to obtain the full characterization of proteoforms (Fornelli et al., 2018). Although bottom-up proteomics has several advantages, it is important to note that only a fraction of the peptides generated from protein digestion are normally detected in MS, which could create ambiguities thus, getting a protein inference from the fraction of peptides can be like a puzzle with many pieces missing (Tao et al., 2019). With the hope of overcoming the challenges and pitfalls associated with bottom-up and top-down approaches, in recent years, there has been an introduction of another approach called the middle-down approach. Like bottom-up proteomics, in the middle-down approach the proteins are digested but rather in a restrictive manner using special proteases so that longer peptides can be generated compared to the ones generated from bottom-up proteomics for enhanced sequence coverage (Pandeswari and Sabareesh, 2018). While this dissertation acknowledges the benefits and drawbacks of utilizing both top-down and bottom-up approaches and the availability of new approaches like middle-down, it will only focus on biological research using the bottom-up approach.

The bottom-up approach of proteome analysis has progressed from qualitatively cataloging thousands of proteins in numerous systems to revealing the quantitative changes in cellular and molecular functions in normal and perturbed/modified conditions. Moreover, MS can now be used to identify proteoform, different molecular isoforms of the proteins from a single gene resulting due to post-transcriptional and post-translational modifications, that extends beyond what's encoded in the genome (Smith et al., 2013). Rapid advancement in the MS instrumentation allows for accurate and sensitive measurement of proteoforms from complex biological matrices such as the human gut, fecal material, rhizosphere, etc. Further detail on the bottom-up proteomics approach will be discussed in chapter 2.

1.2 Need for post-translational modifications perspective in biological research

Post-translational modifications (PTM) are most often an enzymatic-derived modification of proteins after it is synthesized (**Figure 1.1**). PTMs can be highly dynamic and well-regulated and thus represent a major level of cellular and molecular regulation. Throughout a protein's life cycle, it may undergo numerous modifications that affect its assembly, localization, function, stability, degradation, etc. (Millar et al., 2019). PTMs can occur rapidly and be reversible such as phosphorylation, which can take place in seconds, or they can be slow and irreversible (or fixed), such as some forms of glycosylation (Urban, 2022, Venne et al., 2014). The modifications can occur either on the amino acid side chains or at C- or N- terminus of the proteins (Millar et al., 2019). For example, linking sugar or oligosaccharide on the side chain of amino acids during glycosylation can have a profound impact on protein solubility, protein-protein interaction, folding, and signaling. Similarly, modifications such as post-translational cleavage can result in new C and N terminus of the protein, which could provide different functions and stability to the modified protein. Commonly studied PTMs such as glycosylation, phosphorylation, oxidation, and SUMOylation modify the existing functional group of the amino acid and thus extend the diversity and properties of the 20 standard amino acids (Millar et al., 2019). While these are commonly studied PTMs, it is important to note that there are more than 300 different types of PTMs identified (Creasy and Cottrell, 2004, Han et al., 2011, Duan and Walther, 2015). The identification and quantification of modified isoforms of proteins are crucial for a complete understanding of cellular processes in a biological system.

Phosphorylation is one of the, if not the, most well-studied post-translational modifications. Protein phosphorylation is highly dynamic and can modulate cell signaling, protein assembly, localization, function, stability, and protein-protein interactions (Ardito et al., 2017, Vyse et al., 2017, Kim et al., 2004). This modification is catalyzed by the kinases that transfer a phosphate group from adenosine triphosphate (ATP) to targeted amino acids, mainly serine, threonine, and tyrosine (Urban, 2022). The phosphorylated residue is recognized and bound by a specific domain of the proteins that help to transduce signals downstream (Urban, 2022). In a similar context, the phosphatase family of proteins can remove the phosphorylation modification from proteins (Low et al., 2021, Urban, 2022). The dynamic function of the kinases and phosphatases are responsible

for generating a reversible phosphorylation modification. Through this dynamic modification, various downstream genes and signaling pathways can be regulated in a highly sensitive and specific manner. Phosphorylation modification and its importance in regulating plant-fungal interaction are further discussed in chapter 03.

1.3 The challenges with PTM studies and recent advancements in the identification of PTMs

Although the identification and quantification of PTMs are essential for a comprehensive understanding of cellular function, holistically analyzing all the PTMs isoforms is a challenging task. Unlike genes and proteins, PTMs cannot be predicted because they are not encoded by the genome (den Ridder et al., 2020). As such, deeper DNA sequencing to improve the reference database has little to no effect on PTM identification. Moreover, PTMs can have low stoichiometry due to having different modification sites with different modification efficiency, low abundance, and poor ionization efficiency, all of which further add complexity to their identifications. Similarly, the database searching methods which are utilized to identify and quantify the peptide sequence from the mass spectra are restrictive for the PTM search. This is mainly because the common database searching tools currently utilized for proteomics research require prior information on the PTMs present in the sample. If this parameter is not defined for the purpose of global PTM analysis and all the potential modifications are considered, computational search space increases exponentially, which challenges the computational and statistical resources (Ahrne et al., 2010, Na et al., 2012, Ma and Lam, 2014).

Besides the challenges associated with global PTM profiling, there are still challenges associated with the study of single modification-focused studies like phosphorylation. Approximately 30% of proteins are predicted to undergo phosphorylation modification (Cohen, 2000). It is of utmost importance to develop proper analytical methods that allow the accurate and sensitive characterization of phosphorylated proteins (Urban, 2022). However, due to the low stoichiometry of regulatory phosphorylation sites, the high complexity of the phosphoproteome, the low abundance of phosphorylation protein isoform, and the poor ionization efficiency of

phosphopeptides achieving sufficient coverage to cover all phosphorylation sites of interest is a challenging task.

To improve our ability to profile all available modifications in biological samples, several algorithms that allow the open modification search have been developed (den Ridder et al., 2020). These algorithms don't require prior information about the modification and allow the protein search without specifying any modification, which aids in identifying unexpected and unknown modifications in biological samples. These alternative algorithms utilize different strategies for PTM profiling, which include *de-novo* sequencing, sequence-tagging, or multi-round search. PEAKS software is one of the popular vendor-based software that integrates *de-novo* sequencing and multi-round search using the PeaksPTM to enable the identification of all modifications from the Unimod database (Han et al., 2011). Unimod database is a comprehensive community-supported protein modification database for mass spectrometry applications that contains information on more than 300 different PTMs (Creasy and Cottrell, 2004).

For a detailed understanding of a single modification like phosphorylation, several strategies can be implemented to circumvent the challenges. The most utilized technique is phosphopeptide enrichment. Phosphopeptide enrichment techniques are indispensable for system-wide phosphoproteome studies. Some of the commonly adopted methods of phospho-enrichment include (i) affinity-based enrichment (discussed in detail in chapter 02), ii) ion exchange chromatography, and (iii) phospho-specific antibodies (Low et al., 2021). The selective enrichment of phosphopeptides and phosphoproteomics along with high-resolution mass spectrometry have shown to provide unprecedented characterization and quantification ability of phosphorylated proteins and peptides in complex samples which aid in mapping and quantifying intracellular kinase signaling networks (Vyse et al., 2017, Urban, 2022, Riley and Coon, 2016, Olsen and Mann, 2013)

In addition to global proteomics, getting either a holistic perspective of all potential PTMs or a detailed perspective on one type of PTM from a biological system can provide additional insights into the biological processes and regulatory mechanisms of that biological system. Recent advancements in experimental as well as bioinformatic workflows to study PTMs in biological

systems provide an exciting opportunity to understand the important level of regulation that adds further complexity to the central dogma of information flow (Li et al., 2014, Marlow et al., 2016, Zhang et al., 2016, Gupta et al., 2007, Brown et al., 2017). In this dissertation, mass spectrometry-based global proteomics will be utilized along with optimized experimental and bioinformatic methods for both holistic characterization of all potential PTMs as well as detailed characterization of one type of PTM (phosphorylation) in order to gain a deeper molecular understanding on the bioenergy plants with the broader goal of improving bioenergy research.

1.4 Introduction to the bioenergy plant research and its importance

Worldwide consumption of energy has increased rapidly (tripled since the 1960s), which is faster than the rate of population increase (Hein, 2005). Due to this, there has been an increasing concern about how such high energy demand can be regularly met by the finite, and slowly depleting non-renewable supply of fossil fuels (Karp and Shield, 2008). Moreover, with increasing energy demand, fear about greenhouse gas emissions and climate change has also been heightened (Karp and Shield, 2008). Undoubtedly, there is a need for an alternative energy resource. Bioenergy generated from plant biomass can be an alternative energy resource that has the potential to be a significant part of the global energy economy.

Bioenergy research focuses on the use of perennial grasses and trees to produce biofuels like ethanol, biodiesel, hydrocarbon, and high-valued bioproducts like glycerin (Karp and Shield, 2008). Currently, bioenergy accounts for around 10% total energy supply and provides four-fold more energy than solar and wind combined (Reid et al., 2020). Biofuel production for transportation use has been consistently increasing in the last couple of decades and is projected to increase further. Between 2018 to 2023, bioenergy is expected to make up around 30% of the total growth in renewable energy (Reid et al., 2020). The high yield of bioenergy from plant biomass could substantially help to alleviate the global concern of energy security, climate change, and environmental pollution (Karp and Shield, 2008). It would also reduce the use of toxic or harmful chemicals for energy production, create more jobs, and improve the trade balance.

The sustainable high yield of bioenergy relies mainly on two factors i.e., improved growth and distribution of the bioenergy crops and efficient conversion of biomass feedstock to biofuel and high valued bioproducts. Therefore, substantial research has been focused on improving plant fitness and growth as well as improving the biomass conversion to biofuels and bioproducts. This dissertation will mainly focus on utilizing high-performance mass spectrometry and advanced bioinformatic techniques to gain system-wide insight into bioenergy plants and how the environment impacts the plant traits to benefit the overall goal of a sustainable high yield of bioenergy. The efficient plant growth relies on improving the assimilation, recycling, and remobilizing of carbon and nitrogen, water uptake, stress tolerance, and disease resistance (Karp and Shield, 2008). These factors can be impacted by either biotic (e.g., fungi, bacteria, nematodes) or abiotic (e.g., drought, light, heat) environment in or around which the plant is grown. Similarly, lignin modification has been a key strategy for reducing biomass recalcitrance. For bioenergy research, Poplar species of plants have been one of the most important model organisms due to its wide distribution in North America, rapid growth, easy propagation, adaptability to a range of planting sites, and range of microbiome it hosts (Bryant et al., 2020). Nonetheless, other plants like *Brachypodium distachyon* (purple false brome) and *Panicum virgatum* (switchgrass) are also utilized as model plants for perennial grasses.

1.4.1 Understanding bioenergy plants and their environment for improved fitness and growth of bioenergy plants

Improved plant biomass can directly benefit the overall goal of a sustainable high yield of bioenergy. While studying plants to improve their growth and fitness, one must also consider their environment and how the change in environment impacts overall growth and fitness. Plants exist in a soil–plant–atmosphere continuum, and they constantly coordinate various mechanisms simultaneously to respond to the changes in the environment at any given time for optimal growth, fitness, and survival (Zhou and Shao, 2008). Abiotic factors like water have been proven to influence plant fitness. Water deficiency in the plants' environment can substantially reduce the plants' total yield (Zhou and Shao, 2008). Similarly, biotic factors equally impact plant growth and development. For example, the vast majority of plants host fungi in their close vicinity. Some fungi can form a netlike structure called mycorrhizae which benefit the plant through nutrient

acquisition and disease resistance (Blasius et al., 1986, Labbe et al., 2019, Roth-Bejerano et al., 1990, Buscot et al., 2000). Also, some fungi can be parasitic and lead to the demise of the plant.

1.4.1.1 Plants' biotic environment

Biotic factors like microbes are abundantly present across all environments. In or outside plants, niche-level microbiomes exist that significantly differ across habitats such as root, stem, leaf, or soil (Cregger et al., 2018). Regardless of their niche, plants association with biotic factors are closely associated with key functions, including water and nutrient acquisition, stress response, and host defenses (Pii, 2015, De Zelicourt, 2013, Berendsen et al., 2012, Pineda et al., 2015). Plant-microbe interactions can occur above ground or in the aerial region (phyllosphere), below ground in the root vicinity (rhizosphere), and within plant tissue (endosphere) (**Figure 1.3**). While the study of both aboveground and belowground microbes are equally important to get a holistic perspective of a plant's biotic environment, the rhizosphere region is considered to be a microbiome hotspot that represents Earth's one of the most complex microbial ecosystems (Hinsinger and Marschner, 2006, Pierret et al., 2007, Jones and Hinsinger, 2008, Raaijmakers et al., 2009). The rhizosphere is a narrow nutrient-rich region or environment around the roots that is influenced and regulated by the plant itself (Zhang et al., 2017). While the importance of studying different plant niches for a holistic understanding of plants' biotic environment is clear, this dissertation will mainly focus on belowground microbes or rhizosphere microbiomes (rhizomicrobiome) since the majority of biotic interactions occur below ground as stated above.

The rhizomicrobiome influences the plants' health and fitness and enables it to adapt to the environment (Venturi and Keel, 2016). The effect of rhizomicrobiome relies heavily on chemical exudates released by both plants and microbes. These chemical exudates can be diverse molecules ranging from primary metabolites such as sugars, amino acids, and organic acids to secondary metabolites such as terpenoids, alkaloids, and phenolics (Venturi and Keel, 2016, Canarini et al., 2019). To what extent these rhizosphere chemical exudates affect the microbial interactions, community structure, and function are the focus of several ongoing research projects. Most of the studies utilize metabolomics for the identification and characterization of rhizosphere chemical exudates (van Dam and Bouwmeester, 2016) and amplicon-based studies like 16S rRNA gene

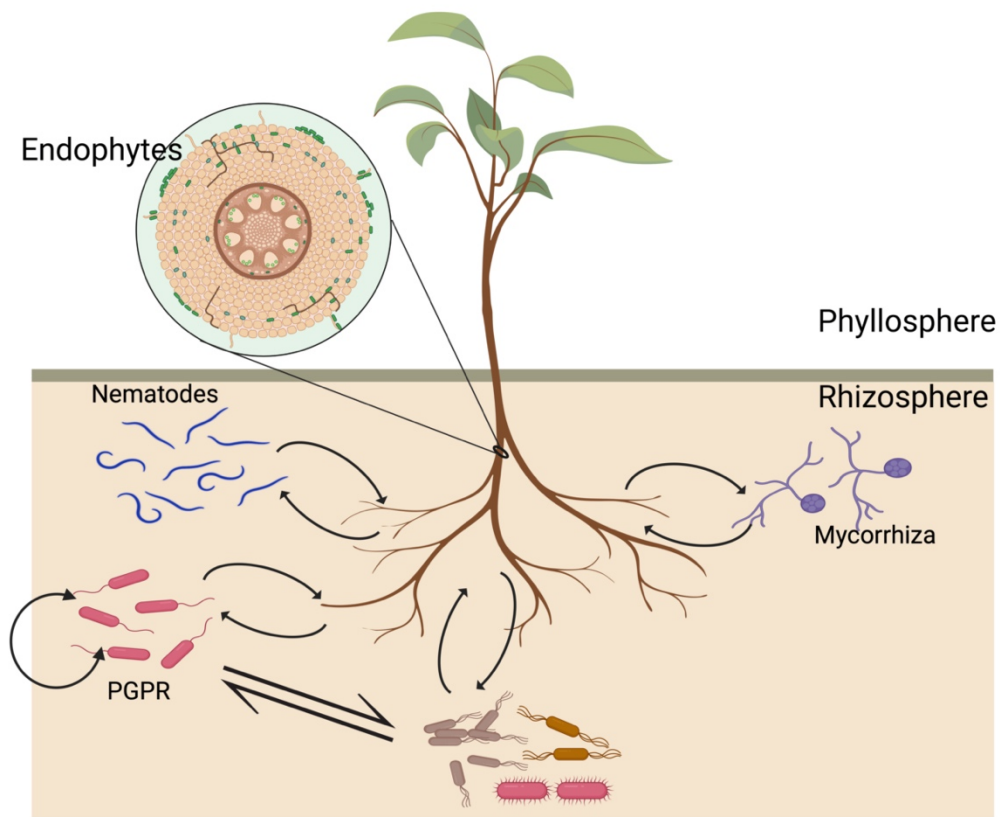


Figure 1.3: Plants' biotic environment and their complex interactions.

A plant can interact with microbes belowground (Rhizosphere), aboveground (Phyllosphere), or within the plant tissue (Endophytes). Among these niches, the rhizosphere is the microbiome hotspot that is often considered Earth's most complex microbial ecosystem. While the plant can interact with microbes that shape the biotic environment of the plant, microbes can also interact with each other which plays an equally important role in microbiome assembly and shaping the microbial community.

amplicon sequencing for assessing microbial diversity (Regalado et al., 2020). Although amplicon studies are useful for estimating microbial diversity, they don't provide information regarding the functional potential and activity of the microbiome. Metatranscriptomics has been the most widely adopted technology to assess the functional responses of both plants and microbes during plant-microbe interactions (Mark et al., 2005, Spaepen et al., 2014, Santos-Medellin et al., 2022, Nerva et al., 2022) and interaction with other environmental factors (Yamazaki et al., 2021, Zhang et al., 2020c, Gonzalez et al., 2018). However, as stated earlier in this chapter, changes in mRNA abundances do not necessarily correspond to changes in protein levels. Therefore, metaproteomics is recognized as a necessary measurement for characterizing changes in proteins, but also serves as an ideal technique for the accurate estimation of biomass from viable and functioning cells (Kleiner et al., 2017, Kleiner, 2019). It can also allow for the characterization of microbial membership, their functional roles in plants' biotic environment, and responses to interspecies interactions (Kleiner, 2019). However, metaproteomics hasn't been well implemented for the study of plants' biotic environments. This dissertation will utilize mass spectrometry-based metaproteomics and PTM analysis to study bioenergy plants' biotic environments. The interactions in the rhizosphere can be categorized into two main categories: microbe-microbes and plant-microbe. The plant-microbe interaction can be studied from the plant's perspective or the microbe's perspective, each providing a unique insight into their response to the interaction.

1.4.1.1.1 Plants' biotic environment: Microbe-microbe interactions

Plant-associated rhizospheric microorganisms can produce a wide variety of quorum sensing (QS) signaling compounds that help to regulate their gene expression as a response to environmental conditions, mainly cell density (Fuqua et al., 2001, Bakker et al., 2013). Through these chemical compounds, plant-associated microbes can regulate biological processes such as virulence factors production, secondary metabolites synthesis, biofilm formation, and motility (Fuqua et al., 2001). Microbes can also produce secondary metabolites (e.g. antibiotics) that can function as intra- and interspecies signaling molecules as well as the crucial factor regulating the final shape and structure of the stable microbial community (Venturi and Keel, 2016). Among the types of secondary metabolites, many volatile organic compounds (VOC) produced by microbes can act as chemical weapons to kill other microbes in the environment (Schmidt et al., 2015). Moreover,

VOCs can also act as intra/interspecies signals that can impact the overall rhizomicrobiome by modulating the gene expression related to microbial processes, such as biofilm formation and stress tolerance (Venturi and Keel, 2016). Although the characterization of signaling compounds during plant-associated microbe-microbe interaction has been the focus of several studies, studies that focus on understanding how the inter-microbial interactions shape microbial assembly and what molecular mechanisms are used by microbes to cooperate, compete, and persist in the complex microbial community is currently lacking.

A deeper understanding of inter-microbial interaction is crucial because it can not only be a contributing factor to microbial community composition and function but also dictates plant health and development. Mass spectrometry-based metaproteomics can be a suitable approach for plant-associated microbial interaction studies because, as mentioned earlier, metaproteomic provides an accurate estimation of biomass from viable and functioning microbial cells which can be used for the characterization of microbial membership (Kleiner et al., 2017, Kleiner, 2019). Moreover, metaproteomics can be utilized to reveal the mechanism of how microbes cooperate, compete, and persist during community assembly (Kleiner, 2019). Further exploiting the metaproteomics data using advanced bioinformatic tools to globally assess post-translational modifications (PTMs) in proteins provides unique insights into a largely unexplored level of microbial regulation/adaptation (Han et al., 2011). Therefore, implementing metaproteomics and PTM analysis for the study of bioenergy plant poplar-associated microbial interactions will be the focus of chapter 03. To the best of the author's knowledge, this is the first study to implement metaproteomics and PTM analysis for the study of plant-associated microbial interactions.

1.4.1.1.2 Plants' biotic environment: Plant-microbe interactions (Plant's response)

In addition to perceiving and responding to the signals produced by themselves or other microorganisms, microbes can also exchange signals with the plant (Venturi and Keel, 2016). Through the diverse signaling molecules, microbes can influence the health and fitness of their plant host. Several studies have focused on studying plant's response to the beneficial microbe which can promote growth and development via the induction of plant defenses against biotic and abiotic stressors, and the acquisition of soil nutrients (Lugtenberg and Kamilova, 2009, Shores et

al., 2010). Some of the beneficial rhizospheric microbes studied in detail are mycorrhiza, rhizobia bacteria, plant growth-promoting rhizo-fungi (PGPF) (Pieterse et al., 2014, Backer et al., 2018, Gouda et al., 2018).

Common signaling molecules that microbe releases are nodulation (Nod) and mycorrhizal oligosaccharides or lipochitooligosaccharides (Myc) factors, which can suppress salicylate-dependent plant defense pathways and initiate symbiosis (Oldroyd, 2013, Zamioudis and Pieterse, 2012). Various mycorrhiza can also produce proteolytic cleavage products (PCPs) that act as mutualistic effectors promoting plant-fungal beneficial interaction. One example of PCPs from the ectomycorrhizal fungus *Laccaria bicolor* is MiSSP7 (Plett et al., 2011). MiSSP7 has a molecular mass of 7 kDa size, and is released from fungal hyphae upon root contact. This PCP enters the plant cell and localizes in the nucleus, where it interacts with a protein that alters jasmonic acid signaling and facilitates plant-fungal symbiosis (Plett et al., 2011). Similar PCPs have also been identified in other mycorrhizal fungi that can alter hormonal signaling (Kloppholz et al., 2011). Rhizospheric microbes from pseudomonads can also produce antimicrobial molecules that can elicit plant defense responses. 2,4-Diacetylphloroglucinol (DAPG) from pseudomonads has been shown to induce phytohormone-dependent systemic resistance against bacterial and fungal pathogens (Weller et al., 2012, Iavicoli et al., 2003). This microbial metabolite can also affect plant root development through an auxin-dependent signaling pathway (Brazelton et al., 2008). On a similar note, antibiotic pyocyanin which can function as an intra/inter-microbial species signal can also induce systemic resistance and regulate root development with a similar mechanism as DAPG (Ortiz-Castro et al., 2014, Mavrodi et al., 2013). Taken together, the microbe or microbe produced signaling molecules during plant-microbe interactions can influence various biological processes in plants. Uncovering how the plants perceive microbe or microbial signals and transduce them into the plant signaling pathways to generate a precise response remains a focus of several ongoing research projects.

An ectomycorrhizal fungus *Laccaria bicolor* colonizes the root of the bioenergy plant *Populus trichocarpa* which is mediated by a G-type lectin receptor-like kinase (PtLecRLK1) (Labbe et al., 2019). During the ectomycorrhizal interaction, minerals, water, and nutrients are exchanged between plants and fungi which benefits both species. Engineering the expression *PtLecRLK1* in

other non-host plants results in *L. bicolor* root colonization. However, how *PtLecRLK1* regulates ectomycorrhizal symbiosis is still unknown. A comprehensive understanding of how *PtLecRLK1* regulates ectomycorrhizal symbiosis allows researchers to engineer the systems that can enhance plant productivity and fitness for bioenergy production. *PtLecRLK1* being a kinase that regulates protein phosphorylation, mass spectrometry-based phosphoproteomics analysis would be an ideal approach to reveal the *PtLecRLK1*-mediated signal transduction. Deciphering the *PtLecRLK1*-mediated plant response during plant-microbe interaction will be the focus of chapter 04.

1.4.1.1.3 Plants' biotic environment: Plant-microbe interactions (Microbe's response)

Plants serve as hosts to numerous commensals, mutualists, or parasitic microbes in the rhizosphere. The phytochemicals secreted by the plants are crucial for the plant-microbe interactions and to shape the microbial community (Zhang et al., 2015). The best understood plant-microbe signaling and interaction from a microbial perspective is between legume plants and nitrogen-fixing rhizobial microbes (Venturi and Keel, 2016). During this interaction, the initial plant-microbe communication is mediated through the plant-produced flavonoid compounds such as 2-phenyl-1,4-benzopyrone derivatives (Redmond et al., 1986). The plant-produced flavonoid compounds can then induce the bacterial nodulation (Nod) genes that facilitate the production and release of lipo-chitooligosaccharides (LOCs), commonly known as Nod factors (Rossen et al., 1985). The perception of these Nod factors by plant receptor kinases allows the plant to initiate nodule formation, which then gets colonized by rhizobial bacteria (Oldroyd, 2013). These microbes modulate plant growth and development by fixing atmospheric nitrogen in nodules and transporting usable forms of nitrogen to plants. Another example of microbe response to plant-microbe interaction is through the strigolactone. This molecule is secreted from plant roots which acts as a signal to facilitate arbuscular symbiosis (Ruyter-Spira et al., 2013, Schmitz and Harrison, 2014). Strigolactones have been demonstrated to have various functions. It can help stimulate mycorrhizal hyphae and can also interfere with auxin transport (Akiyama et al., 2005, Waldie et al., 2014). Similar to strigolactones, cutin monomers are also known to act as a signaling molecule that aids in arbuscular symbiosis (Wang et al., 2012, Gobbato et al., 2012). Besides its role in initiating arbuscular symbiosis, cutin can also act as signaling metabolites in appressoria formation in pathogenic fungi (Liu et al., 2011).

QS by rhizospheric microbes can regulate various processes such as microbial competence, virulence, and the production of secondary metabolites. It can aid microbes to colonize plants and their environment. Some plant-derived metabolites have been shown to impact microbial QS by behaving as agonists or antagonists of QS systems (Bauer and Mathesius, 2004, Keshavan et al., 2005). More recently, many different plant-associated bacteria have been identified to possess QS receptors, LuxRs, that do not bind or respond to QS signals but rather to plant derived compounds (Gonzalez and Venturi, 2013, Subramoni et al., 2011). One of the plant-derived compounds identified from bioenergy plant poplar is ethanolamine derivative, N-(2-hydroxyethyl)-2-(2-hydroxyethylamino) acetamide (HEHEAA) (Coutinho et al., 2018). This plant-derived metabolite is a potent effector of LuxR-type protein called PipR in poplar root endophyte *Pseudomonas* sp. GM79. The mechanism by which HEHEAA regulates various biological processes and traits important for poplar colonization will be discussed further in chapter 05.

1.4.1.2 Plants' abiotic environment: Response to the diurnal cycle

Similar to plants' biotic environments, plants' abiotic environments are also complex and intertwined with their health and fitness. Plants are sessile and thus must constantly adapt to the changing abiotic factors such as water, nutrient availability, humidity, light, wind, CO₂, salt, pollutants, and temperature for optimal growth and development (Pereira, 2016). Plants sense and respond to abiotic factors through changes in metabolism and physiology (Des Marais et al., 2013). These phenotypic plasticity allow the plant to exhibit multiple phenotypes to acclimate to the given environment (Des Marais et al., 2013). Moreover, the plants' biotic environment and its interaction with them can also be shaped and impacted by the physical world that surrounds them. Any deficiencies in the abiotic factors crucial for survival such as water, nutrient, light, temperature, etc. can severely restrict growth. The availability of abiotic factors can determine the types of organisms that can exist in that environmental setting, and how abundant they may become. Most of the abiotic factors are dynamic and vary throughout the day, seasons, or years. Circadian rhythms are known to synchronize plant biological processes like gene expression, photosynthesis, and flowering to external variation and play a key role in a plant's phenotypic plasticity (Kugan et al., 2021, McClung, 2006). The characterization of this rhythmic phenomenon at the proteome and PTMs level is currently lacking for a bioenergy model plant poplar. Identifying and characterizing

the proteins and PTMs that are temporally regulated in poplar using high-resolution mass spectrometry and advanced *de novo*-assisted data processing provides a more complete understanding of how plants respond to environmental cues and will be the focus of chapter 6.

1.4.2 Reducing the recalcitrance in bioenergy plants for the efficient biomass conversion

One of the major hurdles for the sustainable high yield of bioenergy has been to effectively convert lignocellulosic biomass to biofuel and high valued bioproducts. Lignocellulosic biomass is the most abundant sustainable feedstock mainly consisting of polysaccharides (i.e. cellulose and hemicelluloses) and an aromatic lignin polymer (Zoghalmi and Paes, 2019). Lignocellulosic biomass has the potential to be an economical substitute for traditional nonrenewable energy resources (Zoghalmi and Paes, 2019). However, effectively accessing the cellulosic sugars from the lignocellulosic biomass that can be converted to biofuels and other valuable bioproducts is hampered by the lignin (**Figure 1.4**). Lignin is a natural, branched phenolic heteropolymer abundantly present in the plant's secondary cell walls (**Figure 1.4**) (Boerjan et al., 2003, Ralph et al., 2004). It tethers hemicelluloses to cellulose in the plant cell wall and is the second most common biopolymer after cellulose, comprising ~30% of organic carbon (Boerjan et al., 2003, Ragauskas et al., 2014). It is crucial for structural and physical strength to support plant growth and development, water and solutes transportation, and plant defense against a range of biotic and abiotic stresses (Boerjan et al., 2003, Whetten and Sederoff, 1995, Zhao and Dixon, 2014, Barros et al., 2015, Hilal et al., 2004, Fan et al., 2006, Delessert et al., 2004, Malinovsky et al., 2014, Lee et al., 2019). Lignin either limits the access of enzymes to cellulose by physically blocking them or irreversibly binds to the enzymes due to its hydrophobic properties during enzymatic hydrolysis (Zoghalmi and Paes, 2019).

To overcome the recalcitrance of lignified cell wall, thermochemical pretreatment is required, which removes the lignin prior to the enzymatic hydrolysis of lignocellulosic biomass (Chen and Dixon, 2007, Li et al., 2016). However, thermochemical pretreatment is an energy-inefficient and time-consuming process. An efficient approach to controlling lignin formation in plants has been a major focus of bioenergy research in the last decades. To that end, genetic modification of lignin biosynthesis pathways (**Figure 1.5**) has been utilized to minimize lignin content and/or alter lignin

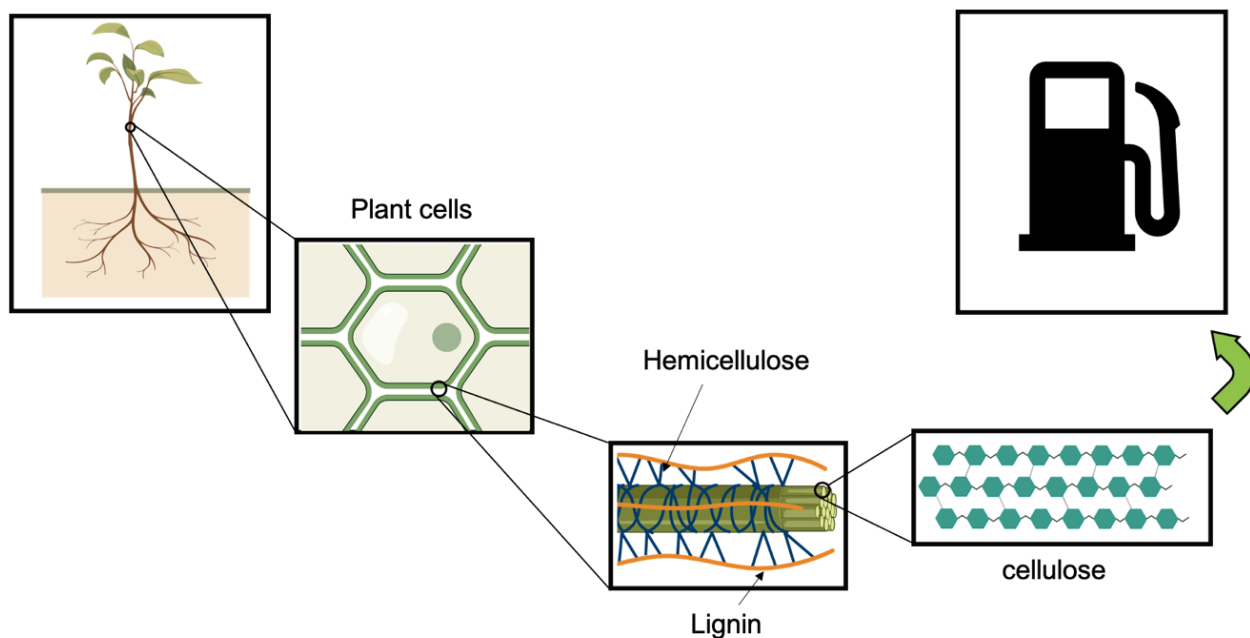
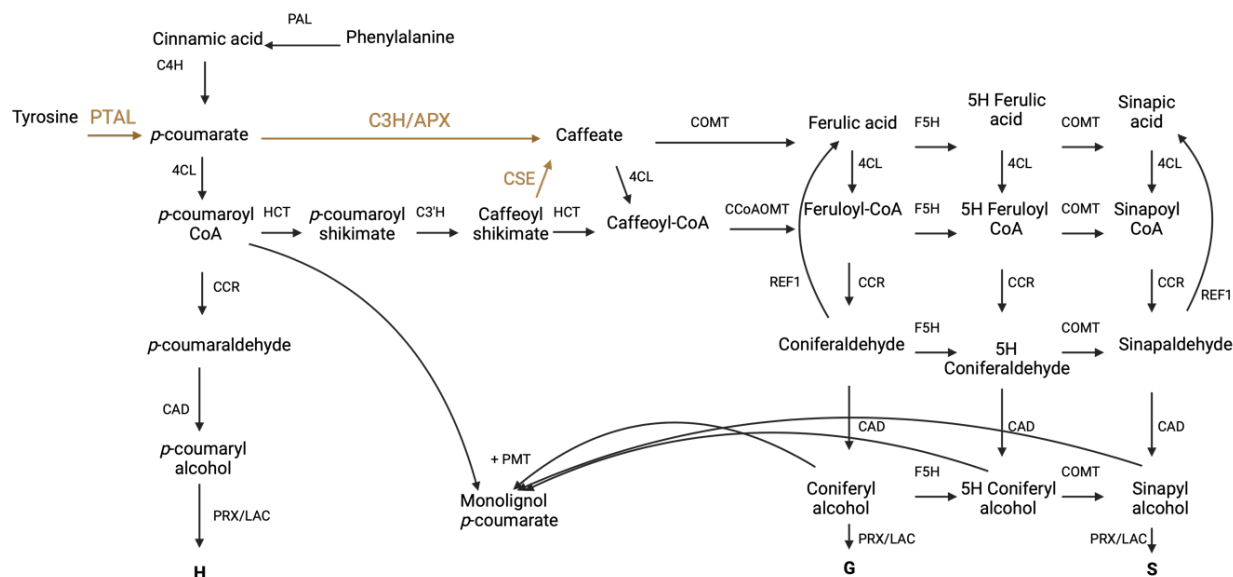


Figure 1.4: Conversion of bioenergy plant biomass to bioproducts.

Plant cell wall contains long chains of sugars (cellulose) that can be converted to fuels and chemicals. Cellulose, however, are embedded within a matrix of hemicellulose and lignin which provide a barrier for sugar conversion to biofuels. The recalcitrant properties of lignin have been one of the major hurdles to the sustainable high yield of bioenergy.



A scheme of the canonical lignin biosynthesis pathway in black color with grass-specific pathway highlighted in gold color. PAL: L-phenylalanine ammonia-lyase, C4H: cinnamate 4-hydroxylase, COMT: caffeate/5-hydroxyferulate 3-*O*-methyltransferase, F5H: ferulate 5-hydroxylase/coniferaldehyde 5-hydroxylase, 4CL: 4-hydroxycinnamate:CoA ligase, HCT: 4-hydroxycinnamoyl CoA:shikimate/quinic acid hydroxycinnamoyltransferase, C3'H: 4-coumaroyl shikimate/quinic acid 3'-hydroxylase, CSE: caffeoyl shikimate esterase, CCoAOMT: caffeoyl CoA 3-*O*-methyltransferase, CCR: cinnamoyl CoA reductase, CAD: cinnamyl alcohol dehydrogenase, PMT: *p*-coumarate monolignol transferase, LAC: laccase, PRX: Peroxidase, REF: reduced epidermal fluorescence, PTAL: bifunctional L-phenylalanine/L-tyrosine ammonia-lyase, C3H/APX: 4-coumarate 3-hydroxylase, APX: Ascorbate peroxidase, CHS: chalcone synthase.

substructures to relieve the lignin's recalcitrant properties for biofuel production (Li et al., 2016, Chen and Dixon, 2007, Carmona et al., 2015). The efforts to modify lignin biosynthesis to relieve the recalcitrant properties often introduce several pleiotropic effects, such as stunted plant growth and associated yield penalty (Ha et al., 2021, Muro-Villanueva et al., 2019, Vanholme et al., 2019). This has been a major bottleneck for exploiting biotechnology to modify lignin to reduce biomass recalcitrance. Unfortunately, the precise molecular mechanisms underlying the negative effect on growth and development in lignin-modified plants are not well understood. Future studies aiming at a system-wide level understanding of altered biological processes in lignin-modified plants are expected to help pinpoint the lignin-modification associated growth defects. Despite the underlying mechanism being elusive, in the last few years, several hypotheses have been postulated to explain the lignin modification-induced negative growth phenotype.

1.4.2.1 Current hypotheses for lignin modification induced negative growth (dwarf phenotype)

Lignin deposition in the stems confers mechanical strength. In the secondary cell walls of tracheary elements, lignin deposition provides the rigidity and hydrophobicity that allows the plant to endure the negative pressure of water conductance (Turner and Somerville, 1997, Muro-Villanueva et al., 2019). Some plants with a mutation in the lignin biosynthetic gene exhibited deformed walls in tracheary elements indicating vascular collapse (Franke et al., 2002, Schilmiller et al., 2009). Therefore, one hypothesis is that negative growth phenotypes are due to collapsed vessels which impact the plant's ability to transport water and solutes. In a recent study, however, cinnamoyl CoA reductase 1 (CCR1) and caffeoyl shikimate esterase (CSE) mutant barrel clover (*Medicago truncatula*) were found not to have collapsed vessels (Man Ha et al., 2019). CCR1 is a key enzyme in a later stage of lignin biosynthesis which converts cinnamoyl-CoAs into their corresponding cinnamaldehydes and CSE hydrolyzes caffeoyl shikimate intermediates in the earlier stage of lignin biosynthesis. These mutants were lignin deficient but were not water-stressed or exhibit drought-responsive gene expression which would be expected if the vascular collapse induced osmotic imbalance was a direct cause of stunted growth (Man Ha et al., 2019).

Research has suggested that negative growth could also be due to the hyperaccumulation of inhibitory molecules that are produced as lignin pathway intermediates (Bonawitz and Chapple, 2013, Bonawitz and Chapple, 2010, Muro-Villanueva et al., 2019). In *Arabidopsis*, the C4H mutation accumulates high levels of cinnamate-derived compounds which impairs the auxin transport and signaling that is crucial for plant growth and development (Schilmiller et al., 2009, Wong et al., 2005, Liu et al., 1993). Similarly, the CCR1 mutation in *Arabidopsis* is attributed to overaccumulation of ferulic acid which antagonizes the reactive oxygen species (ROS) burst needed to exit cell proliferation, leading to defective leaf development in *ccr1* (Xue et al., 2015). However, subsequent research has refuted the direct link between these molecules and lignin-modified dwarfism (Li et al., 2010, De Meester et al., 2018).

Alternatively, deficiencies in metabolites that directly/indirectly facilitate plant growth has been suggested to contribute to the dwarfism observed in plants with modified phenylpropanoid metabolism (Bonawitz and Chapple, 2013, Bonawitz and Chapple, 2010, Muro-Villanueva et al., 2019). In *Arabidopsis* double mutants of caffeoyl coenzyme A 3-O-methyltransferase 1 (CCoAOMTs) and caffeic acid O-methyltransferase 1 (COMT), negative effect on growth (dwarfism) can be attributed to dehydrodiconiferyl glucoside (DCG) deficiency which behaves like cytokinin and helps promote growth and cell division (Do et al., 2007, Binns et al., 1987).

Finally, the tradeoff between growth and defense could also lead to dwarfing. Plant cell wall integrity is crucial for cell development and defense against abiotic and biotic stressors. Plants possess a sensing mechanism that upon sensing alteration in cell wall integrity can elicit transcriptional changes that assist in overcoming damage (Hamann, 2015). Lignin modification alters the cell wall composition and integrity. If the plant detects the defect in cell wall integrity, it initiates a plant recovery response and activates the defense-related genes (Vaahtera et al., 2019). The production of defense compounds reallocates the resources needed for proper plant growth which could lead to dwarfing (Stamp, 2003).

Despite these prevalent hypotheses, many lignin-modified plants don't demonstrate the physiological and molecular changes as hypothesized. Thus, the precise molecular mechanisms underlying the negative effect on plant growth and development in low lignin mutants and

transgenics are still elusive. Recently, a bifunctional peroxidase enzyme, 4-coumarate 3-hydroxylase (C3H/APX), has been shown to provide a parallel route to the shikimate shunt pathway for the conversion of 4-coumarate to caffeate in the early steps of lignin biosynthesis. Knockdown of C3H/APX (C3H/APX KD) expression reduces the lignin content in bioenergy model grass *Brachypodium distachyon*. However, like many other lignin-modified plants, C3H/APX KDs show unpredictable pleiotropic phenotypes, including stunted growth, delayed senescence, and reduced seed yield. As stated earlier, the collapse of xylem vessels which impairs water and solute conductance in the lignin-modified plant has been hypothesized as one of the likely causes of perturbed growth; however, C3H/APX KD knockdown plants do not exhibit abnormal vessel walls. The precise molecular alterations underlying the negative growth effect observed due to C3H/APX knockdown remain to be explored. Chapter 07 of this dissertation will focus on utilizing mass spectrometry-based proteomics along with other -omics approaches to understand the lignin-associated pleiotropic effect in C3H/APX knockdown plants.

1.5 Dissertation summary

Mass spectrometry has become one of the indispensable analytical tools for systems biology. Due to its ability to comprehensively identify and quantify the proteome of an organism to provide information regarding molecular processes and cellular function it has been widely used in the various areas of systems biology. Bioenergy research has been increasing and is expected to be one of the major contributors to the world's energy supply. For that, it is crucial to have a detailed molecular understanding of bioenergy plants, the factors that impact their growth and fitness, and the factor that limits the efficient conversion of bioenergy plants to biofuel. This dissertation aimed to gain detailed system-wide insight into the bioenergy plants and their dynamic environment by leveraging mass spectrometry-based proteomics and post-translational modifications analysis.

Because the plants' biotic environment impacts plants' growth and fitness, chapters 3-6 discuss the effort made in this dissertation to understand the plants' biotic environment. Chapter 3 shows the first published result on the feasibility of utilizing mass spectrometry-based metaproteomics and global post-translational modifications analysis for studying bioenergy plant-associated microbiome interaction. Chapter 4 characterizes the plant's response to the plant-microbe

symbiotic interaction. In this chapter, the analysis was focused on one type of PTM (phosphorylation) to characterize the lectin receptor-like kinase-mediated signal transduction. Similar to Chapter 4, Chapter 5 focuses on the plant-microbe interaction but instead characterizes the microbial response. Like plants' biotic environment, plants' abiotic factor plays a key role in the regulation of plants' biological processes. Chapter 6 characterizes the proteins and PTMs that are regulated by the diurnal cycle in bioenergy plant poplar to gain an understanding of how the plant senses and responds to environmental cues, such as light. While the dynamic environment around which the plants grow dictates their growth and fitness, the plant cell wall component, lignin, is a major factor influencing the efficient conversion of lignocellulosic biomass into biofuels. Modifying the enzymes in lignin biosynthesis, however, causes stunted plant growth and associated yield penalty. Therefore, Chapter 7 focuses on understanding the precise molecular mechanism underlying the stunted growth and development in lignin-modified bioenergy grass.

CHAPTER 2: EXPERIMENTAL, ANALYTICAL, AND COMPUTATIONAL METHODOLOGIES FOR PROTEOMICS AND PTM ANALYSIS

2.1 Overview of bottom-up proteomics and post-translational modifications workflow

As mentioned in chapter 1, bottom-up proteomics analyzes the peptides that are more uniform in chemistry, have a smaller range of molecular masses and charge states, ionize more uniformly and fragment in a predictable manner compared to the analysis of intact proteins in top-down proteomics (Rogers and Bomgarden, 2016, Dupree et al., 2020, Tao et al., 2019). Thus, it has become a more popular choice for proteomics experiments. A standard bottom-up proteomics experiment consists of three main steps i.e. (i) proteomics sample preparation, (ii) peptide separation and tandem mass spectrometry data acquisition, and (iii) data analysis and biological interpretation (**Figure 2.1**). Each of these three steps can be further divided into several other sub-steps (Dupree et al., 2020).

In the proteomics sample preparation step, the biological samples are first disrupted to extract all the proteins expressed within the samples. The proteins extracted are then digested to obtain a peptide mixture (Dupree et al., 2020). The peptide separation and data acquisition step separate the peptide generated using certain physicochemical properties and generate tandem mass spectra for the peptides (Tao et al., 2019). The final data analysis and biological interpretation step involve mathematical, statistical, and bioinformatics analysis to identify and quantify the peptides and proteins, detect the major changes in protein abundance between conditions and map them to biological processes or pathways to get a meaningful interpretation of MS data (Tao et al., 2019).

Although post-translational modification (PTM) focused experiments such as phosphoproteomics follow all the steps of bottom-up proteomics, due to the low stoichiometry of regulatory phosphorylation sites and low abundance of phosphorylated protein isoforms, a few additional steps need to be added to these experiments (**Figure 2.1**) (Urban, 2022). Firstly, the peptides generated from the proteomics sample preparation step must be enriched for phosphopeptides

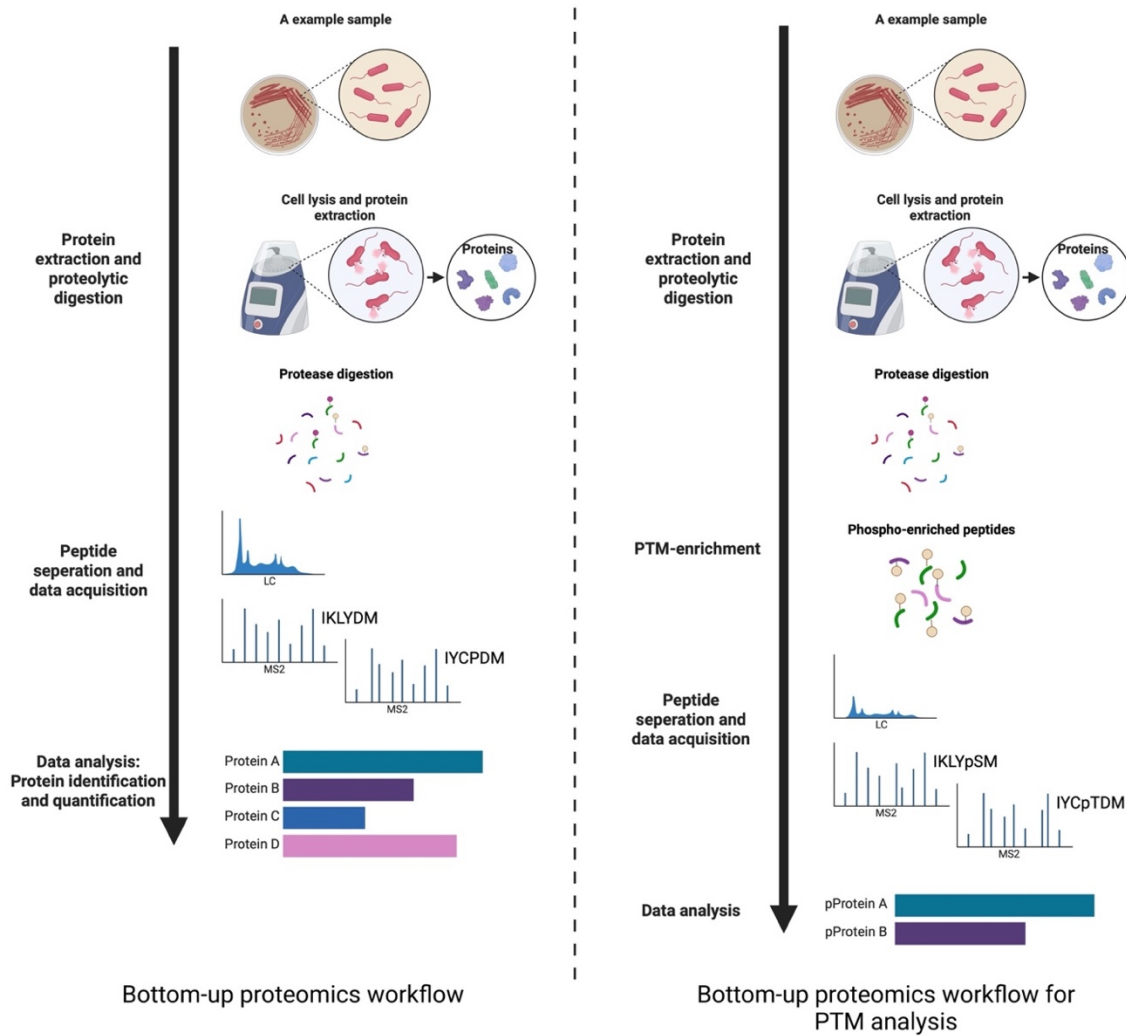


Figure 2.1: Typical workflow of bottom-up proteomics and post-translational modification analysis experiment.

In bottom-up proteomics, proteins are first extracted from biological samples, then digested using protease into peptides. The digested peptides are then separated using the LC system and analyzed in mass spectrometry. The acquired mass spectra from the analysis are matched to existing peptide sequences using reference database search tools/software and this information is further used to identify and quantify the proteins. For the post-translational modification analysis, a similar approach to proteomics is used; however, the PTM enrichment is often added to improve the issues of low stoichiometry and low abundance associated with PTMs.

before the peptide separation and tandem mass spectrometry data acquisition step (**Figure 2.1**). Next, after the data acquisition, the bioinformatic pipeline must be modified to suit the phosphoproteomics analysis. If the goal is to characterize the entire PTMs rather than a detailed perspective on one type of PTM, then the selective enrichment step is not needed and most of the optimization should be focused on the bioinformatics pipeline.

There is a plethora of techniques that can be utilized for each step of bottom-up proteomics and PTM analysis. While acknowledging the commonly utilized technique, this chapter will mainly focus on reviewing the experimental, analytical, and bioinformatic methods utilized in this dissertation.

2.2 Proteomics sample preparation

2.2.1 Considerations before proteomics sample preparation

To perform a comprehensive analysis of entire proteins in a given sample, proteomics sample preparation is the first and most critical step (Graves and Haystead, 2002). Failure to successfully perform high-quality sample preparation renders the sensitive and high-throughput downstream steps pointless. For a successful sample preparation campaign, one needs to provide careful consideration of multiple factors before starting the sample preparation campaign. The first factor to consider is the sample type. Prokaryotes and eukaryotes have different cell sizes and proteome complexity, which can result in different complexity of peptides that can determine the type of LC-separation needed. Similarly, plant cells contain multiple layers of polymers such as cellulose, hemicellulose, and lignin which makes them relatively harder to disrupt compared to animal cells thus, the cell lysis and disruption step need to account for this biological fact. In a similar context, the disruption of gram-positive bacteria can also be difficult due to the presence of a thick peptidoglycan and lipoteichoic acid outer layer compared to gram-negative bacteria (Brown et al., 2015). Sample type also dictates the starting biomass needed for the successful proteomics measurement. For instance, working on the plant root sample requires 4 to 5-fold more starting biomass compared to isolated microbial cell culture. Generally, for the plant root samples starting dry biomass should be at least a few milligrams while microbial cell culture can be around a

milligram. Besides the sample type, the biological questions being asked also determines the starting biomass. For example, if the goal of the research is to characterize the specific PTM, this analysis will require around 100 times more starting biomass samples compared to the proteomics for the same biological system.

The way samples are collected and preserved should also be considered before the proteomics campaign. While working on the microbial sample, the sample should be harvested at the same growth phase otherwise the results could be influenced due to the difference in metabolic and growth rate of the microbial organism across the experimental conditions. Similarly, for the time course study, cells or tissues should be carefully harvested at different time points. All samples collected should be flash frozen immediately after collection at -80 C to prevent the possible proteolytic degradation or cleavage of proteins by cellular proteases. Moreover, flash frozen also halts metabolic processes, so that the sample can be reflective of the condition sampled. The researcher should also know about buffers, salts, carrier proteins, and other potential contaminants in the samples before starting the sample campaign. Mass spectrometry is quite sensitive to charge-carrying species; thus, these contaminants could interfere with the measurement.

Finally, careful consideration should also be given to the biological and technical replicates needed for a study to obtain reproducibility. Biological replicates are the measurements of distinct samples that show biological variation while technical replicates are the repeated measurements of the same sample that determine the variation due to the instrument (Tsvetkov et al., 2019). While both are important, having sufficient biological replicates is crucial for any statistical data analysis and meaningful biological interpretation. The minimum number of biological replicates needed for proteomics is three; however, having more than three is useful.

These key initial considerations before even starting the sample preparation campaign can dictate the success of a proteomics experiment thus, need to be thought thoroughly.

2.2.2 Biological samples to protein content

Initially, in the proteomics sample preparation, biological samples are separated into protein and nonprotein contents such as metabolites, lipids, and nucleotides. To access the biomolecule

contents (protein and nonprotein) from the biological samples, cells are first disrupted with various techniques such as bead beating, sonication, or French press. The cell disruption method will vary based on the biological sample type and the type of proteome analysis. Cell lysis or disruption is done directly in a buffer solution containing anionic detergents like sodium dodecyl sulfate (SDS) or sodium deoxycholate (SDC), strong denaturants like urea or guanidine, or non-ionic detergents like Triton X-100, NP-40, digitonin (Feist and Hummon, 2015). These detergent solutions help to breach the cell membrane to release the cellular content, inactivate cellular proteases and help to solubilize the membrane proteins. The protein content from the resulting cell lysate is separated from non-protein content using the protein extraction method such as centrifugation, trichloroacetic acid (TCA)/acetone extraction, chloroform/methanol extraction, or phenol /ammonium acetate extraction methods (Jiang et al., 2004, Mechin et al., 2007, Isaacson et al., 2006). These protein extraction methods can efficiently separate proteins from other biomolecules that can interfere with the downstream steps, such as salts, lipids, DNA/RNA, and detergents, which are often introduced for cell lysis (Tao et al., 2019). A recently developed protein aggregation capture method can also be used to extract the proteins which utilizes magnetic beads to capture the precipitated proteins. The protein extraction method is generally performed in chemically reduced and alkylated proteins.

Extracted proteins in the sample preparation step are chemically reduced and alkylated to disrupt all cysteine-cysteine disulfide bonds. This will help to denature the protein and prevent them from refolding, thus aid in efficient protein digestion (Kinter and Sherman, 2005). Dithiothreitol (DTT) or tris(2-carboxyethyl) phosphine (TCEP) are the common chemicals utilized for the disulfide bond reduction and chloro- or iodoacetamide (IAA) are the common chemicals utilized for the alkylation of a cysteine residue. The alkylation step prevents the reformation of disulfide bonds. Usually, the DTT to IAA is introduced in the 3:1 ratio. During the alkylation step, care must be taken to prevent the over-alkylation or non-specific alkylation of the protein. Moreover, IAA is light-sensitive and easily degraded in light. Therefore, the alkylation step is done at room temperature in the dark for usually no more than 30-60 minutes. While this is the general protocol, there are some special cases where cysteine–cysteine disulfide bonds can be preserved for structure interpretation or protein-protein interaction (Lu et al., 2015).

2.2.3 Protein digestion and cleanup

Protease is added to the protein mixture that has been reduced and alkylated to digest protein into peptides. The protease digest converts a complex protein mixture into a much more complex peptide mixture. For example, proteolytic digestion of 10,000 proteins can generate around 200,000-500,000 peptides. This might seem counterintuitive; however, this is better for the analytical measurement because the resulting peptides are relatively more uniform in size, shape, and charge that can be separated easily in liquid chromatography. Protein digestion can be carried out at different protein to protease ratios and also with different combinations of proteases (Tao et al., 2019). Frequently utilized proteases include trypsin, Lys-C, CNBr, chymotrypsin, and Arg-C. Among these, trypsin, a highly specific serine protease, has been the gold standard protease for protein digestion. Around 96% of the datasets deposited in the Global Proteome Machine Database (GPMDB) have been generated using trypsin protease (Fenyo and Beavis, 2015). Trypsin is a very effective protease with a high catalytic activity that is most active at 37°C with an optimum pH of 8. It cuts proteins at the C-terminal end of lysine (K) and arginine (R) amino acids with the exception being K or R with proline (P) is the carboxyl side (Kiser et al., 2009, Paizs and Suhai, 2005, Tao et al., 2019). It generates relatively short peptides (~10 amino acids long) with at least two positively charged residues, which makes these peptides well-suited for downstream peptide fragmentation and identification by LC-MS/MS (Paizs and Suhai, 2005). Trypsin generates peptides of ~10 amino acid length because on average the K or R occur every 10 amino acids across the length of a protein. The wide use of trypsin has led to optimized analytical tools based on tryptic peptide properties. Alternatively, LysC protease can be used under denaturing conditions since it is active in up to 8 M urea and cuts at the C-terminal side of lysine residues. Similarly, CNBr can be another alternative to trypsin. It can be used for membrane protein digestion since most hydrophobic domains of membrane proteins lack enough lysine or arginine residues for trypsin digestion (Washburn et al., 2001). However, the main disadvantage of CNBr is its toxicity. All reactions and sample handling with CNBr should be performed in a fume hood. It is important to note that protease digestion may not always proceed to completion, therefore it is best practice to consider peptides that might have occurred from the missed cleavage. After digestion, the peptide mixture is cleaned and often desalted using a C-18 column. Peptide cleanup is an important

step that removes expected and/or unexpected contaminants that could have been introduced during proteomics sample preparation steps (Waas et al., 2019). Common contaminating substances from the sample preparation steps are detergents, particles from molecular weight cutoff filters, and sample enrichment particles (Waas et al., 2019). The removal of these contaminating substances is crucial otherwise, it could adversely affect the data acquisition quality, LC column materials integrity, and MS performance. Reversed-phase C18 resin column has been a popular choice for peptide cleanup and desalting that effectively removes salts as well as helps to concentrate peptides or exchange buffers (Waas et al., 2019).

2.2.4 Protein/Peptide enrichment for PTM analysis

If the overall goal of the study is to profile the entire PTMs, then the selective enrichment step is not required because that will defeat the goal of the study. However, if the goal is to get a detailed perspective on one type of PTM such as phosphorylation, glycosylation, ubiquitination, etc., this step improves PTM identification. For this dissertation, both global PTMs analysis and one type of PTM-focused analysis will be utilized. Global PTMs analysis will be performed for projects discussed in Chapters 3 and 6. One PTM-focused experiment will be utilized for phosphorylation modification in Chapter 4.

In some cases, proteins with phosphorylation modification are directly enriched from the cell lysate. However, enrichment after protein digestion is the commonly used method. In the case of phosphoproteomics, proteins or digested peptides are enriched using affinity purification. For affinity purification either immobilized-metal affinity chromatography (IMAC) or metal oxide affinity chromatography (MOAC) can be used, as they bind the phosphopeptides much more selectively than the more abundant non-modified peptides. Therefore, selective enrichment can substantially improve phosphoprotein identification by multiple orders of magnitude (Andersson and Porath, 1986, Posewitz and Tempst, 1999).

The IMAC approach relies on the high affinity of phosphate groups to cations like Fe^{3+} , Ti^{4+} , and Zn^{2+} (Urban, 2022). The method has been successfully implemented for both offline and online separation of phosphoproteins and phosphopeptides. For the MOAC method, metal oxides like

titanium dioxide (TiO₂) are utilized as a chelating agent (Urban, 2022, Han et al., 2008). To prevent nonspecific chelation, esterification can be executed on acidic residues before enrichment, which significantly improves the specificity of the enrichment (Han et al., 2008, Ficarro et al., 2002). Besides IMAC and MOAC, immunoprecipitation has also been used for phosphorylated protein enrichment (Ballif et al., 2008, Rush et al., 2005). In this method, highly specific antibodies are used which allow proteins containing phosphorylated amino acids to be enriched with high specificity (Tao et al., 2019). The enrichment with this method could be biased toward a certain type of phosphorylation or sequences due to the high specificity of antibodies (Sugiyama et al., 2015, Delom and Chevet, 2006). Therefore, to mitigate this issue multiple antibodies targeting different phosphorylation sites could be used together.

2.3 Fundamentals of peptide separation and tandem MS/MS data acquisition

In the proteomics sample preparation step, a complex protein mixture is digested to obtain an even more complex peptide mixture. For example, proteolytic digestion of 20,000 serum proteins can generate 600,000 peptides excluding PTMs (Anderson and Anderson, 2002). Since MS cannot detect all peptides at once, these peptides must be separated. Peptide separation before MS analysis has become an indispensable analytical technique for bottom-up proteomics. In the past, gel electrophoresis was the gold standard for protein separations which could be combined with mass spectrometry for protein identification (Patterson and Aebersold, 1995, Shevchenko et al., 1996). Using this technique, the proteins can be first separated and visualized in the gels. Selected protein spots can then be excised from the gel for subsequent digestion and characterization with MS. However, more recently high-resolution liquid chromatography has been a technique of choice for this purpose due to its ease of use, minimal sample handling, relatively more reproducibility, and comprehensive nature. While acknowledging the historical relevance of the gel mass spectrometry experiment, this dissertation will solely focus on the proteomics experiment utilizing high-resolution high-pressure liquid chromatography.

As stated previously, digested peptides from the bottom-up sample preparation step can have a high dynamic range of concentration. Thus, liquid chromatography separation of peptides can significantly improve the dynamic range of detection and resolving power of MS in a complex

sample by simplifying the temporal samples that are presented to the ESI source. Thus, the peptide mixture obtained from the earlier steps is separated in liquid chromatography. This helps to fractionate peptides which improves the number of MS/MS data acquired for the peptide mixture and helps with the problem of ion suppression from peptide coelution (Dupree et al., 2020). Peptide mixtures can have a wide distribution of physicochemistry; to better resolve this complexity, different types of separation columns can be used in LC either by itself or in combinations (Mant et al., 2007, Neverova and Van Eyk, 2002). The common separation columns utilized in bottom-up proteomics experiments includes reversed-phase (RP), strong-cation exchange (SCX), and hydrophilic interaction liquid chromatography (HILIC) (Mant et al., 2007, Neverova and Van Eyk, 2002). Moreover, the combination of the different separation columns like SCX and RP in Multidimensional Protein Identification Technology (MudPIT) is equally common which separates peptides effectively by surface charge and hydrophobicity (Issaq et al., 2005, Washburn et al., 2001).

2.3.1 Peptide separation using reversed-phase (RP)

At present, reversed-phase nanoflow LC is the most commonly used technique for peptide separation due to its reproducibility and robustness (Aguilar, 2004, Solovyeva et al., 2018). This method involves the loading of digested peptides onto a single reversed-phase nanoflow capillary column where the peptides are separated by exploiting the peptide hydrophobicity (Tao et al., 2019). In RPLC, the stationary phase is the nonpolar phase usually made of silica material with C18 chains attached to it (Wohlbrand et al., 2013). The peptides generated from the previous step are first loaded to the separation column in an acidic mobile phase containing a low percentage of organic solvent (usually acetonitrile, ACN) (Wohlbrand et al., 2013). After peptides interact and retain in the nonpolar stationary phase, they are gradually eluted with increasing content of the organic solvent in the mobile phase over time (Wohlbrand et al., 2013). This causes peptides with weaker hydrophobic interaction with the nonpolar stationary phase to elute earlier and stronger hydrophobic interaction to elute later.

Once separated by RP, peptides are directly electrosprayed from the capillary tip into the mass spectrometer. This system is set to deliver gradients of high water to high organic over a reverse-

phase capillary column made with C-18, C-8, or C-4 resin of < 2-5 μm size over a specified time of ~1-3 hours. The solvent utilized generally includes a small amount (0.01-0.1%) of organic acid like formic or acetic acid which aid in ionization. In that context, ion-pairing reagents like trifluoroacetic acid (TFA) should not be used because of the potential deleterious effects on electrospray. The single reverse-phase column can have a range of performance metrics, which can be optimized depending on the experiment such as inner diameter which can range from 75 μm - 500 μm , lengths which can range from 10- 25 cm, and flow rates which can range from 0.100- 10 μL per minute. NanoLC, compared to traditional HPLC, can have a flow rate at the nanoliter/min range which allows for minimal sample loss, better separation, higher sensitivity, and better handling of low abundant samples. The efficient separation combined with the high-resolution mass spectrometry can provide deeper identification and coverage even with a single pulsed reversed-phase LC-MS/MS system (Villalobos Solis et al., 2019). This method is ideal for single organism proteomics. The phosphoproteomics experiment discussed in Chapter 4 utilized high-performance liquid chromatography peptide separation using the reverse phase.

2.3.2 Multidimensional Protein Identification Technology (MudPIT)

Depending on the biological source of proteins, enzymatic digestion can result in a highly complex peptide mixture that requires high dynamic range separations. For instance, the peptide mixtures from soil samples or fecal samples are at least an order of magnitude higher in complexity compared to the peptide mixture from a single microbial isolate. In these cases, the combination of orthogonal separation techniques like SCX and RP, can help to better separate peptides and achieve higher proteome coverage (**Figure 2.2**) (Issaq et al., 2005, Washburn et al., 2001, Motoyama et al., 2006). Typically, SCX and RP are coupled in the MudPIT system.

First, the short triphasic microcolumn is made. This is prepared by pressure-loading ~3-5 cm of SCX resin (usually 5 μm particle size; 100 Å pore size) in between ~3 cm of C18 reversed-phase resin (usually 5 μm particle size; 125 Å pore size) (**Figure 2.2**). These columns are often referred to as the back column (**Figure 2.2**). The back column can also be biphasic made with SCX followed by RP. The first short C18 reversed-phase resin in the triphasic back column helps to load and desalt peptide. With the help of an organic gradient made with ~80% acetonitrile and

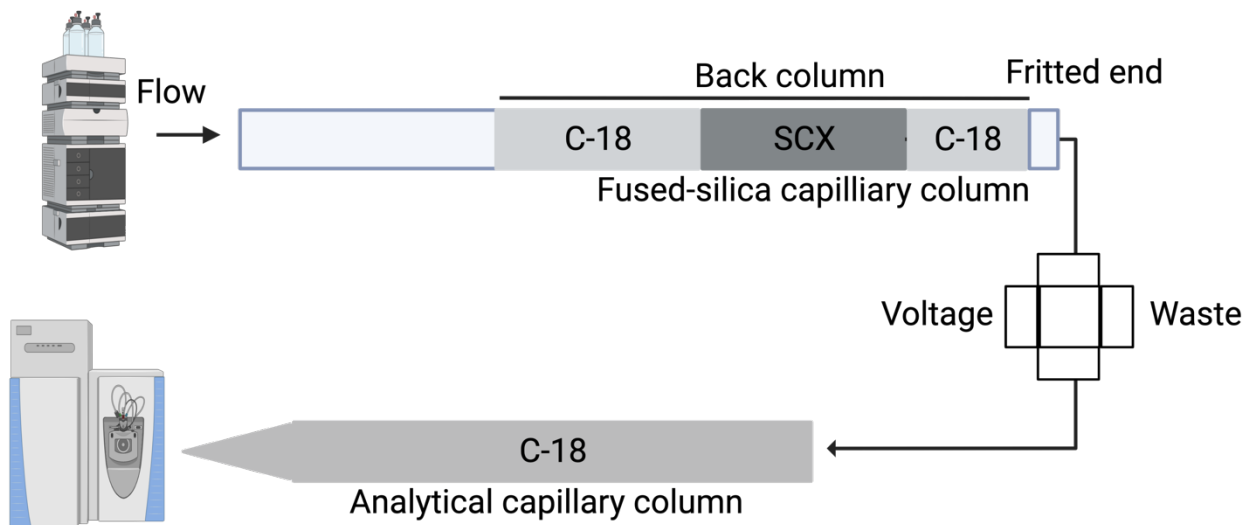


Figure 2.2: Example of a typical Multidimensional Protein Identification Technology (MudPIT) setup.

In MudPIT, the peptide mixture is first loaded and separated onto a back column that is connected to an analytical C-18 capillary column with an electrospray end. With MudPIT, peptides are efficiently separated by the orthogonal separation technique that increases the separative capacity of the chromatographic run by combining dissimilar techniques.

0.1% FA, peptides are pushed to the SCX phase. The SCX phase usually contains aliphatic sulfonic acid groups that are negatively charged, thus tightly binding any strongly basic peptides or analytes (Edelmann, 2011). Peptide interaction with SCX resin is proportional to the total number of positive charges on peptides (Liebler, 2001). To recover the peptides, elution is carried out gradually with mobile phase pulses at increasing salt concentrations (Liebler, 2001, Wohlbrand et al., 2013). Each salt cut or pulse neutralizes different strengths of ionic interaction thus releasing different groups of peptides to the short C18 reversed-phase resin in the triphasic column (Delahunty and Yates, 2007). Throughout this process, a solvent with salt continues to go to the waste. The MudPIT plumbing system splits the flow from the LC system to either the mass spectrometer or to waste. This split flow design in the LC system allows most of the salt to go to the waste thus preventing ion interference from the salt. In addition, split flow allows the HPLC to obtain nanoflows of around 300-400 nL/min because 99% of the flow is directed to the waste.

The back column is connected to a front analytical column made with C-18 with an electrospray end directly placed in front of the mass spectrometer (**Figure 2.2**). Therefore, the peptides released with each salt pulse are pushed to the C-18 front analytical column. The number of salt pulses and the RP gradient length can be adapted as needed to accommodate different complexity of the sample (Delahunty and Yates, 2007). In this dissertation, MudPIT experiments with 3 salt pulses (35mM, 50mM, and 1M Ammonium acetate) and RP gradient lengths of 90-180 minutes were utilized. Once peptides are separated into two orthogonal phases and eluted, they are ionized and enter the mass spectrometer. Most tryptic peptides in acidic pH contain a net charge of +2 or larger, and they can therefore be separated well by SCX (Edelmann, 2011). Peptides in the front analytical column are then eluted into the MS across the reverse phase gradient procedure like the RP-LC strategy described earlier.

For all bottom-up proteomic and PTM experiments discussed in this dissertation except Chapter 4, MudPIT was employed to separate peptide mixtures. This 2-dimensional chromatography methodology can be utilized either off-line or as an on-line system in combination with MS. As the name suggests, off-line loading of the samples are done separately using a pressure bomb before loading to the MS, while on-line loading of the samples are coupled directly to the MS and are carried out with autosampler (Zhou et al., 2004). Off-line loading is flexible, allows the samples

to have some salts and MS incompatible buffers, and can help minimize potential sample losses that could occur due to an autosampler or sample loop within it, however, this method is time-consuming and not as high-throughput as on-line loading (Zhou et al., 2004). All the projects within this dissertation utilize the on-line loading of samples due to their high-throughput nature.

2.3.3 Peptide Ionization

A mass spectrometer instrument contains three key components i.e., i) ion source to convert the analytes to gas phase ions, ii) mass analyzer to measure the m/z of the analytes, and iii) detector to qualitatively identify and quantify the ions at each m/z (Dupree et al., 2020). Peptides separated by LC are first ionized to the gas phase before entering the mass spectrometry. Although there are numerous methods available for ionization, two ionization methods dominate the proteomics studies. They are matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) (Karas et al., 1987, Karas and Hillenkamp, 1988, Fenn et al., 1989). Previously, mass spectrometry was primarily limited to molecules with masses < 5 kDa. However, these two ionization methods allowed the ability to ionize analytes with molecular masses >100 kDa. This has increased the accessible molecular mass by more than an order of magnitude, thus have revolutionized mass spectrometry for biological studies like peptide, protein, and even oligonucleotide studies.

MALDI uses a pulsed laser to desorb biomolecules that have been embedded in a spectrally absorbing matrix compound typically containing a small organic acid. It can provide nondestructive vaporization and ionization of many large and small molecules. MALDI results in mostly singly charged species, such as $(M+H)^+$, although in some cases doubly charged species can also be observed. MALDI is the preferred choice of ionization for “dirty” samples because they are reasonably tolerant of the presence of salts, buffers, and other additives. It has been increasingly used as an ionization source in mass-spectroscopy-based imaging platforms. Although MALDI is an important and widely used ionization method, this dissertation will mainly be focused on chapters that utilize the ESI-based ionization method.

ESI is commonly used in analytical mass spectrometry for bottom-up proteomics due to its ease of use and notable sensitivity. ESI is considered a soft ionization technique because analyte ions retain little to no residual energy and no in-source fragmentation of analytes occurs upon ionization (Banerjee and Mazumdar, 2012). ESI allows the interrogation of both positively charged and negatively charged analytes; however, the positive mode is generally utilized for bottom-up proteomics. Typically, peptide solutions in the nanomolar to micromolar concentration range are prepared in water-acetonitrile solution usually mixed at a 50:50 ratio by volume. To protonate the peptides in solution 0.1 – 1.0 % formic or acetic acid is added. ESI technique employs a high voltage (~2-6 kV) needle to transfer ions from the solution to the gas phase (Banerjee and Mazumdar, 2012). In the positive mode, a high electric field at the ES tip yields a meniscus enriched in positive charges while negative ions, if present, will migrate away from the tip (Kearle and Verkerk, 2009). These polarizing forces cause the formation of a cone-shaped spray, referred to as the Taylor cone, that points towards the counter electrode of the MS (**Figure 2.3**). The unstable end of the cone breaks from the other and emits a jet spray of positively charged droplets. These fine droplets are charged with excessive positive ions which repulse each other (**Figure 2.3**). The repulsion can break the droplet into a smaller droplet. As the charged droplet travels toward the mass spectrometry, the solvent starts to evaporate (**Figure 2.3**). This causes droplets to reduce in volume while maintaining the constant charge which further increases the repulsion forces between charges at the surface. This process led to the explosion of the droplets into small, charged secondary or progeny droplets. The explosion is called Coulomb explosion or Coulomb fission, and this occurs at the Rayleigh limit (Frs, 1882). Ultimately, this leads to the very small, charged particles entering the mass spectrometer in the gas-phase.

The resulting ions are multiply charged ions denoted by $(M+nH)^{n+}$ where n is the number of charges. The multiply charged ions extend the effective mass range of the mass spectrometer that measures m/z of analytes to accommodate molecules of Kilodalton to Megadalton thus making it feasible for biomolecules like proteins and polypeptide fragments (Fenn et al., 1989). ESI process is very sensitive to charge-carrying molecules therefore, the analyte solution should be free of salts, carrier proteins, or other small molecule interference. Usually, the final separation column

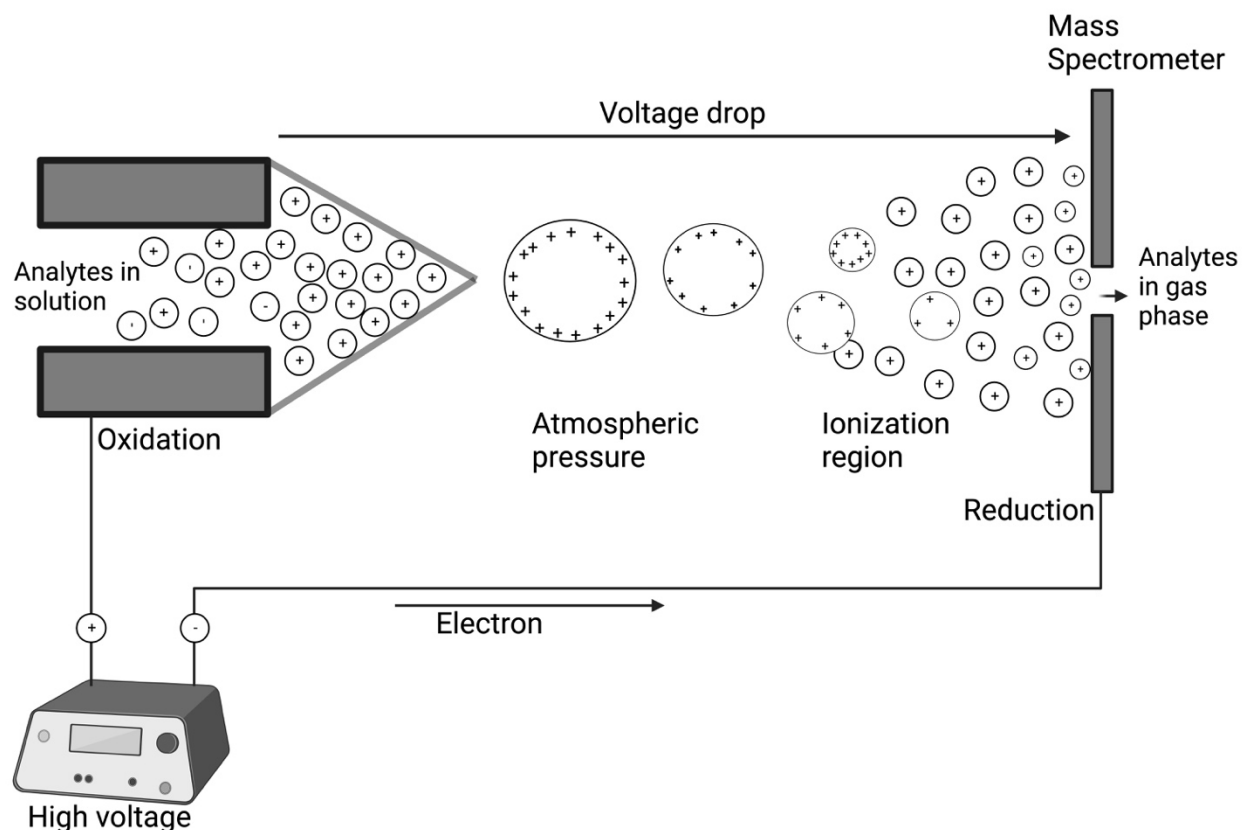


Figure 2.3: A schematic illustration of the electrospray ionization process.

Charged analytes in solution are sprayed from the front or analytical column emitter towards the mass spectrometer. The droplet released contains hundreds of peptide ions initially. As these droplets travel from the spray tip to the mass spectrometer, the solvent evaporates from the droplets. The volume reduction due to solvent evaporation forces ions with the same charge to be together until charge repulsion causes an explosion of the droplets. The high voltage placed between the emitter and the mass spectrometer pulls the charged ions into mass spectrometry in the gas phase.

(analytical column) before introducing peptides into ESI is the reverse phase which aids in removing salts and other small molecule interference.

2.3.4 Mass analyzer

Peptide ionization, as discussed earlier, is a vital step for transferring the analyte into the gas phase. Likewise, the method to detect these ions is equally critical and is the function of a mass analyzer. Various types of mass analyzers exist which mostly share the common characteristic of manipulating the ion motion by electric and magnetic fields for selective detection. The figures of merits can be a useful way to evaluate the different mass analyzers. The mass spectrometry figures of merit are mass resolution, mass accuracy, mass range, detection limits, dynamic range, and the scan or acquisition speed (Zhou et al., 2004).

(i) Mass resolution

Mass resolution can be defined as $m/\Delta m$, where m is the mass of the ion and Δm is typically the peak width mostly measured at full-width half-maximum (FWHM) (**Figure 2.4**) (Murray et al., 2013). Besides the FWHM method, the 10% valley method can also be utilized for the calculation of mass resolution. According to the 10% valley method, mass resolving power is $M/\Delta M$ for two singly charged ion signals of equal height in a mass spectrum at masses M and $M-\Delta M$ separated by a valley of 10% height of either peak (**Figure 2.4**) (Hart-Smith and Blanksby, 2012). The resolving power is defined as the ability to differentiate adjacent peaks in the mass spectrum with a small difference in m/z . A mass analyzer with a high mass resolution is desired because it can differentiate the isotope clusters of a compound which enables the accurate detection of the monoisotopic ion (Wohlbrand et al., 2013). Mass resolution impacts the mass accuracy and sensitivity of the instrument (Wohlbrand et al., 2013).

(ii) Mass accuracy

The measured mass is the experimentally determined mass, and the theoretical mass is its calculated mass. Mass accuracy is the comparison of the measured mass to the calculated

Formula:

$$\text{Mass resolving power} = M/\Delta M$$

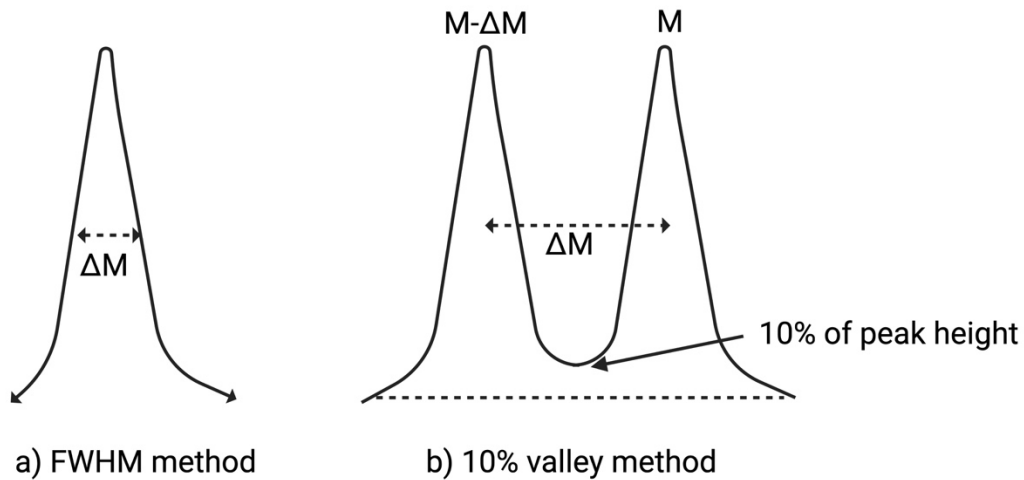


Figure 2.4: Formula and different methods of calculating mass resolving power.

Mass resolving power is calculated by $M/\Delta M$. A) Full width at the half maximum height method. M is the mass of the ion peak and ΔM is the value at full width at the half maximum height. B) 10% valley method. M and $M - \Delta M$ are masses of two singly charged ion signals of equal height in a mass spectrum separated by a valley of 10% height of either peak.

mass. As mentioned earlier, mass resolution impacts mass accuracy. Generally, the low-resolution has an accuracy of ± 0.1 u (nominal mass determination) for single-charge ions and the high-resolution instrument has an accuracy better than ± 0.01 Da (Niessen and Falck, 2015, Wohlbrand et al., 2013). For high-resolution instruments, mass accuracy can be represented as an absolute mass error (measured mass – theoretical mass, in mu) or as a relative error in parts-per-millions (ppm) calculated by (measured mass – theoretical mass) $\times 10^6$ / theoretical mass (Niessen and Falck, 2015). To achieve the appropriate mass accuracy, mass spectrometry needs to be calibrated frequently with standards recommended by the manufacturers.

(iii) Mass range

Mass range is the range between the largest and smallest m/z that can be measured by a mass analyzer. Often only an upper limit is stated when quoting the mass ranges. Increasing the m/z limit will often compromise lower m/z measurements (Hart-Smith and Blanksby, 2012). The mass range is critical when gauging the suitability of a given mass analyzer for the measurement of large analytes such as polymer samples (Hart-Smith and Blanksby, 2012).

(iv) Detection limits

The detection limit is the ability to detect the smallest amount of sample that can be measured with a signal-to-noise ratio of at least 3:1 (Zhou et al., 2004). This factor is crucial while working on the trace analysis. In general, modern mass spectrometers can provide very low background noise and have the ability to detect analyte ions with good signal quality.

(v) Dynamic range

Dynamic range is the range between the least abundant and the most abundant component that can be detected in a single sample. For the proteomics and metaproteomics measurement, the dynamic range is the most important factor because the medium to high

abundance proteins can mask most low abundance proteins (Zhou et al., 2004). A wider dynamic range instrument is desired however, this can be somewhat relaxed by incorporating separation technologies before MS detection (Zhou et al., 2004).

(vi) Scan speed

Scan speed or the acquisition speed is the time taken for a given mass analyzer to record a complete mass spectrum (Zhou et al., 2004). Scan speeds of the mass spectrometry determine its duty cycle which is a percentage of the ions produced in the ion source that are effectively analyzed by the analyzer (De Hoffmann and Stroobant, 2007). Fast scan speeds are required for high-performance bottom-up proteomics. In full-spectrum (or full-scan) mode, mass spectra are continuously acquired between a low m/z and a high m/z within a preset time (mostly ≤ 1 s) (Niessen and Falck, 2015).

For biological studies, linear ion trap (LIT), Orbitrap, quadrupole, Fourier transform ion cyclotron resonance (FT-ICR), and time of flight (TOF) are some of the major mass analyzers (**Table 2.1**). In this dissertation, a Q-Exactive instrument with a quadrupole mass filter and Orbitrap mass analyzer is utilized due to its high mass resolution (up to 140,000 at m/z 200) and accurate mass ions detection (<3 ppm in m/z range between 50-6000) during fast chromatographic separation (**Figure 2.5**). In addition, the high collision dissociation (HCD) feature of this instrument enables sensitive peptide identification and quantitation. These features of the Q-Exactive instrument make it an ideal instrument for deep proteomic/metaproteomics as well as PTM analysis for various projects including the ones in this dissertation. This instrument consists of a few key components that include a bent flatapole, a quadrupole mass filter, curved linear trap (C-trap), a high energy collision cell (HCD cell), and an Orbitrap analyzer (**Figure 2.5**). The charged peptide ions from the ESI source enter and transmit through the S-lens and bent flatapole to quadrupole. The bent flatapole acts to filter all neutral masses and high-velocity particles because the neutral and high-velocity particles can't be controlled to bend at a 90-degree angle. The quadrupole filters the ions according to their m/z ratios and sends them to C-trap which uses radio frequency (RF) and nitrogen gas to cool and align ions. The ions are injected into the orbitrap for the MS1. In the DDA mode for the MS/MS, the top 'n' most abundant precursor ions are selected by quadrupole and

Table 2.1: Typical figures of merit for five common mass analyzers.

Figures of merit	Orbitrap mass analyzers	Linear ion traps
Mass resolving power	10 000–150 000	1000–10 000
Mass accuracy	2–5 ppm	50–100 ppm
Mass range	6000	4000
Linear dynamic range	10E3-10E4	10E3-10E4
Others	Moderate cost and low space	Low cost and low space
Figures of merit	TOF mass analyzers	FT-ICR mass analyzers
Mass resolving power	1000–40 000	10 000–1000,000
Mass accuracy	5–50 ppm	1–5 ppm
Mass range	>100 000	>10 000
Linear dynamic range	1.00E+06	10E3-10E4
Others	Moderate cost and moderate space	High cost and large space
Figures of merit	Quadrupole	
Mass resolving power	100–1000	
Mass accuracy	100 ppm	
Mass range	4000	
Linear dynamic range	1.00E+07	
Others	Low cost and low space	

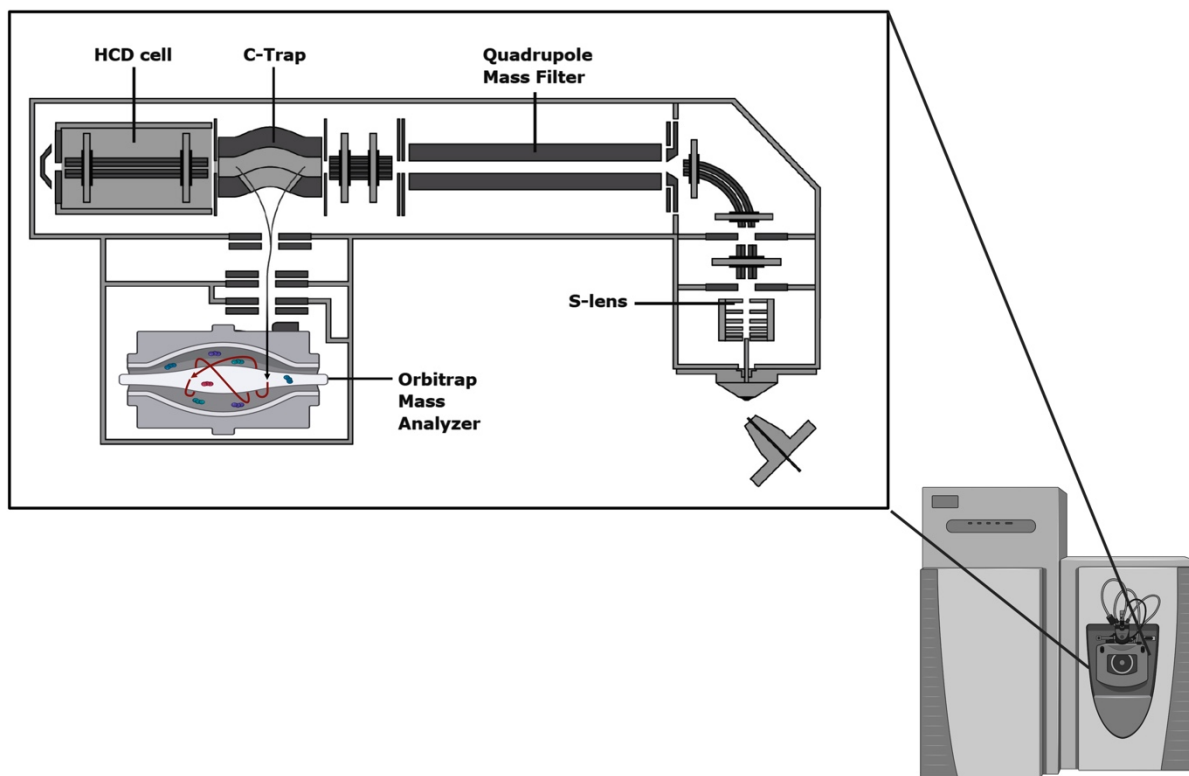


Figure 2.5: Detail schematic of the Thermo Scientific Q Exactive Plus mass spectrometer.

This instrument was utilized for proteomics and PTM analysis in chapters 3-7. The image was originally generated by Thermo Scientific which was further modified in BioRender.

passed to the HCD cell through C-trap which fragments the precursor ions (Niessen and Falck, 2015). The fragmented ions are then again sent to be injected into the orbitrap through C-trap for MS/MS measurement. In Orbitrap, peptide ions orbit around a central electrode with a frequency proportional to $(m/z)^{-1/2}$ (Makarov, 2000). The signal produced by peptide ions is measured repeatedly which enhances the accuracy (Makarov, 2000). The frequency signals are then converted to highly accurate m/z values using Fourier transformation (FT) to yield high-resolution mass spectra.

2.3.5 Peptide fragmentation

Peptide fragmentation plays a crucial role in improving proteome coverage. Depending on the method of choice, one can get different information and perspectives about the structure and composition of an analyte. The two common peptide fragmentation methods are collision-induced dissociation (CID) also known as collisionally activated dissociation (CAD) and electron transfer dissociation (ETD). ETD methods result in fast fragmentation and use either electrons from anions like negatively charged fluoranthene or thermal electrons for fragmentation. They are generally softer and help to preserve the labile bond associated with PTMs. Despite this, CID is the most used method of fragmentation due to its high efficiency, predictable fragmentation, and ease of use. CID or CAD involves three key steps i.e., parent ion isolation, dissociation, and fragment ion detection. During ion isolation, all unwanted ions are removed except for a selected parent ion mass/charge ratio. This isolated parent ion is accelerated translationally by applying appropriate voltages to collide with a target gas such as helium, argon, or nitrogen. Upon impact with the target gas, the translational energy of the ion is converted to internal energy, which causes fragmentation. The extent of fragmentation can be controlled by the amount of excitation energy. In low energy CID, peptides are typically fragmented by cleavage along the peptide backbone. Cleavage can occur at multiple sites and a systematic alphabetic code is used to designate the specific cleavage site (**Figure 2.6**). Depending on the cleavage site, fragment ions that retain the charge on the N-terminus end are designated as “a, b, or c” type ions meanwhile if fragment ions that retain the charge on the C-terminus end are designated as “x, y, or z” type ions (Biemann, 1988, Roepstorff and Fohlman, 1984). The commonly produced ions by CID during proteomics analysis are y- and b-type which leave charges on the N- and C-terminus.

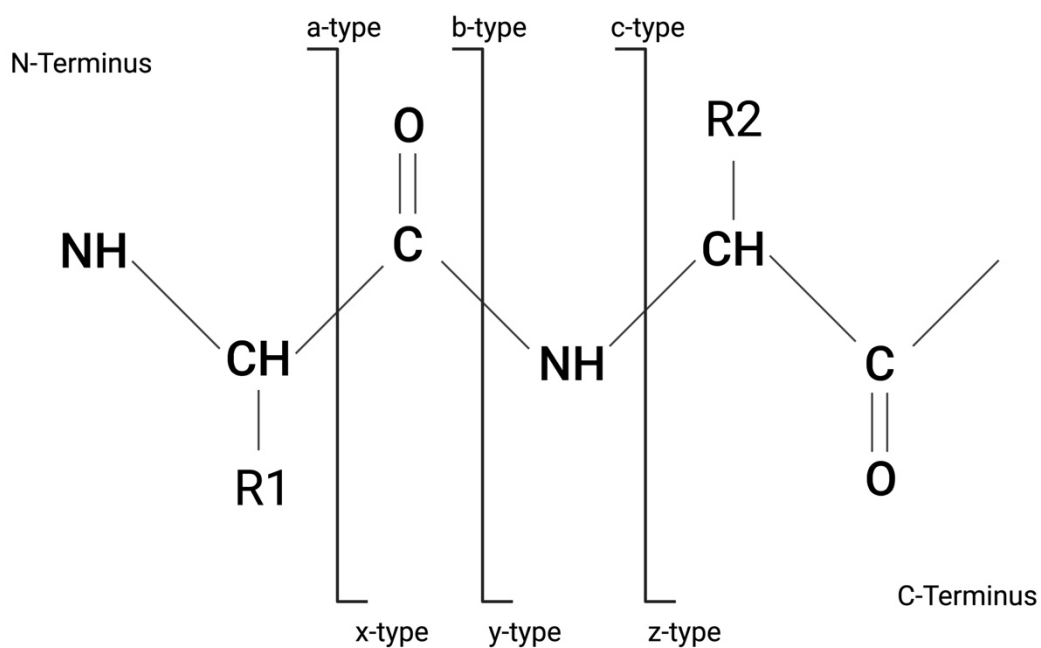


Figure 2.6: A systematic alphabetic code and potential cleavage sites for a peptide sequence.

Fragment ions that retain the charge on the N-terminus end are designated as “a, b, or c” type ions based on the cleavage site. Fragment ions that retain the charge on the C-terminus end are designated as “x, y, or z” type ions based on the cleavage site.

2.3.6 Data acquisition Mode

Mass spectrometry data can be acquired using the data-independent acquisition (DIA) mode or the data-dependent acquisition (DDA) mode (Doerr, 2015, Bateman et al., 2014). In DIA, a mass spectrometer collects data by sequential isolation and fragmentation of specific m/z windows (Doerr, 2015, Venable et al., 2004). All peptide species or analytes within the pre-specified m/z (generally around 10-25 m/z) window are fragmented together and acquired on the same mass spectra (Venable et al., 2004, Li et al., 2020). Detectable ions within an m/z window can contain multiple peptide ions, which can make peptide identification challenging using DIA (Hu et al., 2016). Moreover, this approach can generate a larger amount of data than DDA and requires extensive deconvolution of spectra generated making it computationally challenging and expensive. On the other hand, the DDA mode first scans the parent ions and then selects the subset of topmost abundant ions for fragmentation with a window of $\sim 2-3$ Da. Most modern instruments can scan one parent ion and fragment the top 10 most abundant ions within 2 seconds (Hu et al., 2016). Since DDA samples only the subset of most abundant peptides, it often misses the low-abundance peptides (Venable et al., 2004). Mass spectra acquired with DDA usually contain one peptide per spectra, which simplifies peptide identification. Additionally, dynamic exclusion can avoid MS from the repeated collection of spectra from the same peptide ion over the set period. This allows the mass spectrometer to scan for unique peptide species which can improve the detection range. While there are benefits of using the DIA mode and recent advancement in software and hardware that makes this approach appealing, this dissertation will contain the proteomics experiment with only the DDA acquisition mode.

2.4 Bioinformatics analysis of proteomics and PTMs data

The raw mass spectrometry data acquired for each detected peptide including MS1 and MS2 is analyzed by bioinformatics tools that search reference databases provided by the user to match the spectra and quantify their abundance. The data analysis step significantly influences the overall biological interpretation and thus needs to be executed with a high degree of accuracy. Moreover, a modern-day mass spectrometry system can generate a large amount of data (more than 10 gigabits files per day), which along with the need for high accuracy, present a substantial

computational challenge for data analysis. To alleviate this challenge, numerous vendors based, or open-source data analysis algorithms and tools have been developed in the last couple of decades that have helped thrust the field of bottom-up proteomics further.

The foundation of bottom-up proteomics is peptide identification by the mass spectrometer. Manual interpretation of MS/MS can be used to decipher the peptide sequence from the MS data however, this is a tedious and time-consuming process. With the invention of a computer program like SEQUEST, automated matching of experimental mass spectra to peptide sequences has been possible (Eng et al., 1994). Initially, this was possible mainly due to two timely technological advancements i.e., availability of genome sequencing data that can be used to generate the protein reference database and computer algorithms that can match MS/MS peptide spectra to peptide sequences (Eng et al., 1994). Automation of peptide identification through computer algorithms allows researchers to detect thousands of peptides in a short timeframe which in turn improves the efficiency and scale of proteomics research. Nowadays primarily three major computational search approaches are utilized for peptide identification i.e., reference database search, spectral library search, and *de-novo* peptide sequencing/*de-novo* assisted database searching.

2.4.1 Reference database search

In the reference database search approach, theoretical mass spectra are created by *in-silico* digestion of the protein reference database. A protein reference database is generated from genome sequencing and contains information on all the proteins the specific organism can produce. Experimental tandem mass spectra obtained from mass spectrometry measurement are compared with these theoretical mass spectra generated from a reference database at the spectrum level. The first automated peptide identification algorithm is SEQUEST (Eng et al., 1994). This algorithm computes two main calculations i.e., XCorr and DeltaCN for confirming whether the experimental peptide sequence is a perfect match for a fragmentation spectrum. XCorr is a statistical calculation of correlation between theoretical mass spectra generated from a reference database and experimental tandem mass spectra obtained from mass spectrometry measurement. DeltaCN is the difference between the best and the second-best peptide-spectrum match (PSM) possibilities. Though several database search programs such as MS Amanda, SequestHT, MASCOT,

X!Tandem, Comet, Crux, and MSFragger have been developed for bottom-up proteomics, they all use similar principles with different scoring schemes and algorithms and all require a preexisted reference database generated from the genomes data. These reference database search algorithms first compare each mass spectrum to all the theoretical spectra derived from the reference database with predefined enzyme specificity and precursor mass to generate PSMs (**Figure 2.7**). The best experimental and theoretical spectra matches will be calculated and ranked. During the database search, decoy sequences generated from reversed or random protein sequences are appended to the reference database. These peptide hits on the decoy sequence are used as a true negative which is used as an internal standard to estimate and adjust the false discovery rate (FDR). The reference database must contain the sequences of all expected proteins because none of the database search algorithms can identify the peptides or proteins that are missing in the reference database. Therefore, a proper reference database is crucial for quality proteomics results. Fortunately, with the development and accessibility of next-generation genomic sequencing techniques, any genome can be sequenced at a reasonable cost. These genome sequencing with unprecedented capability can be used to predict and build a curated protein reference database.

2.4.2 Spectral library searching

A spectral library search is an alternative to the traditional database search. In spectral library search, the experimental spectra are compared with the existing standard spectra library to find possible matches (**Figure 2.7**). This search method is more sensitive and reproducible since most spectral library search algorithms use a dot product calculation that utilizes information about the m/z value as well as the intensity of the fragmentation ions (Tao et al., 2019). However, the biggest hurdle for this technique is to generate an enormous amount of peptide standards to create a comprehensive library. For a given biological sample, the proteome can be in the order of thousands. When the trypsin or any protease is utilized for peptide generation the complexity increases from thousands to tens of thousands. This number increases exponentially with the incorporation of PTMs. Moreover, the peptide fragmentation can be impacted by the ionization type, mass analyzer used, and the fragmentation mode which needs to be incorporated into the spectral library. Generally, high confidence peptides from DDA mode are used to construct a spectral library for future peptide identification.

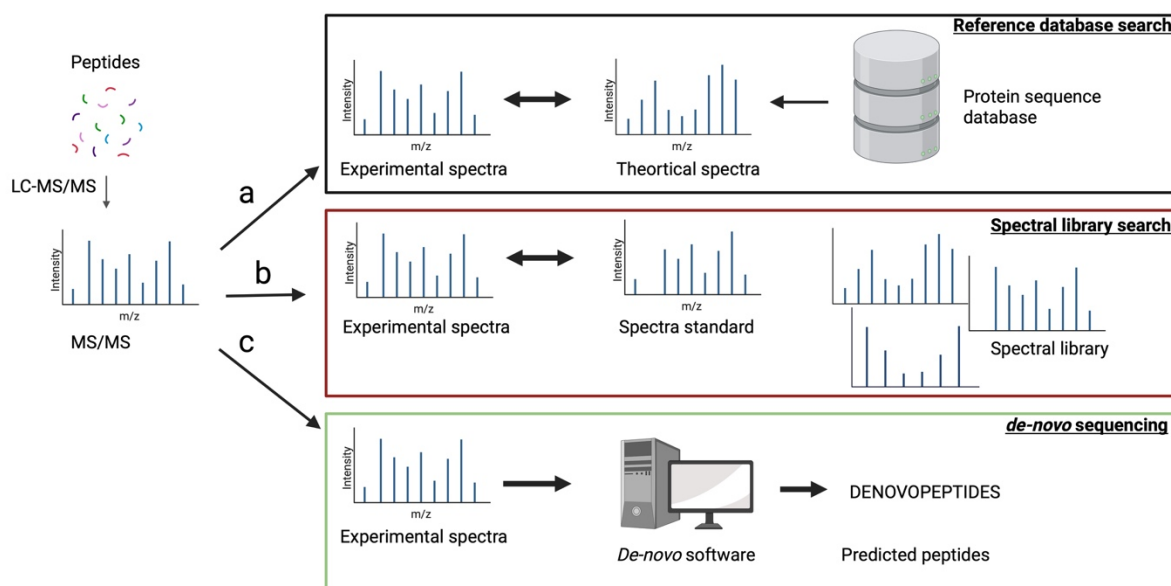


Figure 2.7: Three widely utilized peptide identification strategies.

(a) Reference database search compares the theoretical spectrum generated *in silico* using pre-set parameters from a protein sequence database to the actual experimental spectrum. (b) Spectral library search compares the experimental spectrum to the spectral library acquired through the peptide standard. (c) *De novo* sequencing utilizes a machine learning algorithm to directly derive peptide sequences from the experimental spectrum independent of the reference database or spectral library.

2.4.3 *De-novo* peptide sequencing and *de-novo* assisted database searching

Peptides fragment in a predictable manner. While the widely used traditional database or spectral library searching approach identifies peptides by matching theoretical and experimental spectra, *de novo*-based peptide sequencing provides a suitable alternative. In the *de novo*-based peptide sequencing approach, amino acid sequences are determined for mass spectra independent of the reference database with the help of *de novo* algorithms (**Figure 2.7**) (Li and Tang, 2016). *De novo* peptide sequencing searches for a peptide in a space containing all possible peptides therefore it is not limited to either a known sequence or known spectral library (Li and Tang, 2016). Traditionally, *de novo* algorithms used to be less accurate, and less efficient due to the non-restricted search space and expensive with the computational time and resources when compared to reference database searching. However, with the introduction of high-resolution mass spectrometry and improved *de novo* algorithms with enhanced machine learning, the output of *de novo* sequencing software has been considerably improved. Despite the improvement, *de novo* sequencing by itself still falls behind the database searching and spectral library analysis in terms of both accuracy and specificity. Therefore, the current application of *de novo* sequencing has mainly been in the form of *de-novo* assisted database searching which utilizes both *de novo* sequencing and traditional database searching to improve the identification and quantification.

Several software packages use *de novo* algorithms for *de novo* sequencing or *de-novo*-assisted database searching. Among these, PEAKS has become an industry standard for automated *de novo* sequencing (Zhang et al., 2012). PEAKS is a commercial software by Bioinformatic Solutions and provides the capability of incorporating *de novo* sequencing results into database searches (Zhang et al., 2012). In total, PEAKS performs multiple steps to identify peptides from tandem MS spectra. Notably, PEAKS PTM can be used to identify common modifications in proteins and peptides by allowing the search against all possible modifications from the Unimod database. Similarly, PEAKS SPIDER can be used to detect any possible *de novo* sequencing errors and homology peptide mutations. This pipeline has been used in chapter 06 and has been described in further detail.

2.4.4 Peptide/Protein Quantification: Labeled and label-free quantification

Bottom-up proteomics uses peptide-level abundance to infer the abundance at the protein level. Changes in abundance reflect the cellular and molecular response to the environment or treatments. Bottom-up proteomics can provide quantitative information for thousands of proteins which allow us to get useful insights involved in certain cellular processes or conditions. There are two primary methods of peptide/protein quantification i.e., labeled and label-free quantification methods. In labeled quantification, samples are modified by different weight isotopes before analyzing in mass spectrometry. There are two kinds of labeling approaches i.e., metabolic labeling and chemical labeling (Beynon and Pratt, 2005, Parker et al., 2004, Smolka et al., 2001, Cahill et al., 2003, Yao et al., 2001, Thevis et al., 2003). In metabolic labeling, samples are grown using isotope-labeled amino acids so that the heavy isotope atoms are incorporated in cell growth (Beynon and Pratt, 2005). In chemical labeling, peptides are labeled with isotope-labeled chemical probes after cell lysis and protein digestion (Parker et al., 2004, Smolka et al., 2001, Cahill et al., 2003, Yao et al., 2001, Thevis et al., 2003). While labeled-based quantifications represent the key technique of protein quantification, this dissertation will mainly focus on label-free quantification.

Unlike labeled-based quantification methods which are negatively impacted by added processing steps, inefficient labeling, difficulty in the analysis of low-abundance proteins, and higher expenses related to labeling reagents, label-free quantification can be fast, efficient, and cost-effective protein quantification. Unlabeled samples are analyzed independently and then compared across conditions for relative quantification. Label-free quantitation can be done using three main approaches i.e., spectral count, ion intensity, or area under the curve (AUC). Spectral counting uses total PSM counts of protein as an abundance value while ion intensity or AUC method uses the signal intensity of peptide peaks or the area of the peptide peak belonging to a particular protein for an abundance value.

2.4.5 PTM data analysis

In principle, the data acquisition protocol for proteomics and PTM analysis is similar. The key feature of mass spectra containing the PTM compared to mass spectra without PTM is a mass shift

on specific b- or y-type of ions that is equal to the mass of modification and mass of amino acid (**Figure 2.8**). For example, methylation will have a mass shift of +14.016, acetylation will have a mass shift of +42.011, and phosphorylation will have a mass shift of +79.966 in addition to a mass of amino acid. Besides this, PTMs also shift the peptide's molecular weight. The downstream data analysis tools should be adapted to look for the mass shift in peptide's molecular weight and fragment ions. There are software platforms such as MaxQuant, Proteome Discoverer, and PeptideShaker that allow the common PTM analysis. This can be particularly useful to analyze PTM-enriched data such as phosphoenriched data. However, if the goal is to get the global PTM perspective, PEAKS and MSFragger can be valuable. PEAKS utilizes the PEAKS PTM node to identify the entire PTMs from the Unimod database, including phosphorylation based on the mass shift (Ma et al., 2003). PTM data analysis such as phosphoproteomics should not only calculate the phosphorylated peptide abundance but also the site of phosphorylation and localization confidence of the modification. Several localization scoring algorithms have been developed such as delta score, D-score, PTM-score, phosphoRS, and A-score. In this dissertation, Proteome Discoverer has been utilized for the phospho-enriched data analysis in Chapter 4. Meanwhile, PEAKS has been utilized for the global characterization of PTMs in Chapters 3 and 6.

2.4.6 From protein identification and quantification to biological interpretation

2.4.6.1 Data processing and quality control filtration

The result of bottom-up proteomics can be massive and often overwhelming. To answer the fundamental scientific question using the interrogation of protein abundance values between conditions, there is a need to integrate different aspects of bioinformatics and statistics. Generally, the output of protein results at a certain FDR threshold from various search methods are normalized, filtered for stochasticity and missing values are imputed before the downstream analysis. Normalization helps to remove the systemic bias in LC-MS/MS results that could have been introduced during the various steps of sample processing and data generation such as differences in the amounts of sample loaded, changes in ionization efficiencies, degradation of packing material in chromatographic columns, and drifts in the performance of mass spectrometers over time. Stochasticity filtration with a certain threshold can improve the quality of results by

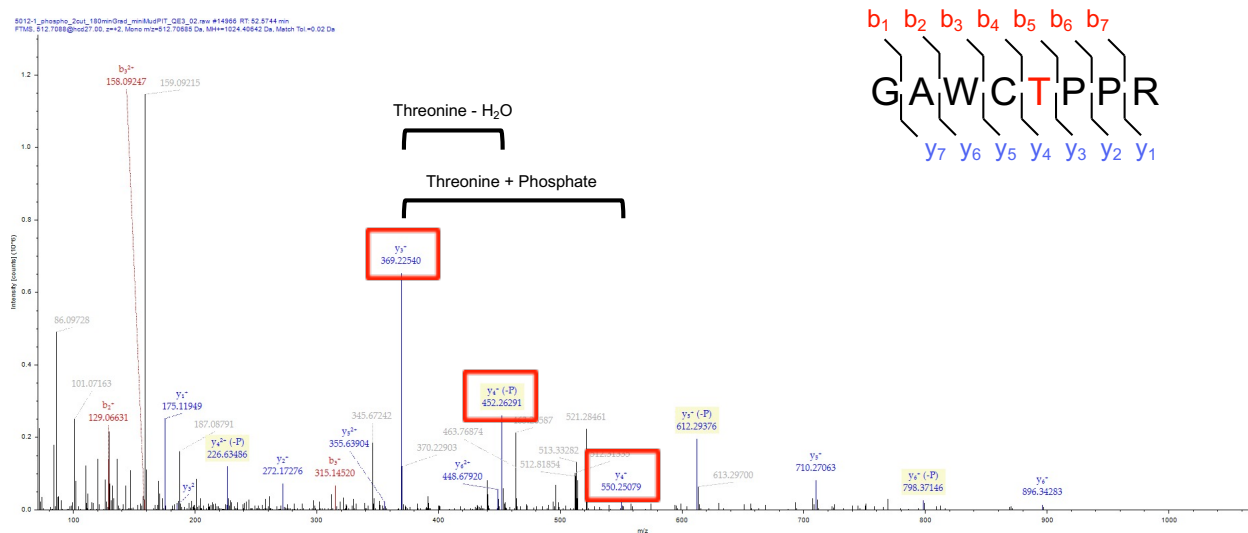


Figure 2.8: Tandem mass spectrum illustrating a specific mass shift caused by phosphorylation modification.

The sequence and type of ions, either b- or y-, are shown in the right corner. The modification occurs in Threonine highlighted in red. This modification causes a mass shift (between y₄ and y₃ in the figure) in a mass spectrum that equals the mass of the modification and the mass of amino acid. Generally, fragmentation with modification and loss of modification can be observed (shown in the red box).

removing the stochastic hits and help focus on the results that are well replicated. For example, filtering the proteins that don't have at least 2 quantified values out of 3 replicates in at least one condition can greatly reduce the protein hits from random sampling. Even after filtration for stochasticity, a key challenge of missing protein abundance values exists. In global proteomics, as much as 50% of proteins can have missing value in final datasets which could be due to biological, analytical, or unknown effects (Lazar et al., 2016, Webb-Robertson et al., 2015). This could negatively impact the biological interpretation of the experiment as it prevents the complete and accurate extraction of quantitative protein values and functional information (Webb-Robertson et al., 2015). One solution to this challenge is to remove the data with missing values; however, this can lead to a drastic reduction of data size and incomplete or misinformation on crucial biological insights, therefore alternative ways of handling the missing value are desired. Fortunately, there are several methods and approaches developed to tackle this issue. This dissertation will contain the experiment with the use of the random tail imputation (RTI) approach (Webb-Robertson et al., 2015). The assumption with this approach is that proteome data are normally distributed, and the majority of the missing data are left-censored (towards the low intensity or abundance) due to the limit of detection of the instrument and can be drawn from the tail of the distribution (Webb-Robertson et al., 2015). For the proteomics data utilized in this dissertation work, the RTI method is utilized where the distribution of mean and standard deviations simulating low abundance values below the noise level was computed first. The missing values were then imputed with random numbers generated from this distribution.

2.4.6.2 Statistical analysis

Proteomics studies can identify a large number of proteins however, the majority of them might be related to the key biological processes that don't change between the experimental conditions. While the proteins that don't change in abundance could be important for cellular function, finding the proteins that change in abundance between experimental conditions can provide more meaningful information to answer biological questions. Statistical analysis can be carried out to identify the proteins with significantly changing abundance between conditions. Depending on the number of experimental conditions, either the student T-test (two conditions) or ANOVA (more than two conditions) can be implemented. In proteomics, p-values of 0.05 or 0.01 are frequently

used as a significance threshold. False positives can be particularly problematic in proteomic studies where many candidate proteins are statistically tested. Therefore, while doing the statistical test one needs to be mindful of the false positive rate. Often along with the p-value, one needs to calculate the false discovery rates or q-values (Noble, 2009). This can be done using several algorithms like Benjamini-Hochberg or permutation-based FDR (Noble, 2009).

2.4.6.3 Data visualization and interpretation

Undoubtedly, high-throughput proteomics can rapidly generate large-scale data. Visualization of these data is a key aspect of both the analysis and understanding of them (Gehlenborg et al., 2010, Michal and Schomburg, 2013, Levine and Davidson, 2005, Nishizuka, 1984). Many visualization methods and tools exist; however, the challenge is to generate clear yet meaningful, cohesive visualizations that give biological insight, without being overwhelmed by the intrinsic complexity of the data (Gehlenborg et al., 2010). There are multiple ways to visualize the qualitative and quantitative information from bottom-up proteomics. Qualitative data such as the number of peptides, proteins, and PSMs identified can be visualized with bar charts with error bars. Likewise, similarities and differences between the proteins/peptides identified in different conditions can be clearly illustrated with Venn diagrams, DiVenn, or even UpSet plots. While these are the common way of visualizing qualitative data, they by no means are the only ways of visualizing qualitative data.

Quantitative proteomics data between two conditions can be presented through volcano plots in which the fold change is plotted versus the p-value. Volcano plots make proteins with high fold changes and low p-values easy to visualize and thus are well suited for illustrating changes in proteomics data. If there are more than 2 conditions with multiple replicates and a large number of protein rows, heatmaps with different color intensities and gradients can be utilized. In heatmaps, data can be clustered and visualized along with the groups of proteins with similar expression profiles which makes it easy to recognize or detect the pattern. Moreover, PCA plots can be utilized to observe the global differences between replicates and conditions in two- or three-dimensional space. Some of the common tools and software that can be utilized for proteomics data visualization and interpretation are listed in **Table 2.2**.

Most of the proteomics tools and software data visualization and interpretation can be used for PTM like phosphoproteomics analysis as well. In addition, there are several bioinformatics tools dedicated to translating phosphopeptide data into meaningful biological output. Some of these resources include PTMs proteomic databases, kinase, and phosphatase activity inference tools, phosphorylation sites tools and resources, kinase-substrate relationships and kinase inhibitors, and the role of kinases in biological signaling (Kamath et al., 2011, Wiredja et al., 2017, Wirbel et al., 2018, Deznabi et al., 2020, Mischnik et al., 2016, Ravikumar et al., 2016, Kanshin et al., 2017, Domanova et al., 2016, Sacco et al., 2018).

2.5 Summary

This chapter described the general methods and techniques utilized in this dissertation along with some alternatives for bottom-up proteomics experiments. Briefly, various types of samples have been utilized in this dissertation that includes plant tissues, microbial culture, plant tissues co-cultured with microbes, or complex microbial communities all grown in a specific growth condition. For the projects in this dissertation, the starting materials have been mainly flash-frozen tissue materials, cell pellets, or cell cultures. Proteins have been extracted and processed from biological materials using two different techniques. In Chapter 3, the protein aggregation capture method has been utilized whereas other chapters utilized the chloroform-methanol extraction method. For all projects, at least three biological replicates have been used. Other experimental approaches have been adjusted based on the sample complexity and the biological question trying to answer. For example, Chapter 5 utilizes the phosphoproteomics protocol with selective enrichment for phosphopeptides because the goal is to get a detailed perspective on phosphorylation modification whereas other projects utilized general proteomics or metaproteomics protocol to characterize the global proteome changes.

A combination of 2-dimensional (Chapter 3 and Chapter 5) or 1-dimensional (Chapter 4, Chapter 6, and Chapter 7) liquid chromatography coupled with tandem mass spectrometry has been utilized for the projects described in this dissertation. For all the projects, the Q-Exactive Plus instrument with orbitrap as the main mass analyzer has been used. All the projects have utilized the top 10 data-dependent acquisition mode. Raw MS files have been searched using either reference

database search tool (Chapters 3-5 and Chapter 7) or *de-novo* assisted search tool (Chapter 6). All the projects utilized peptide and protein level false discovery rate of <1% and have been normalized by LOESS followed by the mean-centered method. Stochasticity is filtered and missing values are imputed using all results before the downstream analysis. Significance is calculated mainly using two statistical approaches i.e., T-test for two conditions (Chapter 4) and ANOVA (all Chapter except 4) for more than two conditions which are further FDR corrected. For biological interpretation of the MS result, a diverse set of software and tools have been utilized which are listed in **Table 2.2**.

Table 2.2: Some useful tools and software for proteomics data analysis and data interpretation.

The description column contains the information on their capability and/or their primary purpose and the URL column provides a link to access these tools.

Tools or software	Description	URL
AgriGO	Database for gene ontology analyses	http://systemsbiology.cau.edu.cn/agriGOv2/
BioRender	Web application to create scientific figures	https://biorender.com
ClueGO	Cytoscape plug-in that visualizes the non-redundant biological terms	https://apps.cytoscape.org/apps/cluego
ClustVis	Web tool for visualizing clustering of multivariate data	https://biit.cs.ut.ee/clustvis/
Cytoscape	Network analysis; extensive list of plug-ins for advanced visualization	http://www.cytoscape.org/
DiVenn	Web tool to visualize and compare gene/protein lists	https://divenn.tch.harvard.edu
g:Profiler	Web tool for functional enrichment analysis	https://biit.cs.ut.ee/gprofiler/gost
Heatmapper	Web tool to interactively visualize the data in the form of heat maps	http://www.heatmapper.ca
InfernoRDN	Graphical front-end to R for common data analysis	https://pnnl-comp-mass-spec.github.io/InfernoRDN/
iPath	Web-based tool for the visualization, analysis and customization of various pathway	https://pathways.embl.de
JMP PRO	Analytics software for data analysis and visualization	https://www.jmp.com/en_us/software/predictive-analytics-software.html
KEGG	Database resource for biological pathway	https://www.genome.jp/kegg/
MapMan	Visualization of abundance data on metabolic pathways	http://tinyurl.com/MapManApp/
MetaboAnalyst	Comprehensive platform dedicated for metabolomics data analysis but can also be used for proteomics	https://www.metaboanalyst.ca
Perseus	Software platform for interpreting protein quantification, interaction and post-translational modification data	https://maxquant.net/perseus/
StringDB	Web tool for Protein-Protein Interaction Networks Functional Enrichment Analysis	https://string-db.org
UpSet Plot	Web tool for visualizing Intersecting Sets	https://upset.app
WikiPathways	Database of biological pathways	https://www.wikipathways.org/index.php/WikiPathways

CHAPTER 3: METAPROTEOMICS REVEALS INSIGHTS INTO MICROBIAL STRUCTURE, INTERACTIONS, AND DYNAMIC REGULATION IN DEFINED COMMUNITIES AS THEY RESPOND TO ENVIRONMENTAL DISTURBANCE

A version of this chapter was originally published by Him K. Shrestha:

Him K Shrestha, Manasa R Appidi, Dr. Manuel I Villalobos Solis, Dr. Jia Wang, Dr. Dana L Carper, Dr. Leah Burdick, Dr. Dale A Pelletier, Dr. Mitchel J Doktycz, Dr. Robert L Hettich and Dr. Paul E Abraham (2021). "Metaproteomics reveals insights into microbial structure, interactions, and dynamic regulation in defined communities as they respond to environmental disturbance." *BMC microbiology*. 21.1 1-17.

This is the collaborative work with researchers from the Plant-microbe interface (PMI) group at Oak Ridge National Laboratory. Our PMI collaborators conceptualized the original study, defined the constructed microbial communities, carried out microbial growth and sampling. This project is the extension of our previous publication: Jia Wang, Dr. Dana L Carper, Dr. Leah H Burdick, **Him K Shrestha**, Manasa R Appidi, Dr. Paul E Abraham, Dr. Collin M Timm, Dr. Robert L Hettich, Dr. Dale A Pelletier and Dr. Mitchel J Doktycz (2021). *Computational and Structural Biotechnology Journal*. 1917-1927

Even though the microbial abundance estimation during community assembly of the 10-member defined community was characterized using 16S rRNA sequencing, the molecular mechanisms shaping this microbial community were not known. Him K. Shrestha and Manasa Appidi led this project to understand the molecular mechanisms utilized by microbial members during the community assembly. Him K. Shrestha's contribution included: 1) conceptualization of proteomics extension of the project, 2) Sample processing, metaproteomics measurement, data analysis, figure generation 3) writing, editing, and revising manuscript for publication.

Note: All the supplementary information including supplementary figures and tables are accessible in doi.org/10.1186/s12866-021-02370-4.

3.1 Introduction

Similar to other eukaryotic organisms, plants have complex host-associated microbiomes that impact fitness and productivity. Microbial interactions within or outside plant tissues are intimately connected to essential processes including water and nutrient acquisition, stress response, and reducing disease or herbivory by priming host defenses (Pii, 2015, De Zelicourt, 2013, Berendsen

et al., 2012, Pineda et al., 2015). Extensive research supports the co-evolution of binary interactions between plants and their associated microbes, yet the complex ecological interactions taking place in nature between microbes and the evolution of microbe-microbe interaction mechanisms remains poorly understood. Furthermore, it remains unclear as to what extent inter-microbial interactions shape microbial assemblages and function in nature and what molecular mechanisms are used by microbes to cooperate, compete, and persist in complex microbial consortia.

Inter-microbial interactions exist in various forms such as resource competition, synergism, antagonism, and these interactions can be altered through the environment. To fully understand the microbial dynamics in its environment, it is crucial to decipher these diverse and dynamic interactions and predict their competitive and cooperative potentials. Several computational and mathematical modeling approaches have been developed to predict the functional contribution of individual microbes in a community including their metabolic functions, inter-species interactions, and community dynamics. Generally, modeling approaches utilize ordinary differential equations (ODEs), annotated genomes and sequence read abundances (Kumar et al., 2019a). Generalized Lotka–Volterra (gLTV) models based on ordinary differential equations have been widely used in understanding temporal dynamics of microbial interactions. Compositional Lotka-Volterra (cLTV), a recent non-linear modeling approach based on relative abundances is shown to accurately predict microbial trajectories over time in a community (Joseph et al., 2020). Constraint based genome-scale metabolic models are used in predicting the extent of resource competition and microbial metabolic interactions, which is facilitated by flux balance analysis (FBA) (van der Ark et al., 2017). Besides these approaches, data-driven inferences of microbial interaction networks, such as ecology guided models that predict the metabolite cross-feeding interactions based on metagenomic and metabolomic measurements, are also valuable in understanding the community dynamics (Goyal et al., 2021). Taken together, microbial modeling approaches provide a guide to start understanding the functional state of microbial communities, as well as of their competitive and cooperative metabolic interactions. However, despite the usefulness of these tools, the key to fully understand the dynamics of microbial communities and the underlying principles within them, is to integrate predictive models with experimental data such as those obtained from

experiments using defined microbial consortia (Goyal et al., 2021, Joseph et al., 2020, Kumar et al., 2019a, van der Ark et al., 2017).

In general, there are two complementary methodological frameworks to study microbial communities: holistically in natural environments or in laboratory-controlled delineated conditions. Both approaches have led to important discoveries in plant microbiome research (Tecon et al., 2019). Using the holistic approach to study the total rhizosphere proteome can provide fundamental information on environment-plant-microorganism interactions, however soil protein extraction is extremely challenging due to the complexity of soil matrix which led to low quantity and quality of the extracted proteins (Tartaglia et al., 2020). On the other hand, by first building a detailed understanding of relatively simple subsets of a microbial community, emergent properties that dictate microbial composition and function can be tested in more complex systems by adding different building blocks (e.g., hosts, community members) for different environments (e.g., liquid, soil). As more complex systems are studied, certain principles are expected to become less relevant, and these realizations begin to uncover higher-order interactions.

Cultivated microbes assembled into defined communities offer a definable landscape to explore microbe-microbe interactions and microbial mechanisms and, when coupled to omics-based technologies, presents a powerful approach to achieve higher predictive power for genotype-to-phenotype associations. Acquiring high-resolution, quantitative data reflecting *in situ* conditions is a crucial starting step to understand microbial community dynamics. Estimation of microbial community dynamics has most commonly been generated by 16S rRNA gene amplicon sequencing; however, these approaches have inherent limitations (Poretsky et al., 2014). Microbial community members experience active vs. dormancy dynamics in their environment. Typically, at any given point in time, only a subset of microbial members is active, while others are in a state of dormancy and behave with strongly reduced metabolic rates (Stolpovsky, 2011, Bowsher et al., 2019). It is important to distinguish between active and inactive microbial taxa to understand their functional contributions to ecosystems. Therefore, pairing both 16S rRNA transcript and 16S rRNA gene sequencing attempts to address this issue by normalizing the measured rRNA levels by the abundances of individual members (Bowsher et al., 2019). However, given that rRNA content or RNA/DNA ratios and growth rates do not always correlate, researchers has suggested

that 16S rRNA/rRNA gene sequencing is best interpreted as potential microbial activity (Bowsher et al., 2019).

Metaproteomics has been demonstrated to be a powerful method for accurate estimation of biomass from viable and functioning cells (Kleiner et al., 2017, Kleiner, 2019). Additionally, metaproteomics allows the large-scale identification and quantification of proteins from microbial communities which helps to characterize microbial membership, their functional roles and interspecies interactions in the community (Kleiner, 2019). Furthermore, recent advancements in optimized bioinformatic pipelines for metaproteomics exploiting *de novo* peptide sequencing facilitate the identification and characterization of post-translational modifications (PTMs) in proteins, which provides unique insights into a largely unexplored level of microbial regulation/adaptation (Han et al., 2011).

Herein, the objective of this study is to integrate a reductionist approach with high-resolution metaproteomics to study microbe-microbe interactions during the assembly of ten taxonomically diverse bacterial members frequently observed in the rhizosphere of *Populus* species. The assemblage process is interrogated for the 10-member microbial consortia co-cultured in a defined (MOPS + glucose) or a complex (R2A) growth medium. For this defined community (DefCom), we aim to i) quantify changes in relative microbial population sizes for the community when passaged in different growth environments, ii) characterize changes in microbial proteomes during the formation of stable communities, and iii) reveal molecular mechanisms driving the community's assembly and structure.

3.2 Materials and methods

3.2.1 Preparation of 10-member microbial communities

Bacterial strains used in this study are *Populus deltoides* derived *Bacillus* sp. BC15, *Caulobacter* sp. AP07, *Duganella* sp. CF402, *Pantoea* sp. YR343, *Paraburkholderia* sp. BT03, *Pseudomonas* sp. GM17, *Rhizobium* sp. CF142, *Sphingobium* sp. AP49, *Streptomyces* sp. YR139, and *Variovorax* sp. CF313 (Brown et al., 2012, Wang, 2021, Blair et al., 2018, Carper et al., 2021) (**Supplemental Figure 3.1A**). The growth rates (OD_{600nm}) for these 10 microbial strains have been

previously measured in R2A and MOPS+ 0.2%glucose media (Wang, 2021). Equal volumes of all 10 isolates with the same normalized OD₆₀₀ were mixed in 10 mL of R2A complex medium (Teknova, # R0005) (Reasoner and Geldreich, 1985) and 10 mL of MOPS minimal medium (Neidhardt et al., 1974) supplemented with 0.2% glucose at 30°C with shaking at 200 rpm. The cultures were transferred into fresh media by diluting 1:10 every 48 hours for a total of 15 transfers. The remaining cultures were pelleted by centrifugation at 12,000 rpm for 15 min, and the spent culture supernatants were stored at -80°C. Based on the 16S rRNA gene amplicon sequencing results, 8-time points (T0, T1, T2, T3, T4, T5, T10, T15, note: T0 is after 48-hour growth in media) with three biological replicates for both R2A and MOPS+glucose were analyzed for proteomic analysis (Wang, 2021).

3.2.2 Cellular protein extraction

Cell pellets were solubilized in 600 µL of lysis buffer (4% sodium dodecyl sulfate (SDS) (Sigma-Aldrich, #L6026, USA) in 100 mM Tris, pH 8.0) supplemented with 1x Halt Phosphatase Inhibitor Cocktail (Thermo Scientific, #78426, USA). Samples were vortexed and then further disrupted with a Bullet Blender storm 24 (Next Advance) for 5 min at setting #10 using 0.15mm Zirconium oxide beads (Next Advance, #ZROB015) at 3:1 sample to bead ratio. Samples were then placed in a heat-block for 10 min at 90°C and centrifuged at maximum speed for 2 minutes to get rid of the foam. Protein concentration was measured using a Nanodrop One spectrophotometer (Thermo Scientific). Samples were centrifuged again at maximum speed for 10 minutes. The cell lysates were transferred into fresh Eppendorf tubes. Samples were reduced with 10mM dithiothreitol (DTT) (Sigma Life Science, #43815, USA) and incubated at 90°C for 10 minutes and then alkylated with 30 mM iodoacetamide (IAA) (Sigma Life Science, #I1149, USA) for 15 minutes in dark to prevent the reformation of disulfide bonds. In fresh tubes, Sera-Mag beads (GE Healthcare Life Sciences, #GE45152105050250, UK) were added at the 1:1 protein to beads ratio and proteins were extracted by protein aggregation capture method as described previously (Batth et al., 2019). Beads were washed with acetonitrile (ACN), LC-MS grade (EMD Millipore Corp., #AX0156-1) on a magnetic rack, after removal from the magnetic rack, samples were added to tubes and then adjusted to 70% ACN. At this point, proteins started to precipitate and bind to the beads. Samples were let to settle for 10 minutes. After 10 minutes, settled beads were gently resuspended. Samples

were again let to settle for another 10 minutes. Samples were then placed on a magnetic rack. The supernatant was removed using a vacuum system. Samples were further washed with 1 mL ACN and 1mL of 70% ethanol (EMD Millipore Corp., #EX0278-1) while on a magnetic rack. A total of 1mL volume of Tris buffer were added to the sample tubes with proteins bound to magnetic beads. The samples were removed from the magnetic rack. Proteins were digested with two separate and sequential aliquots of sequencing grade trypsin (Thermo Scientific, #90057, USA) at 1:75 (wt/wt) protein:trypsin ratio for overnight, followed by 3 hours at 37°C at constant shaking. The samples were then adjusted to 0.1% TFA (Sigma-Aldrich, #302031, USA), vortexed, and centrifuged at max for 10 minutes. Vivaspin 500,10 kDa MW cutoff filters (Sartorius, #VS0102) were equilibrated with 500 µL of Tris buffer and centrifuged at 12,000g for 15 minutes. After equilibrating, the samples were added to the Vivaspin columns and centrifuged at 12,000g for 15 minutes. Tryptic peptides flow-through were collected and desalted on Pierce peptide desalting spin column (Thermo Scientific, #89852, USA) as per the manufacturer's instructions. Desalted peptides were vacuum dried with a SpeedVac Concentrator (Thermo Scientific) and then resolubilized in 0.1% formic acid (Fisher Chemical, #A117-50). Peptide concentrations were measured using the Nanodrop instrument and transferred to the auto-sampler vials for LC-MS/MS measurement.

3.2.3 Extracellular protein extraction

Spent media samples were adjusted to 2mM EDTA (Thermo Fisher Scientific, #R1021) on ice. The samples were filtered using Vivaspin 20mL Centrifugal Concentrators with 5000 Da molecular weight cut off (Sartorius, #VS2012) by spinning at 4,000 rpm at 4°C for 30 minutes. The Vivaspin 5000 Da molecular weight cut-off filters were washed with 1 mL of 4M urea (Sigma Life Science, #51456, USA) in Tris buffer. The proteins were resuspended on the filter membrane with a final volume of 500 µL 4M urea in Tris buffer. DTT was added to a final concentration of 10 mM for reduction of disulfide bonds and incubated samples at room temperature for 30 minutes. Samples were then alkylated by adding IAA to a final concentration of 30 mM and placed at room temperature in the dark for 15 min. Protein concentrations were measured by Nanodrop. Proteins were digested with Pierce Trypsin Protease, MS-grade at 1:75 wt/wt protein to trypsin ratio, overnight at 37°C with constant shaking. Samples were diluted with Tris buffer to a final

concentration of 2M urea and a second 3h digestion with trypsin was performed using the same conditions. The tryptic peptide flow throughs were collected by spinning at 4,000 rpm for 15 minutes. The samples were acidified with 0.5% formic acid and then desalted using Pierce peptide desalting spin column as per the manufacturer instructions. Desalted peptides were vacuum-dried with a SpeedVac Concentrator and then resolubilized in 0.1% formic acid. Peptide concentrations were measured using the Nanodrop instrument and transferred to auto-sampler vials for LC-MS/MS measurement.

3.2.4 Protein identification and quantification

Each sample was analyzed using two-dimensional (2D) liquid chromatography (LC) on an Ultimate 3000 RSLCnano system (Thermo Fischer Scientific, USA) coupled with a Q Exactive Plus mass spectrometer (Thermo Fischer Scientific, USA). For each sample, an aliquot of digested peptide mixture was injected across an in-house built strong cation exchange (SCX) Luna trap column (5 μ m, 150 μ m X 50mm; Phenomenex, USA) followed by a nanoEase symmetry reverse phase (RP) C18 trap column (5 μ m, 300 μ m X 50mm; Waters, USA) and washed with an aqueous solvent. Cellular peptide mixtures were separated and analyzed across three successive SCX fractions of increasing concentrations of ammonium acetate (35mM, 50mM, and 500mM), each followed by a 100-minute organic gradient (25nL/min flow rate) to separate peptides across an in-house pulled nanospray emitter analytical column (75 μ m X 350 mm) packed with Kinetex RP C18 resin (1.7 μ m; Phenomenex, USA). The extracellular peptide mixtures were separated using the same separation regime, albeit with only one SCX fraction. All MS data were acquired with Thermo Xcalibur (version 4.2.47) using the topN method where N could be up to 10. Target values for the full scan MS spectra were 3×10^6 charges in the 300 – 1,500 m/z range with a maximum injection time of 25 ms. Transient times corresponding to a resolution of 70,000 at m/z 200 were chosen. A 1.6 m/z isolation window and fragmentation of precursor ions were performed by higher-energy C-trap dissociation (HCD) with a normalized collision energy of 27 eV. MS/MS scans were performed at a resolution of 17,500 at m/z 200 with an ion target value of 1×10^5 and a maximum injection time of 50 ms. Dynamic exclusion was set to 20 s to avoid repeated sequencing of peptides. All MS raw data files were analyzed using the Proteome Discoverer software (version 2.3, Thermo Fisher Scientific, USA). Each MS raw data file was processed by

the SEQUEST HT database search algorithm (Eng et al., 1994) and confidence in peptide-to-spectrum (PSM) matching was evaluated by Percolator (Kall et al., 2007). Peptides and PSMs were considered identified at $q < 0.01$ and proteins were required to have at least one unique peptide sequence. Protein relative abundance values were calculated by summing together peptide extracted ion chromatograms.

Proteins with at least one unique peptide were exported from Proteome Discoverer. Protein abundances were log₂-transformed, LOESS normalized and mean-centered across the entire dataset using InfernoRDN software (Polpitiya et al., 2008). From this normalized dataset, protein abundances subset for each microbe were extracted and further mean-centered by InfernoRDN. For this study, pairwise comparisons were performed across different passages (i.e., Time Point 0 vs Time Point 1, Time Point 1 vs Time Point 2, etc. for R2A and MOPS+glucose) to identify the differences between the passages. The analysis was limited to proteins that were identified in at least two out of three biological replicates of at least one sample to improve the robustness of the downstream analysis. The abundance values for proteins with missing values were imputed with random values drawn from the normal distribution (width 0.3, downshift 2.8) using Perseus software v.1.6.12.0 (Tyanova et al., 2016). Student's *t*-tests were performed to identify the differences in protein abundance values between different passages. Proteins are characterized as significantly differentially abundant if they pass the significance threshold of $p\text{-value} \leq 0.05$ and absolute log₂ fold-change difference greater than 1.

3.2.5 PTM identification by PEAKS

Raw spectral data collected from both complex and minimal media sample sets were re-searched by *de novo*-assisted database searches against the 10-member community proteome accompanied with common contaminant proteins using PEAKS DB, PEAKS PTM, and PEAKS SPIDER in PEAKS X Studio (Bioinformatics Solutions, Waterloo, Canada). The peptide and fragment ion mass tolerances were set to ± 10 ppm and ± 0.02 Da, respectively. “Trypsin” was set as the enzyme parameter. Features associated with chimera scan were enabled. *De novo* ALC score was set at $>90\%$. A false discovery rate of 1% was applied to accept the peptide sequences and a minimum of three peptides were required to identify a protein. For PEAKS DB, carbamidomethylation

(+57.02) of cysteine was set as fixed modification and oxidation (+15.99) of methionine was set as a variable modification. The PEAKS PTM algorithm was used to identify other types of modifications by allowing the search against all possible modifications from the Unimod database (Creasy and Cottrell, 2004). Similarly, PEAKS SPIDER algorithm was used to detect any possible *de novo* sequencing errors and homology peptide mutations.

To report the number and percentages of modified proteins, only proteins with unique peptide sequences, that is, without any modifications, that were present in at least two out of three biological replicates per passage were considered. Peptide uniqueness was verified with the Protein Coverage Summarizer Tool (<https://omics.pnl.gov/software/protein-coverage-summarizer>) and the data filtered using the Perseus software (<http://www.perseus-framework.org>) (Tyanova et al., 2016). Unimod categories used to discriminate biologically relevant PTMs to those PTMs that can be the products of sample preparation/handling were: “Post-translational”, “Multiple”, “N-linked glycosylation”, and “O-linked glycosylation”.

The PTM profile tables reported by the PEAKS software were used to interrogate individual proteins and their PTMs. These tables list the summed abundance of modified and unmodified versions of unique and shared peptides that fall within a modified peptide sequence window in a protein. Only summed abundances reported in at least two out of three biological replicates per passage and conditions we considered. Summed abundance values were averaged per passage, and the percentage abundance ratios of modified to unmodified peptides were calculated from them. All PTM related figures presented with the manuscript were created with JMP Pro 14 (https://www.jmp.com/en_ca/software/predictive-analytics-software.html).

3.3 Results

3.3.1 Metaproteomics reveals details of microbial community stabilization and population equilibration as a function of growth media

The ten diverse strains used in DefCom originated from roots of *Populus deltoides* (Brown et al., 2012, Hasim et al., 2018, Levy et al., 2017, Schaefer et al., 2013) (**Supplemental Figure 3.1A**). The 10-member DefCom co-culture was grown in liquid defined media (MOPS + 0.2%

glucose) and complex media (R2A) subjected to growth/dilution cycles every two days until the community reached a stable state (state with minimal fluctuation of microbial community composition and abundance) (**Figure 3.1**). These two media were selected to determine the effect of environmental filtering based on a single carbon source with limited essential nutrients and a more complex media with increased carbon resources and nutrients designed for long incubation periods. Metaproteomic measurements of microbial communities provide a robust and accurate assessment of microbial cell population sizes (Kleiner et al., 2017) because proteins make up 40-60% of bacterial cell biomass (Neurohr and Amon, 2020) and are known to have a linear correlation with cellular mass and volume (Milo, 2013). Relative organism cell population size estimates, as determined by summed protein counts or protein summed abundances, shows that communities stabilized, reaching stable population equilibria in both growth media, but formed distinct population sizes and community structures between the two different growth media (**Figure 3.2A and Supplemental Figure 3.1B**). Interestingly, three discrete community transition phases were observed in R2A media, in which the relative abundance and microbial membership were fluctuating systematically before reaching a stable state. A principal component analysis (PCA) for the metaproteomics dataset supports the observed discrete population changes for R2A media by revealing three major clusters (R0, R1 to R5 and R10 to R15) over the experimental passages (**Supplemental Figure 3.2A**). In the defined media, the 10-member community only experienced two discrete phases, with the community stabilizing after the second passage (**Supplemental Figure 3.2B**).

Organism relative compositions were additionally assessed for extracellular fractions (**Supplemental Figure 3.1C**). The extracellular fraction, though most commonly used to characterize proteins secreted from cells, has a composition notorious for its abundance of proteins resulting from cell lysis (**Supplemental Figure 3.1D**), and therefore provides a useful perspective for detecting recent mortality events. In general, organism relative abundances were highly similar between cellular and extracellular fractions. (**Figure 3.2B**). It can be reasoned that the measured correlation between these two fractions indicates that mortality and growth rates for each member in the community are relatively stable at the 48-hour sampling point.

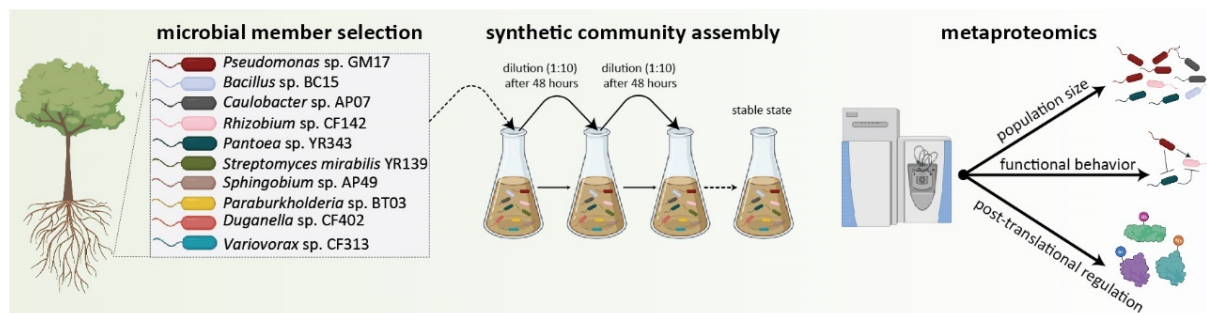
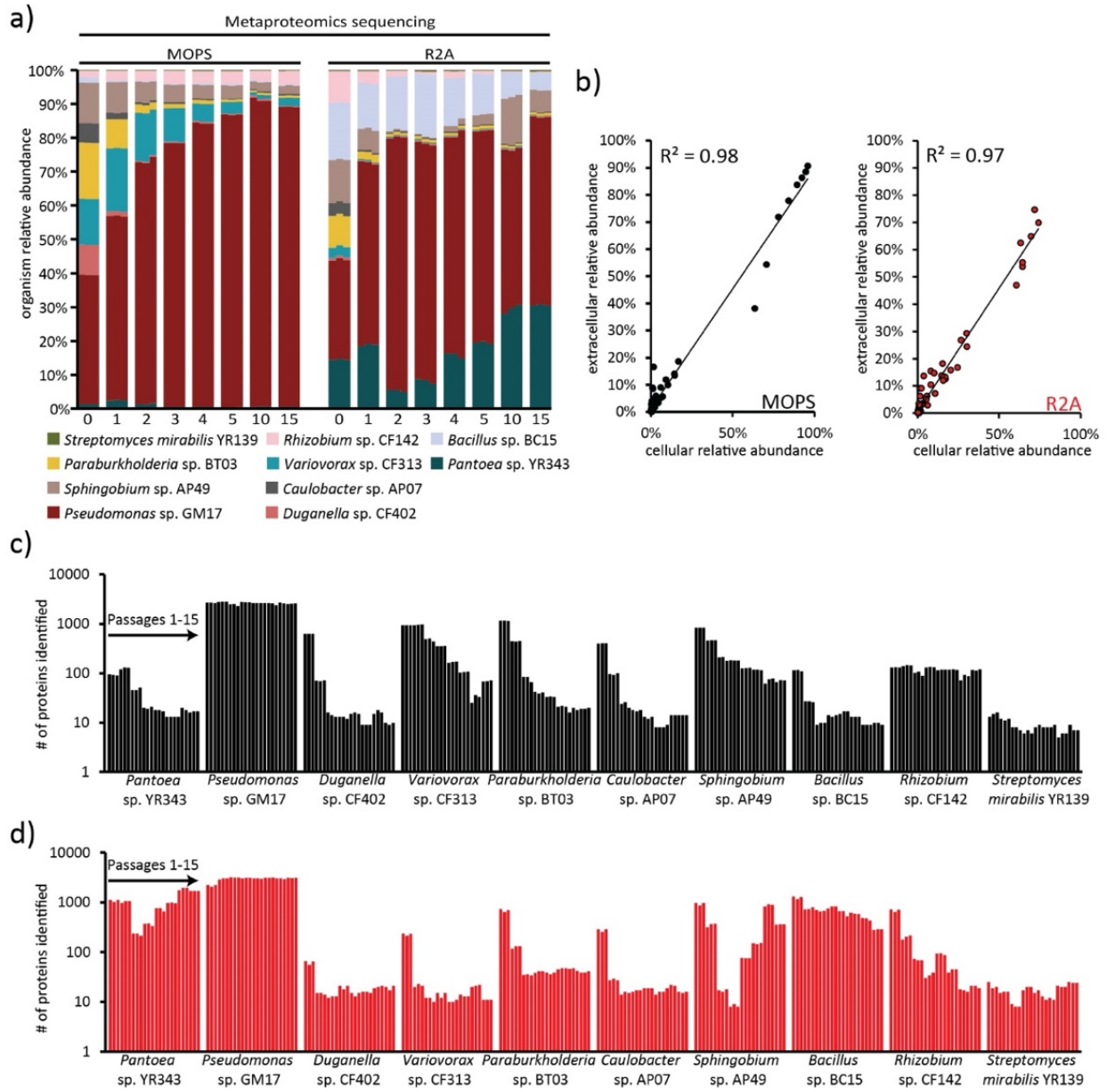


Figure 3.1: Schematic diagram of study design.

Microbial members were selected based on phyla level abundance and tax diversity from natural *P. deltoides* (Pd) rhizosphere communities. These members were inoculated together and diluted 1:10 after 48 hours for 15 passages. Samples were analyzed by high-resolution metaproteomics to understand the community size and structure, molecular mechanisms underpinning microbial behavior and functions, and post-translational regulations. Created with BioRender.com.

Figure 3.2: Assessment of microbial population size.

(A) Microbial community composition from cell pellets using metaproteomics was estimated by total protein count for each community passaged in defined media MOPS + glucose (MOPS) or complex (R2A) media. There was a total of 15 passages per medium and 3 biological replicates per passage. Individual colors in the stacked bar charts represent each microbial member. **(B)** Cellular estimates of organism relative abundance plotted against extracellular estimates of organism abundance for each passage for MOPS (black circles) and R2A (red circles). Each circle represents the averaged abundance across replicates for a single passage. Proteome depth (number of proteins identified) is plotted for each microbe per sample measured in the defined microbial community passaged across **(C)** defined MOPS + glucose (black) and **(D)** complex R2A (red) media.



3.3.2 Quantifying microbial species reveals functional behaviors during community assembly

The ability to holistically study inter-microbial interactions and identify molecular mechanisms used by microbes to cooperate, compete, and persist in complex microbial consortia arguably requires expansive proteome coverage. Using state-of-the-art techniques, proteome coverage for the 10-member DefCom varied between microbes, ranging from tens to thousands of proteins per organism (**Figure 3.2B and 3.2C**) (**Supplemental Table 3.1 and 3.2**). While the measured proteome depth can be affected by growth of microbes and metabolic rate, microbes often adapt their behavior to states of lower metabolic activity during unfavorable growth conditions (Stolpovsky, 2011). As such, valuation of the completeness of a metaproteome measurement is not trivial and should not be used as a filter for interpretation.

To identify how members of this DefCom are responding to changes in the environment, significantly changing proteins were identified per organism using Student's t-test between consecutive passages for both media. In complex media, three organisms, *Pseudomonas* sp. GM17, *Pantoea* sp. YR343, *Bacillus* sp. BC15, showed major proteome abundance changes across the passages (**Supplemental Figure 3.3**). In minimal medium, *Pseudomonas* sp. GM17 was the only organism to show major proteome abundance changes. Given these observations, *Pseudomonas* sp. GM17, *Pantoea* sp. YR343, *Bacillus* sp. BC15 were further interrogated to identify the proteome-level changes in biological processes that influenced their functional behaviors and abundances in the community.

3.3.3 *Pseudomonas* sp. GM17 behaves antagonistically during the community selection process

Pseudomonas sp. GM17 was the dominant microbial member in both R2A and MOPS media, with the highest number of significant proteins in the majority of the pairwise comparisons (**Supplemental Figure 3.3**). The observed proteomes between two media conditions have substantial protein and functional overlap, with about 79% of the proteins expressed in both media conditions and with 17% uniquely identified in R2A and 5% uniquely identified in MOPS (**Figure**

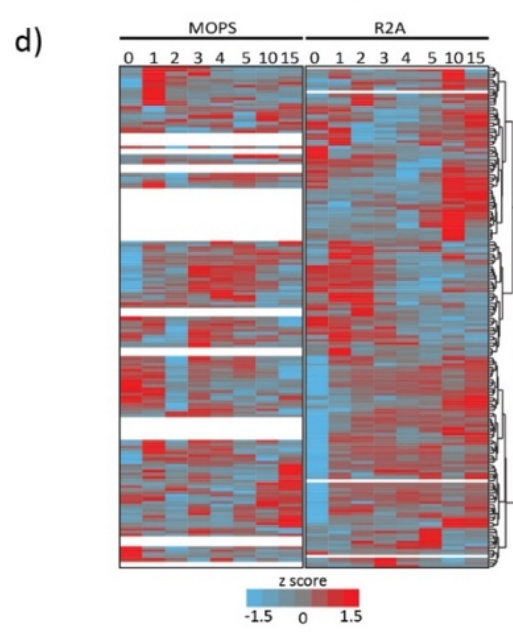
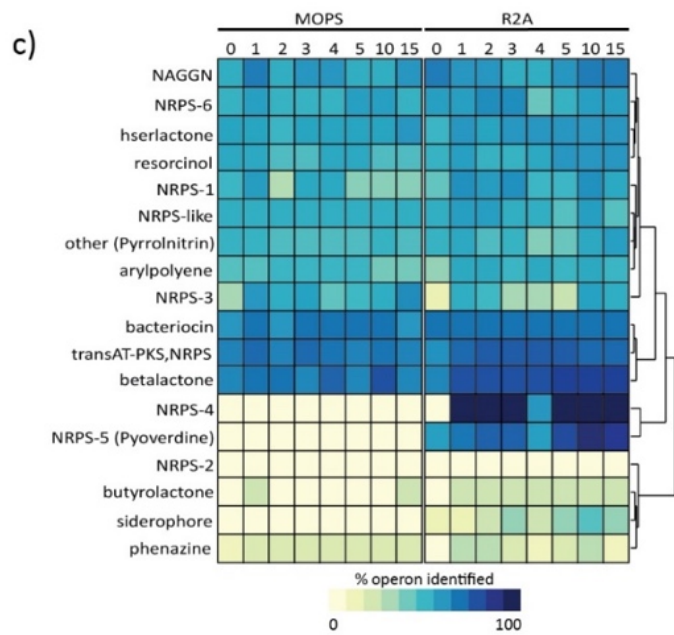
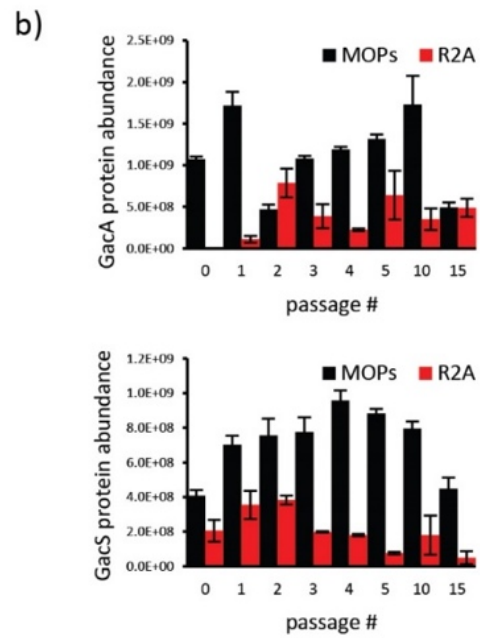
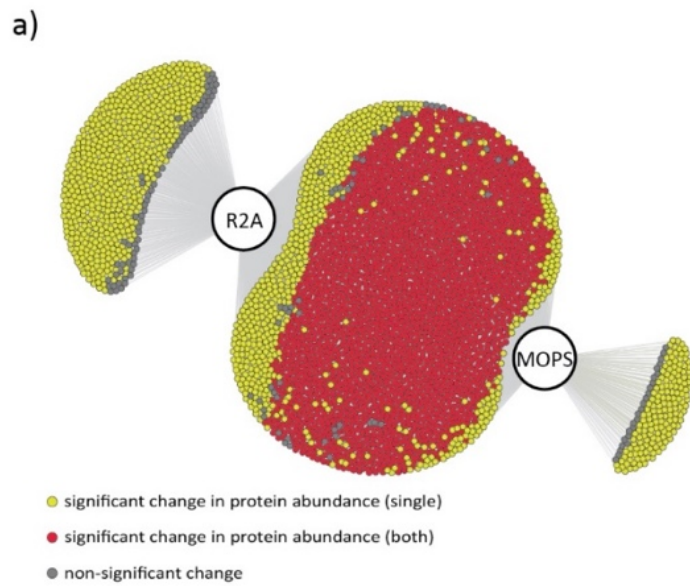
3.3A). Additionally, the cluster of orthologous groups (COG) based annotation of the proteins showed a similar distribution of functional categories in both media suggesting similarity in mechanisms facilitating its dominance (**Supplemental Figure 3.4A**).

In general, *Pseudomonas* species are known to possess a large repertoire of antibiotics and secondary metabolites (SMs) that can equip them with a fitness advantage in multi-species consortium (Cheng et al., 2019). A recent study has shown *Pseudomonas* sp. GM17 to have negative interactions with the majority of the microbial members in this 10-member DefCom (Wang, 2021). In fact, notable changes in relative protein abundances were identified for a sensor histidine kinase GacS (WP_007927747.1), which plays a role in sensing environmental signals and GacA (WP_020295157.1), a response regulator activated by GacS (**Figure 3.3B**). This two-component system is well known for its regulation of biosynthetic gene clusters (BGCs) and is involved in the biosynthesis of antibiotics and SMs in *Pseudomonas*. Based on antiSMASH v5.0 predictions (Chen et al., 2021, Blin et al., 2019), the *Pseudomonas* sp. GM17 genome is suspected to have 18 BGCs encoding for the antibiotics and SMs (**Supplemental Table 3.3**). Of these 18 BGC, proteomics analysis identified proteins from 17 BGC in R2A and 14 in MOPS media (**Figure 3.3C**). Several BGCs in both media are involved in the biosynthesis of antibiotics such as pyrrolnitrin, phenazine, and resorcinol. Pyrrolnitrin biosynthesis proteins are involved in converting tryptophan to pyrrolnitrin. Pyrrolnitrin has been demonstrated to have antimicrobial activity (Selin et al., 2010). Similarly, broad-spectrum antibiotic phenazines are known to enhance the competitiveness of *Pseudomonas* and are involved in antagonistic activity (Timms-Wilson et al., 2000). Likewise, resorcinol is also well known for fungal antagonism, biofilm formation, and biocontrol activity in *Pseudomonas* (Calderon et al., 2019).

Three BGCs were uniquely identified in R2A media, of which two were predicted to encode non-ribosomal peptide synthetase (NRPS) and the other encodes for a siderophore (**Supplemental Figure 3.4B**). NRPS are multimodular enzymes that can produce products with diverse properties such as toxins, siderophores or antibiotics (Martínez-Núñez, 2016). Protein-level interrogation of one of the uniquely predicted NRPS BGC suggests involvement in pyoverdine biosynthesis. Pyoverdine is the fluorescent green-yellowish pigment produced by *Pseudomonas* species and represents a key siderophore (Cornelis and Matthijs, 2002). This siderophore is a very efficient

Figure 3.3: Metaproteomics analysis of *Pseudomonas* sp. GM17 functional behavior during community assembly.

(A) Overlap of proteins identified between both media. Each node represents a unique protein accession, and the color indicates whether the relative protein abundance changed significantly based on ANOVA in one (yellow) or both media (red) or not significant in either (grey). Figure generated using DiVenn 2.0. **(B)** Relative protein abundance for the GacS sensor histidine kinase and the GacA response regulator in MOPS (black) or R2A media (red). Error bars represent standard error for each set of biological triplicates. **(C)** Heatmap (one-way clustering using ward method) illustration for 18 antibiotic and secondary metabolite gene clusters predicted by antiSMASH v5.0. Color gradient represents the percentage of proteins identified for a given gene cluster. **(D)**. Heatmap (one-way clustering using ward method) illustration of relative protein abundances for proteins encoded by the 18 antibiotic and secondary metabolite gene clusters. Color gradient represents a standardized score calculated per protein and white represents proteins that were not quantified in a particular medium.



iron scavenger and helps *Pseudomonas* species adapt to changing environments and niche colonization (Cornelis, 2010). Siderophores, in addition to iron scavenging, are also involved in the formation of complexes with other metals, antagonizing plant root microbes or facilitating antimicrobial activity (Loper, 1991). While the proteomics data show that *Pseudomonas* synthesize multiple antibiotics and SMs in both media, few antibiotic- and SM-related proteins were uniquely identified in R2A, which suggests this environment and community is more competitive against *Pseudomonas* sp. GM17 when compared to minimal media (**Figure 3.3C and 3.3D**). Interestingly, the relative abundance trends of proteins associated with the production of antibiotics and secondary metabolites also differed between media (**Figure 3.3D**).

Taken together, the proteomics analysis of *Pseudomonas* shows an increase in antibiotics, siderophore, and secondary metabolite production during the community selection process, indicating that *Pseudomonas* sp. GM17 is involved in antagonizing other microbial members. Pairwise-microbe interaction screen results reinforced these antagonistic properties of *Pseudomonas* by showing a presence of a zone of inhibition for most members of the community (**Supplemental Figure 3.4C**).

3.3.4 *Pantoea* sp. YR343 is effective at adapting to the changes in the community in R2A media

In the minimal media, *Pantoea* sp. YR343 represent less than 1% of stable community and most proteins identified are related to stress response (**Figure 3.4A and Supplemental Figure 3.5**). Unlike its behavior in minimal media, *Pantoea* sp. YR343 population size changed substantially throughout the R2A passages, with a large reduction in size after passage 2 followed by a continuous gain in population size until representing ~31% of the relative abundance in final stable community (**Figure 3.4A**). To understand how *Pantoea* sp. YR343 is adapting to the competitive or unfavorable environment and becoming a relatively abundant member of the community in R2A media, its respective proteome expression profile was extracted and analyzed (**Supplemental Table 3.4**).

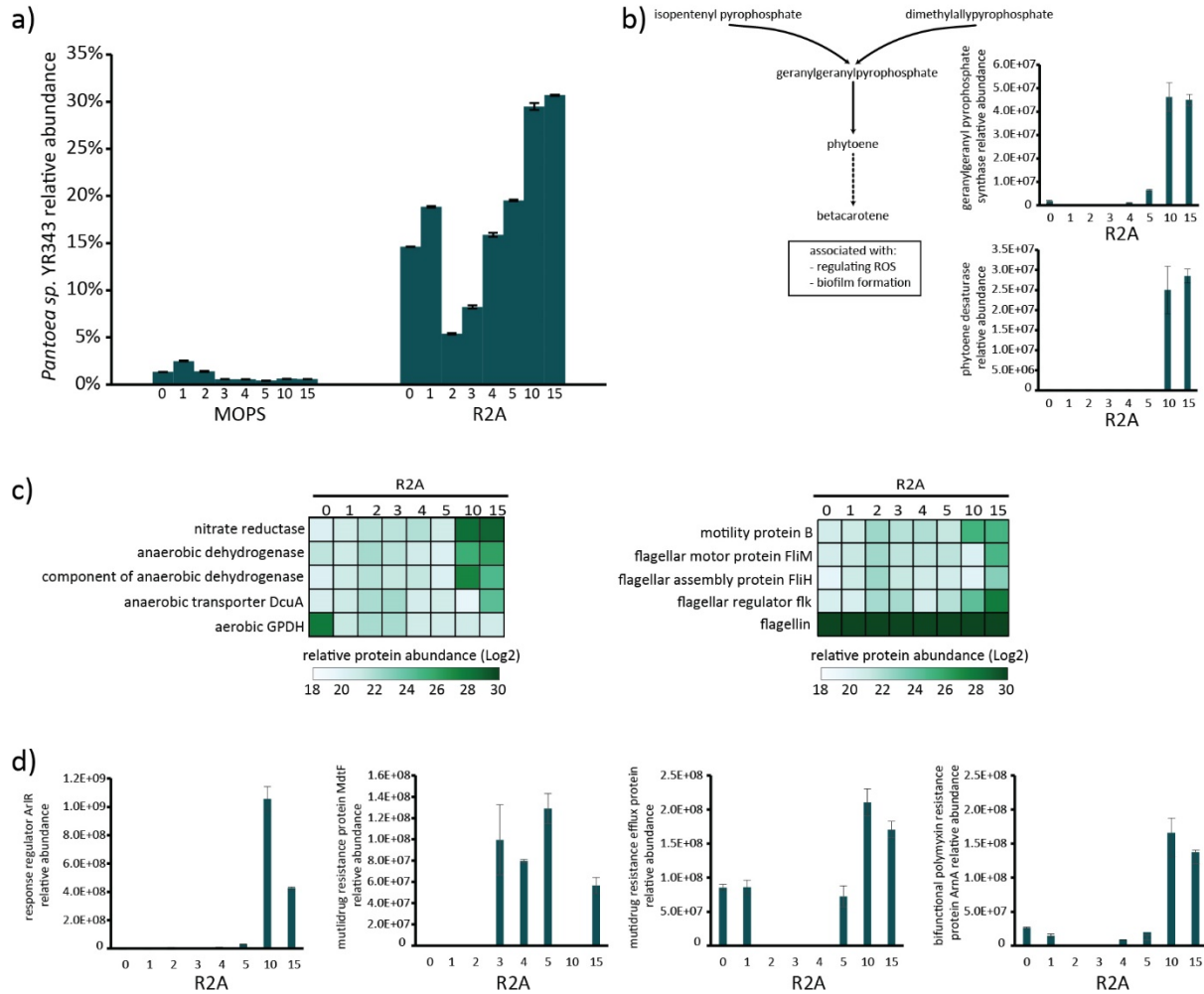


Figure 3.4: Metaproteomics analysis of *Pantoea* sp. YR343 functional behavior during community assembly.

(A) Relative abundance of *Pantoea* organism abundance in MOPS and R2A media based on metaproteomics data. (B) Relative abundance across R2A passages for geranylgeranyl pyrophosphate synthase and phytoene desaturase, two key proteins involve in carotenoid biosynthesis. Error bars represent standard error for each set of biological triplicates. The dashed arrow in the flow diagram represents multiple steps in biosynthesis. (C) Heatmap of relative protein abundance for proteins involved in aerobic/anaerobic respiration and motility. (D) Relative abundances for proteins associated with defense responses to antagonist behaviors. Error bars represent standard error for each set of biological triplicates.

Overall, *Pantoea* sp. YR343 was observed to effectively adapt to the changes in the community grown in R2A media by increasing the abundance of proteins related to stress responses, antibiotic resistance, motility, as well as shifting metabolism from aerobic to anaerobic processes. Increasing the relative abundances of proteins related to carotenoid biosynthesis, such as geranylgeranyl pyrophosphate synthase (J2V561) and phytoene desaturase (J2V5J0), helps to modulate membrane fluidity and aid in the *Pantoea*'s survival against oxidative stress, extremes pH, and toxins (**Figure 3.4B**) (Krinsky, 1989, Liaaen-Jensen and Andrewes, 1972, Edge, 1997). Population size increase coincided with changes in chemotaxis and motility-related proteins; a behavior used to find environmental niches for optimal survival and growth. In an environment seemingly limited in oxygen, notable changes were observed in proteins related to anaerobic growth such as nitrate reductase (Accession ID: J3HM00, J3BZ11) and anaerobic dehydrogenase (Accession ID: J2VI80, J3HP12), suggesting *Pantoea* sp. YR343 survival benefited from an augmented metabolism (**Figure 3.4C**).

In response to the antagonistic behavior of *Pseudomonas* sp. GM17, *Pantoea* sp. YR343 effectively increased the abundance of several crucial defense proteins that protect against inhibition (**Supplemental Figure 3.4B**). Relative protein abundances for several antibiotic/drug resistance proteins, such as multidrug resistance protein MdtF (J3HB42), response regulator ArlR (J3HGU6), bifunctional polymyxin resistance protein ArnA (J3BZA2) and multidrug resistance efflux pump (J3HPW6), increased across passages (**Figure 3.4D**). Multidrug resistance protein MdtF and response regulator ArlR provide protection against a broad range of antibiotic compounds. Similarly, bifunctional polymyxin resistance protein ArnA helps in the resistance against polymyxin that breaks up the bacterial cell membrane and cationic antimicrobial peptides.

3.3.5 *Bacillus* sp. BC15 sporulates as an adaptive response to nutritional competition

Based on protein content, *Bacillus* sp. BC15 was also a relatively abundant member in R2A media, but not in minimal media. Therefore, *Bacillus* sp. BC15 proteins were further investigated to identify which adaptive mechanisms were employed for these cells to thrive in R2A media (**Supplemental Table 3.5**).

Using a rank-based distribution of *Bacillus* proteins based on abundance, spore-related functions were among the most abundant proteins observed (**Supplemental Figure 3.6**). *Bacillus* species are known to cease growth and initiate sporulation under nutrient-limiting conditions and also in the presence of siderophores in the surrounding environment (Higgins and Dworkin, 2012). During sporulation, initially, an asymmetric cell division generates a smaller cell (forespore) and a larger cell (mother cell). The mother cell engulfs the forespore and mediates the development of the forespore into the spore through the production of the spore cortex and the inner and outer coat (Higgins and Dworkin, 2012). Upon spore maturation, the mother cell lyses, releasing the mature spore. This process is achieved in several developmental stages (**Figure 3.5A**). Metaproteomic results identified several *Bacillus* proteins from each stage of sporulation in R2A media, including stage 0 sporulation protein A (Spo0A; A0A1M7EFS7) (**Figure 3.5B**) (**Supplemental Table 3.5**). Spo0A is a master transcription factor that binds to the promoters and regulates gene expression, driving the sporulation events (Molle et al., 2003). Moreover, endospore proteins such as spore coat proteins (A0A1M7B6S6, A0A1M7DEY3 etc.), small acid-soluble proteins (A0A1M7C0T3, A0A1M6VBR5 etc.) were significantly upregulated in complex media (**Figure 3.5C**). Spore coat plays an important function in preventing spore degradation. Small acid-soluble proteins are found in the spore core and help to maintain the chromosomal DNA in a compact state (Higgins and Dworkin, 2012).

Next, we assessed whether *Bacillus* sp. BC15 is becoming more or less metabolically active across the R2A passages. Overall, ~45% of the *Bacillus* proteome expressed in early passages represents proteins associated with spores or sporulation processes (**Figure 3.5D**) and this shifts to less than 10% in passages 10 and 15. As the total abundance of spore-related proteins decrease with increasing passages, proteins associated with translation, ribosome biogenesis, and cellular metabolic and catabolic processes increase in the final passages (**Figure 3.5E**). In general, these results suggest that *Bacillus* sp. BC15 was not actively growing but rather existing in dormancy to survive during unfavorable conditions and persisted in the community by adapting differently until it could become more metabolically active, as represented by increased abundances in major cellular processes in the final passages.

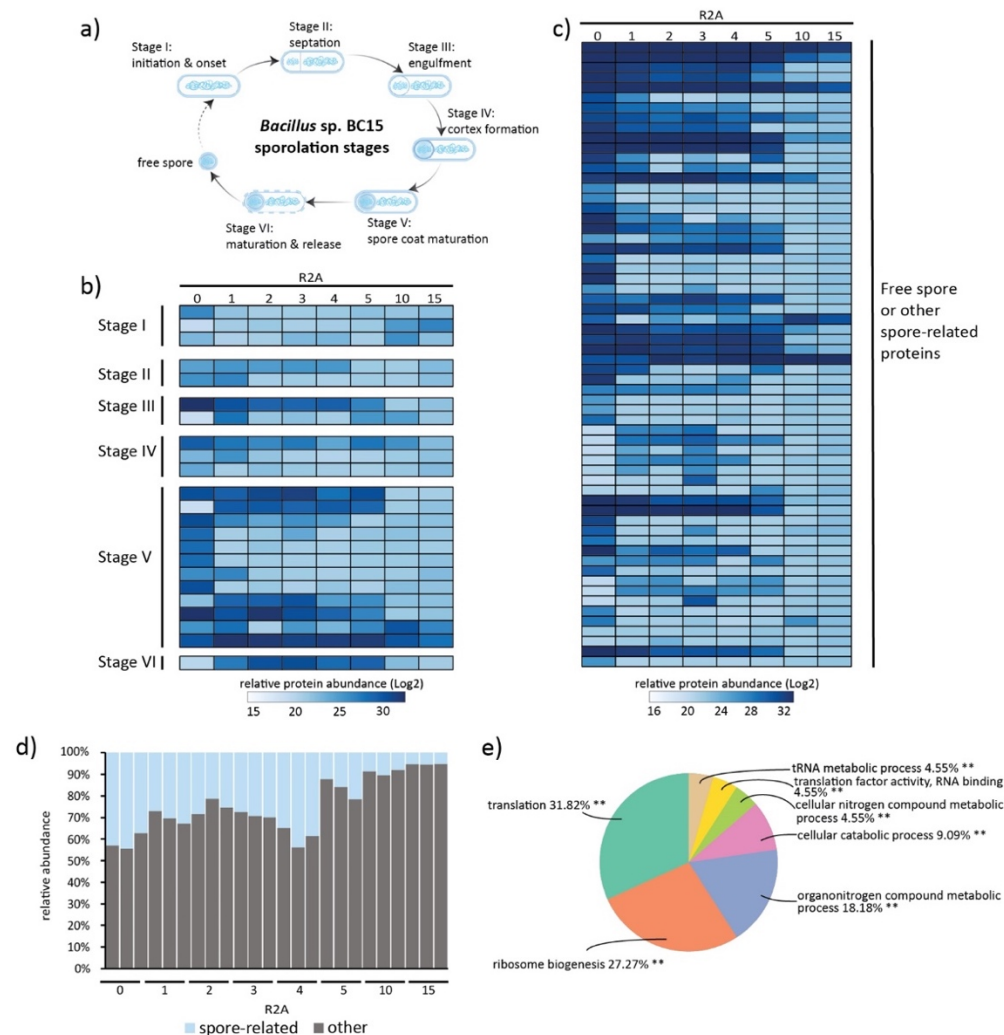


Figure 3.5: Metaproteomics analysis of *Bacillus* sp. BC15 functional behavior during community assembly.

(A) Developmental stages of sporulation in *Bacillus* sp BC15. Under environmental stress or nutrient limited conditions, *Bacillus* undergoes endospore formation which is accomplished across multiple stages of morphogenesis. Relative abundance of the proteins identified (B) across all stages and (C) free spores. Abundances shown are the average of three biological replicates. The color gradient represents the abundance in the Log2 scale. (D) Proteome-wide relative abundance distribution of spore-related proteins compared to all other proteins. (E) The pie chart shows the enriched biological processes (p-value<0.05) of “all other” proteins. Gene ontology enrichment analysis was done using ClueGO.

3.3.6 Post-translational modifications offer critical insights into how microbes are adapting to changing communities and environments

Post-translational modifications (PTMs) of proteins represents one of the most important yet understudied mechanism that microbes use to rapidly perceive and respond to changing conditions (Walsh et al., 2005, Mijakovic et al., 2016, Prabakaran et al., 2012, Macek et al., 2019). Recent advancements in metaproteomic experimental workflows now afford the ability to broadly study PTMs in bacterial isolates and communities (Li et al., 2014, Marlow et al., 2016, Zhang et al., 2016, Gupta et al., 2007, Brown et al., 2017). Therefore, we further interrogated the data to demonstrate that the analysis of PTMs offers new insights into the function and behavior of individual microbial community members as they respond to changing communities and surrounding environments.

In general, PTMs were identified on ~40% of proteins observed in this study (**Supplemental Table 3.6**). Modified proteins are expected to include both biologically relevant PTMs as well as those that are direct products of sample preparation procedures (e.g., carbamidomethylation on cysteine residues and methionine oxidation). Therefore, all observed PTMs were annotated as either a natural or artificial modification using a standardized nomenclature from the Unimod database as a guide (see Materials & Methods section) (Liu et al., 2013, Devabhaktuni et al., 2019). Following this step, we observed that ~10 % of proteins identified in this study were proteins that contained biologically meaningful PTMs.

Amongst the biologically relevant PTMs observed in this study, the most frequently occurring types of PTMs were methylations, dehydrations, and oxidations/hydroxylations. Previous studies have shown these modifications to be quite abundant in proteomes from bacterial isolates as well as environmental communities (Brown et al., 2017, Li et al., 2014, Zhang et al., 2016). The impact that these types of modifications have on proteins can be quite diverse but are frequently associated with altering protein-protein interactions or activity (Erce et al., 2012, Van Staalduinen and Jia, 2015, Ge et al., 2012). Based on annotations provided by cluster of orthologous groups (COG) categories, PTMs observed across this study impact many prominent cellular processes, including translation, ribosomal structure and biogenesis, energy production and conversion, protein

turnover, and chaperon functions as well as amino acid metabolism and transport. Proteome-wide analysis of an abundant member of the community, such as *Pseudomonas* sp. GM17, suggests that changes in communities or the environment do not affect which types of PTMs are predominant or what cellular processes are being modified (**Supplemental Figure 3.7 and 3.8**). Shifting our analysis from a proteome-wide to a protein-centric perspective allowed us to identify PTMs that occur at conserved residues in bacterial proteins (**Supplemental Table 3.7**) (Macek et al., 2019). This examination revealed two types of modifications: proteins having static, seemingly always modified positions; and proteins with positions that are dynamic and reversible. Evidence for site-specific modifications that are static are interesting because they suggest a role critical to a protein's primary function. In this study, one example is a β -methylthiolation modification in ribosomal protein S12 (**Supplemental Figure 3.9**). This modification is localized on an aspartic acid residue universally conserved within bacteria and presumed to be structurally or functionally important because substitutions at this position are lethal (Kowalak and Walsh, 1996, Running and Reilly, 2009, Carr et al., 2005, Noeske et al., 2015, Strader et al., 2004). All bacterial ribosomal S12 proteins sequenced across this entire study contained this modification, reinforcing previous findings that this modification is a static feature of the ribosome S12 protein. Beyond this well-known PTM, we also observed other proteins that have evidence for static modifications and warrant further interrogation to evaluate their essentialities (**Supplemental Table 3.8**).

Unlike static modifications, dynamic modifications are greatly influenced by experimental design and require adequate sampling to identify their roles. The identification and tracking of protein positions that are reversibly modified not only provides insights into a changing functional or structural state of a particular protein, but an extended understanding of how the organisms is perceiving or altering its behavior to changing conditions. For example, we observed a dynamic methylation on a lysine position in the elongation factor thermo unstable protein (EF-Tu) that could imply an organism is experiencing nutrient deprivation. EF-Tu proteins are one of the most conserved and abundant proteins expressed by bacteria (Furano, 1975, Harvey et al., 2019). First reported in *E. coli*, the EF-Tu protein is hypermethylated at a lysine residue at position 56 under nutrient deprivation conditions and this modification is expected to reduce protein synthesis by affecting the ability of the protein to bind and hydrolase GTP (Young and Bernlohr, 1991, Kraal

et al., 1999, VAN NOORT et al., 1986, Polevoda and Sherman, 2007). Across this study, EF-Tu was predominately methylated at the conserved lysine position in *Pantoea* (Accession ID. J3HLV4) and *Rhizobium* (Accession ID. J1SLR3) (**Figures 3.6A and 3.6B**), however this modification was absent in EF-TU proteins observed in *Pseudomonas*, *Bacillus*, *Sphingobium*, *Caulobacter*, and *Duganella*. Based on the relative abundance of these microbes across different passages, we hypothesize that *Pseudomonas* maintains a relatively active protein synthesis rate across the entire experiment, whereas others like *Pantoea* and *Rhizobium* have a PTM-modified EF-Tu with reduced activity due to a lack nutrients. The lack of a modification in this protein for other members like *Bacillus* and *Sphingobium* is likely due to the presence of arginine residues at this position (**Figure 3.6C**), which seems to be the main reason behind the absence of this modification in some bacteria (Kraal et al., 1999). Although, it is possible that the EF-Tu methyltransferase does not function in these microbes. For other members like *Duganella* and *Caulobacter*, we believe that the detection limits of our instrumentation may preclude the observation of the methylated Lys residue, but we cannot completely discard the presence of this modification on their proteins.

3.3 Discussion

In addition to plant-associated lifestyles, members of the rhizospheric microbiome adapt to local changes in environmental variables, such as resource availability, that shape the growth of individual populations and the interactions between species. Stochastic changes in the relative abundance of individual members have been demonstrated to have large downstream effects on community composition that can alter ecological dynamics (Zhou et al., 2013). The difference in stable microbial community composition observed in this study in different growth media could be due to the change in the chemical environment which largely affects the growth rate of microbial members. Interestingly, discrete community transition phases were observed across the passages in R2A media where the membership, as well as their relative abundance fluctuate suggesting the dynamic interaction between microbial members during the community assembly and selection process. Although this study provides the foundational understanding of rhizospheric-associated microbial community, dynamic interaction observed in the synthetic media may not be exact in the rhizospheric environment where factors like root exudates, other microbes effect the

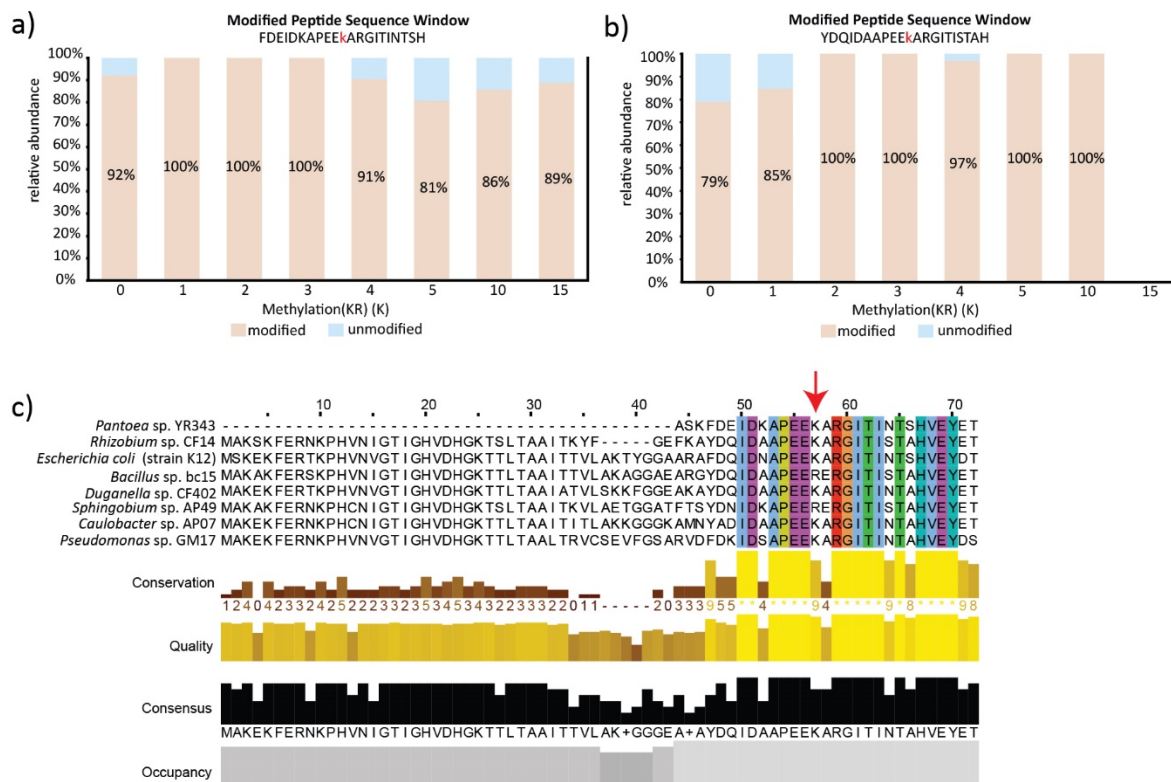


Figure 3.6: Dynamic regulation of methylation modification in lysine (K) residue of EF-Tu in nutrient-deprived organisms.

Percentage abundances of K-methylated modified and unmodified peptides in the Elongation factor Tu proteins of (A) *Pantoea* sp. YR343 and (B) *Rhizobium* sp. CF142 in R2A media. (C) Multiple sequence alignment was performed for all Elongation Factor Tu proteins identified by metaproteomics in this study. The lysine amino acid position where the modification occurs in this study is highlighted with a red arrow. Unlike other organisms, *Bacillus* sp. BC15 and *Sphingobium* sp. AP49 have an arginine (R) residue at this position and no methylation modifications were observed for these protein sequences. For reference, the Elongation factor TU sequence of *E. coli* (strain K12) was added for comparison.

community assembly and selection process. Microbes interact with each other in a variety of ways, these interactions can be positive such as mutualism, synergism, and commensalism; or be negative such as antagonism, parasitism, predation, or competition (Weiland-Brauer, 2021). Metaproteomics results allowed us to understand the microbe-microbe interactions in the DefCom and the molecular mechanisms driving the community level outcome.

Our metaproteomics analysis identified *Pseudomonas* sp. GM17 to be the dominant member of the community regardless of the chemical environment and environmental disturbances. The genus *Pseudomonas* are ubiquitous γ -proteobacteria well known for extreme versatility and adaptability (Sitaraman, 2015). The higher growth rate of *Pseudomonas* sp. GM17 in both media could have provided an initial competitive advantage that led to its dominance. Although *Pseudomonas* sp. GM17 was one of the fast-growing microbes, other microbes such as *Sphingobium* sp. AP49, *Paraburkholderia* sp. BT03 also has a similar growth rate (Wang, 2021). However, these microbes did not represent a significant portion of the stable community. The proteome-wide analysis revealed that *Pseudomonas* expressed multiple proteins involved in the biosynthesis of siderophores that provide fitness advantage as well as antibiotics and secondary metabolites which are known to have biocontrol activity and facilitate antagonistic behavior. During antagonistic interaction the first population that produces inhibitory substances are unaffected or even gain a competition advantage and survive in the habitat while other populations get inhibited. High growth rate as well as antagonistic behavior due to production of antibiotics and secondary metabolites helped in *Pseudomonas* sp. GM17 dominance by outcompeting or preventing the growth of other microbes (Becker et al., 2012).

As members of the community are challenged with unfavorable conditions created either intentionally by other members of the community (e.g., production of antibiotics from *Pseudomonas*) or unintentionally through the collective activity of community metabolism (e.g., limited nutrients, lower oxygen levels, etc.), microbial persistence requires the expression of necessary coping mechanisms. The proteome-wide analysis of *Pantoea* revealed a significantly higher proportion of proteins related to stress responses in minimal media but not necessarily of proteins related to coping mechanisms (**Supplemental Figure 3.5**). Instead, in complex media, *Pantoea* sp. YR343 still experienced the stressful environment created by *Pseudomonas* and other

members of the community but it was able to cope by increasing the abundance of proteins related to carotenoid biosynthesis, motility, and antibiotics resistance. Carotenoids are sterol analogs and studies have shown its ability to modulate membrane fluidity and role in the formation of membrane domains (Lopez and Kolter, 2010, Kumar et al., 2019b). Carotenoids aid in the survival of cells under harsh conditions, such as oxidative stress, extreme pH, and presence of toxins (Krinsky, 1989, Liaaen-Jensen and Andrewes, 1972, Edge, 1997). In *Pantoea* sp. YR343, carotenoid is required to regulate the sensitivity to reactive oxygen species, secretion, biofilm formation, and rhizosphere survival (Bible et al., 2016). The defect in the gene involved in synthesizing carotenoids has been shown to affect the microbe's growth, biofilm formation, and phytohormone production (Bible et al., 2016). Similarly, motility directed by chemotaxis is an important means by which microbes avoid adverse conditions in their environment (de Weert et al., 2002). Microbes sense the presence of nutrients or other harmful chemicals in the environment with the help of chemotaxis related proteins which play an important role in environmental adaptation. Moreover, *Pantoea*'s ability to develop resistance against antibiotic and toxic compounds further helped to adapt in the changing environment. *Pantoea* is a facultative anaerobe that can grow in the presence and absence of oxygen, and our proteomics results also showed the shifting from aerobic to anaerobic respiration/metabolism as one of the coping mechanisms of *Pantoea* in R2A media.

While *Pantoea* effectively adapts to the changes in the environment of R2A media in the DefCom by increasing the abundance of proteins related to antibiotic resistance and motility as well as shifting metabolism from aerobic to anaerobic processes, *Bacillus* on the other hand, sporulated as an adaptive response. Endospores formed from the process of sporulation are morphologically distinct, metabolically dormant, and environmentally resistant, capable of surviving extreme environments (Higgins and Dworkin, 2012, Tan and Ramamurthi, 2014). The identification of endospore proteins and sporulation proteins from various developmental stages implies that although *Bacillus* sp. has been identified as a key member of a stable community in R2A media, it is experiencing severe nutrient stress. A recent study has shown that siderophores can also act as interspecies cues that alter cellular development and accelerate sporulation in *Bacillus subtilis* (Grandchamp et al., 2017). Our results have shown that siderophore-related proteins from

Pseudomonas were highly abundant, thus it is plausible that *Bacillus* sp. is sporulating in R2A media in the response of siderophores produced by the former.

Bacterial members within microbial communities adapt to their environmental conditions not only by modulating the abundance of the proteins they produce, but also by regulating the functions of these molecules. PTMs can modulate protein activity, conformation states, localization, and interactions. PTMs of proteins represents one of the most important, yet understudied mechanism that microbes use to rapidly perceive and respond to changing conditions (Walsh et al., 2005, Mijakovic et al., 2016, Prabakaran et al., 2012, Macek et al., 2019). PTM information allows an extended understanding of how an organism is perceiving or regulating its behavior to changing conditions. For example, the dynamic lysine methylation identified here in EF-Tu proteins has been suggested to reduce the rate of protein synthesis by affecting the ability of the protein to bind and hydrolase GTP under nutrient deprivation conditions (Young and Bernlohr, 1991, Kraal et al., 1999, VAN NOORT et al., 1986, Polevoda and Sherman, 2007). Proteome-wide as well as protein-centric (i.e., static and dynamic) modifications identified in this study provide the proof of concept that optimized bioinformatics pipelines and high-resolution mass spectrometry not only affords the ability to broadly characterize PTMs in a biological system but also provides a level of sensitivity capable of revealing regulatory mechanisms influencing the activity of single proteins and we expect to continue mining this wealth of information to find novel and/or less-studied PTM regulatory mechanisms.

3.4 Conclusions

The understanding of intricate ecological interactions between microbes, the extent to which inter-microbial interactions shape microbial community, and the comprehensive understanding of the mechanisms driving the microbe-microbe interaction represent key aspects of microbial ecosystem, yet these remain not well understood. The reductionist approach presented here provided an informative and useful opportunity to study naturally occurring complex microbial communities in a controlled and tractable laboratory setting. Together with high-resolution metaproteomics, an accurate assessment of microbial population sizes was obtained for this 10-member DefCom while simultaneously providing detailed understanding of proteome- and

protein-level changes that help elucidate the biological process underpinning community assembly across two distinct growth conditions. By identifying building blocks of complex microbial behaviors that exist in defined communities, mechanisms underlying microbe-microbe offer some degree of prediction that can be further tested as more natural settings become integrated into experimental design. Moving forward, we envision data-rich metaproteomics datasets will become an integral source of information in modeling microbe-microbe interactions and a valuable datatype in predictive biology and discovery-based research environments, like The Department of Energy Systems Biology Knowledgebase (KBase) (Arkin et al., 2018).

Abbreviations

2D: Two-Dimensional; ACN: Acetonitrile; BGC: Biosynthetic Gene Clusters; cLV: Compositional Lotka-Volterra; COG: Cluster of Orthologous Groups; DefCom: Defined Community; DTT: Dithiothreitol; EF-Tu: Elongation Factor Thermo Unstable; FBA: Flux Balance Analysis; gLV: Generalized Lotka–Volterra; HCD: Higher-Energy C-Trap Dissociation; IAA: Iodoacetamide; KBase: Department of Energy Systems Biology Knowledgebase; LC: Liquid Chromatography; LOESS: Local regression; MOPS: Potassium morpholinopropane sulfonate; MS: Mass spectrometry; NRPS: Non-Ribosomal Peptide Synthetase; OD: Optical Density; ODE: Ordinary Differential Equations; PCA: Principal Component Analysis; Pd: *Populus deltoides*; PSM: Peptide-spectrum match; PTM: Post-Translational Modifications; R2A: Reasoner's 2A media; RP: Reverse Phase ; SCX: Strong Cation Exchange; SDS: Sodium Dodecyl Sulfate; SM: Secondary Metabolites; Spo0A: Stage 0 Sporulation Protein A

CHAPTER 4: CROSS-SPECIES PHOSPHOPROTEOME PROFILING REVEALS A KINASE-INDUCED SIGNALING MECHANISM USED BY PLANTS TO PERCEIVE BENEFICIAL FUNGUS

A version of this chapter will be submitted for publication by Him K. Shrestha:

Him Shrestha, Dr. Tao Yao, Dr. Zhenzhen Qiao, Dr. Yali Sun, Dr. Wellington Muchero, Dr. Robert L. Hettich, Dr. Jin-Gui Chen, and Dr. Paul E. Abraham (2022). Cross-species phosphoproteome profiling reveals a kinase-induced signaling mechanism used by plants to perceive beneficial fungus. Targeted for *The Plant Journal*. Manuscript in preparation.

This is the collaborative work with researchers from Plant-microbe interface (PMI) group at Oak Ridge National Laboratory i.e., Dr. Tao Yao, Dr. Zhenzhen Qiao, Dr. Yali Sun, Dr. Wellington Muchero, and Dr. Jin-Gui Chen. These collaborators conceptualized the original study, carried out plant-fungus growth and sampling, and helped with follow-up experiments. This project is the extension of our previous publication: Dr. Zhenzhen Qiao, Timothy B Yates, **Him K Shrestha**, Nancy L Engle, Dr. Amy Flanagan, Dr. Jennifer L Morrell-Falvey, Dr. Yali Sun, Dr. Timothy J Tschaplinski, Dr. Paul E Abraham, Dr. Jessy Labbé, Dr. Zeng-Yu Wang, Dr. Robert L Hettich, Dr. Gerald A Tuskan, Dr. Wellington Muchero, and Dr. Jin-Gui Chen (2021). Towards engineering ectomycorrhization into switchgrass bioenergy crops via a lectin receptor-like kinase. *Plant Biotechnology Journal*. 19(12), 2454-2468.

Lectin receptor-like kinase mediated signaling cascades during ectomycorrhization were not well understood. Him K Shrestha is leading a project on Phosphoproteomics analysis to understand the lectin receptor-like kinase regulated signaling mechanism. Him K Shrestha's contributions in this project are: 1) Phosphoproteomics sample processing, 2) Mass spectrometry measurement and data analysis, 3) Identifying the follow up candidates and experiments, 4) Manuscript draft writing and figure generation.

4.1 INTRODUCTION

Beside the inter-microbial interactions discussed in chapter 3, plant-microbe interaction is another key factor of plant holobiont studies. To better understand this interkingdom interaction, it is crucial to understand the response of both organisms involved. Plant tissues, especially in the rhizospheric region, are in close proximity to diverse microbial communities (Bulgarelli et al., 2012, Tringe et al., 2005, Lundberg et al., 2012, Edwards et al., 2015). Plants and their associated microbes have both adapted to use this intimate association for their mutual benefit. Microbes help

to convert critical nutrients such as phosphate and nitrate to a more usable form for plant uptake while plants supply the carbon source derived from photosynthesis to the microbes (Long, 1989, Zhang et al., 2009, Bais et al., 2006). Plant-microbe mutualistic interactions contribute to plant growth, development and phenotype via pathogen resistance, water retention, nutrient uptake and synthesis of growth-promoting hormones (Mendes et al., 2011, Berendsen et al., 2012, Lu et al., 2018, Cook et al., 1995). This host-microbe mutualistic interaction is a dynamic process shaped by complex chemical and molecular events that have remained to be fully understood.

Populus sp. (poplar) is a dominant perennial woody plant with wide distribution in North America (Bryant et al., 2020). Poplar is an important biofuel feedstock due to their rapid growth, ease of propagation, and adaptability to a range of planting sites (Bryant et al., 2020). Black cottonwood poplar (*Populus trichocarpa*) roots are colonized by the beneficial fungus *Laccaria bicolor* (Labbe et al., 2019). During the beneficial plant-fungus interaction, fungal hyphae penetrate the root and grow in between the apoplast space of the epidermal and cortex cells. This forms a network called the Hartig net (Blasius et al., 1986, Labbe et al., 2019, Roth-Bejerano et al., 1990, Buscot et al., 2000). The fungus also forms a hyphal layer in the root tip which forms mantle structures (Blasius et al., 1986, Labbe et al., 2019). Through these physical structures, minerals, water, and nutrients are exchanged between plants and fungi.

Previously research in poplar discovered a gene deletion event of a specific G-type lectin receptor-like kinase (PtLecRLK1) that is associated with decreased root interaction between poplar and *L. bicolor* (Labbe et al., 2019). In an effort to understand the functional role of PtLecRLK1, the *PtLecRLK1* gene was expressed in non-host plants. Interestingly, transgenic expression of *PtLecRLK1* into non-host plants, such as switchgrass and Arabidopsis, display interstitial hyphal growth and Hartig net-like structures induced during *L. bicolor* colonization (Labbe et al., 2019, Qiao et al., 2021). Moreover, the defense responses are suppressed in these non-host plants upon exposure to *L. bicolor*, a key mechanism initiating fungal colonization (Labbe et al., 2019, Qiao et al., 2021). These findings imply the potential of *PtLecRLK1* in engineering mycorrhizal symbiosis for improving plant health or productivity.

PtLecRLK1 is a G-type lectin receptor-like kinase. In plants, G-type LecRLKs has been shown to be involved in various biological processes including the perception of environmental cues, pathogen detection and defense, and plant-microbe interactions (Trontin et al., 2014, Deb et al., 2014, Sun et al., 2020, Bellande et al., 2017, Chen et al., 2006). In Arabidopsis, G-type LecRLKs called S-domain receptor kinase regulates lateral root development under phosphate starvation conditions (Deb et al., 2014). Other G-type LecRLKs have also been shown to be involved in stress signaling upon perceiving plant pathogen-produced molecules and contribute to plant defense (Trontin et al., 2014). Besides this, several other G-type LecRLKs from Arabidopsis and other plant species are involved in plant-pathogen interactions (Wang et al., 2022, Guidarelli et al., 2014, Liu et al., 2015, Cole and Diener, 2013, Cheng et al., 2013, Gilardoni et al., 2011, Kim et al., 2009, Ranf et al., 2015, Chen et al., 2006). There are also evidences of G-type LecRLKs involved in the plant-microbe symbiosis (Navarro-Gochicoa et al., 2003, De Hoff et al., 2009, Bouwmeester and Govers, 2009).

G-type LecRLKs contain N-terminal mannose-binding lectin domain followed by S-locus glycoprotein domain, a PAN-like domain, and EGF domain in the extracellular region (Sun et al., 2020, Bellande et al., 2017). The lectin domain is about 100 amino acid residues and contains a β -barrel structure with three subdomains of four-stranded β sheets, each harboring a unique site for carbohydrate binding (Sun et al., 2020, Bellande et al., 2017, Van Damme et al., 2008). Although initially considered mannose-specific, this domain in some RLKs has also shown to have a strong affinity toward oligomannosides and high-mannose N-glycans (Van Damme et al., 2008). The S-locus glycoprotein domain is about 100 residues-long and plays a role in self-incompatibility (Bellande et al., 2017). The PAN motif of this protein is about 80 amino acids. PAN motif can contribute to protein-protein interaction and/or protein-carbohydrate interaction, while the epidermal growth factor (EGF) domain has been linked to be involved in disulfide bond formation (Naithani et al., 2007). Overall, the extracellular domains allow the RLKs to recognize a great range of signals.

The intracellular region of G-type LecRLKs contains the kinase domain. The C-terminus tail of the kinase domain contains a conserved sequence and is reported to be important for the regulation of RLKs-mediated pathways (Sun et al., 2020, Weiner and Zagzag, 2000). Kinases mediate the

transduction of complex cellular signals through the phosphorylation of the proteins. During phosphorylation, the phosphate group from ATP is transferred to a targeted amino acid mainly serine, threonine, and tyrosine by kinases (Urban, 2022). The modified site can be recognized by a specific domain of the proteins that bind to the phosphorylated residue and signal the downstream signal cascade (Urban, 2022). In contrast, the phosphatase family of proteins can remove the phosphorylation modification from proteins (Low et al., 2021, Urban, 2022). The activity of kinases and phosphatases are responsible for generation of a reversible phosphorylation modification. Phosphorylation is a highly dynamic post-translational modification (PTM) that regulates cell signaling, protein assembly, localization, function, stability, and protein-protein interactions (Ardito et al., 2017, Vyse et al., 2017, Kim et al., 2004). The elaborative downstream protein signaling through phosphorylation modification can lead to the regulation of various genes in a highly sensitive and specific manner. A comprehensive understanding of the phosphorylation signaling cascade facilitating beneficial plant-microbe interaction by specific kinase could enable us to engineer systems that can enhance plant productivity and fitness in marginal lands and address the growing challenges of bioenergy production and environmental remediation.

Phosphoproteomics by mass spectrometry has become a method of choice that provides unprecedented ability to characterize and quantify phosphorylated proteins and peptides in complex samples which aid in mapping intracellular kinase signaling networks (Urban, 2022, Riley and Coon, 2016, Olsen and Mann, 2013). Therefore, it can be a powerful technique to understand the phosphorylation signaling cascade associated with the recognition of *L. bicolor* through PtLecRLK1 expression. However, due to the low stoichiometry of regulatory phosphorylation sites, the high complexity of the phosphoproteome, and the low abundance of phosphorylated peptides compared to non-phosphorylated peptides, achieving deeper and more sensitive measurement to cover amply phosphorylation sites of interest is challenging (Urban, 2022). Optimized sample preparation protocols such as phosphopeptide enrichment with high-performance LC-MS/MS methodology are desired. In this study, we utilized phosphopeptide enrichment along with high-performance LC-MS/MS methodology to i) map the PtLecRLK1-induced signaling cascade that facilitates plant-fungal mutualistic interaction in the non-host species, ii) identify the direct interacting partner of PtLecRLK1 that helps to transduce signal

downstream of *PtLecRLK1* and iii) compare transgenic non-host species to see if the mechanism and interacting partners are conserved cross-species.

4.2 Materials and Methods

4.2.1 Plant and fungal growth

In the potato dextrose liquid medium, free-living mycelium of *L. bicolor* S238N was grown for 7 days. In parallel, plant material was grown and prepared for inoculation with fungus. Tillers of switchgrass and poplar were pre-rooted in water. After ~2 weeks, rooted tillers were moved to dampened soil with *L. bicolor* liquid inoculum. This co-culture was grown for 2 months in the greenhouse. In the greenhouse, plants were grown under a 16-hour light/8-hours dark photoperiod at 24°C. At least three independent biological replicates of inoculated root samples for each of the transgenic and wild type (WT) switchgrass and poplar were collected two months post inoculation. Plants roots were rinsed, collected and flash frozen in liquid nitrogen. Root samples were then homogenization by the TissueLyzer (Invitrogen). The root samples were then processed for mass spectrometry analysis.

4.2.2 Proteomics Sample Preparation

Wild type (WT) and transgenic switchgrass and poplar lines expressing *PtLecRLK1* were grown in the presence of *L. bicolor*. Root tissues were collected for mass spectroscopy with at least three biological replicates. Each replicate was flash-frozen and ground under liquid nitrogen. Samples were solubilized in 3-5 mL of 4% SDS in 100 mM ammonium bicarbonate (ABC) buffer. Samples were lysed by boiling at 90C for 5 minutes followed by sonication. After sonication, samples were centrifuged at 5,000 g for 10 minutes. Next, the supernatant was pipetted for protein extraction. The supernatant was reduced with 10mM dithiothreitol by incubating for 30 min. Then, alkylated with 30 mM iodoacetamide (IAA) at room temperature for 15 min in dark. Protein was then extracted using a chloroform-methanol extraction protocol as previously described (Jiang et al. 2004). The isolated proteins were reconstituted in sodium deoxycholate (SDC) solution made up of 4% SDC in 100 mM ABC. Protein concentration was measured using nanodrop ((Thermo Scientific). Proteins were then digested with two separate and sequential aliquots of sequencing

grade trypsin (Promega) in the ratio of 1:75 trypsin to sample protein ratio for three hours followed by overnight. After digestion, SDC was removed by precipitating with 1% formic acid. The digested peptide was further clean for SDC using ethyl acetate wash. Peptides were then lyophilized via SpeedVac (Thermo Scientific). The lyophilized peptide was desalted on Pierce peptide desalting spin columns (Thermo Scientific) as per the manufacturer instructions. After desalting, peptides were again vacuum-dried with a SpeedVac and then resuspended in 0.1% formic acid. Peptide concentrations were then measured using nanodrop (Thermo Scientific) and 15 µg of each sample was allocated for proteomics measurement. The remaining peptides were vacuum-dried with a SpeedVac and then resuspended in the manufacturer-recommended buffer for phosphopeptide enrichment. Phosphopeptide enrichment was carried out as per the manufacturer's instructions. Finally, the phosphopeptides were vacuum-dried with a SpeedVac and then resuspended in 0.1% formic acid for phosphopeptide measurement.

4.2.3 LC-MS/MS analysis

All samples were analyzed on a Q Exactive Plus mass spectrometer (Thermo Scientific) coupled with an RSLCnano UHPLC system (Thermo Scientific). Phosphopeptides were separated on a MudPIT system with biphasic precolumn (SCX-RP) coupled to an in-house-pulled nanospray emitter of 75 µm inner diameter packed with 25 cm of 1.7 µm of Kinetex C18 resin (Phenomenex). For the phosphoproteome measurement, a single 1 µg injection of phospho-enriched peptides was loaded and analyzed across salt cuts as described previously (Qiao et al., 2021) in a 180 min split-flow (300 nL/min) gradient. MS data were acquired with the Thermo Xcalibur software using the top 10 data-dependent acquisition.

4.2.4 Proteome Database Search

All MS/MS spectra collected for switchgrass were processed in Proteome Discoverer, version 2.4 with MS Amanda and Percolator. Spectral data were searched against the *P. virgatum* and *L. bicolor* reference proteome database from DOE Joint Genome Institute (JGI) downloaded in May 2020 to which transgenic sequences and common laboratory contaminants were appended. In MS Amanda of Proteome Discoverer, following parameter were set up to derive fully tryptic peptides:

MS1 tolerance = 10 ppm; MS2 tolerance = 0.02 Da; missed cleavages = 2. Carbamidomethyl (C, + 57.021 Da) was set as static modification; and oxidation (M, + 15.995 Da) as dynamic modifications. Besides oxidation modification, phosphorylation (STY) (+79.966 Da) was also added as dynamic modifications. The percolator was utilized for the false discovery rate. FDR threshold was set to 1% at the PSM and peptide levels. These peptides were then quantified based on their chromatographic area-under-the-curve (AUC) and mapped to their respective proteins. AUCs were summed to estimate the abundance at the protein-level.

4.2.5 Data Analysis

For differential abundance analysis of proteins and phosphorylated proteins, the protein table was exported from Proteome Discoverer. Proteins were filtered to remove stochastic sampling. All proteins present in 2 out of 3 biological replicates in any condition were considered valid for quantitative analysis. Data was Log₂ transformed, LOESS normalized between the biological replicates and mean-centered across all the conditions. Missing data were imputed by random numbers drawn from a normal distribution (width = 0.3 and downshift = 2.8 using Perseus software (<http://www.perseus-framework.org>). The resulting matrix was subjected to student T-test to assess differentially abundant proteins between the transgenic and wild type experimental groups.

4.2.6 GO and Functional Pathway Analysis

Arabidopsis thaliana homologs for each protein are provided in the JGI annotation of switchgrass (*Panicum virgatum* v5.1, Phytozome v11.0). Since the switchgrass is poorly annotated, the *A. thaliana* database was utilized using Arabidopsis homolog of differentially abundance switchgrass phosphoproteins for both GO and KEGG analysis. The enrichment analysis was performed in Cytoscape plug-in software ClueGO (Bindea et al., 2009). For both analyses, the GO or pathway were considered enriched if they pass significance threshold of $p < 0.05$ and Benjamini Hochberg FDR correction. The GO evidence code was set to all experimental and kappa score to 0.4. All other parameters were set to default.

4.2.7 Network analysis of switchgrass phosphoprotein interactions

The differentially abundant phosphoproteome interaction map was created using Cytoscape 3.7.1 (www.cytoscape.org) (Shannon et al., 2003). The Arabidopsis homolog of differentially abundant phosphoproteome switchgrass phosphoproteins were imported into Cytoscape and protein interactions information were retrieved using the string network node with a confidence threshold >0.7 (Doncheva et al., 2019). The cGMP node with high connectivity was selected and diffused to get top 10th percentile of most relevant nodes. The subnetwork created with the diffusion output rank was then interrogated for functional enrichment using the Cytoscape. Some of the significantly enriched KEGG pathway proteins were highlighted with different broader color.

4.2.8 Bimolecular fluorescence complementation assay (BiFC assay)

BiFC assay was performed in *Populus* protoplasts as described by Zhang et al., 2020 (Zhang et al., 2020b). In brief, the CDSs of PtLecRLK1 and its substrate candidate proteins were cloned into CFP^c (pUC119-CD3-1068) and VENUSⁿ (pUC119-CD3-1076) vectors through Gateway cloning, respectively. 10 mg of CFP^c-PtLecRLK1 plasmids and 10 mg of VENUSⁿ-Substrate plasmids were co-transfected in *Populus* protoplasts. After 18-20 hours dark incubation, the reconstructed YFP signals were detected by a Zeiss LSM 710 confocal microscope. ZEN software was used for image processing.

4.3 Results

4.3.1 Assessment of switchgrass phosphoproteomics results reveal distinct phosphorylation events between fungus inoculated transgenic and WT plant

To uncover how *PtLecRLK1* regulates beneficial plant-fungal interaction, we performed phosphoproteomics analysis in transgenic (host) and wild-type (non-host) switchgrass (**Figure 4.1A**). Across the experimental conditions, we identified 284,588 peptide spectrum matches (PSMs), out of which 75% had phosphorylation evidence (**Figure 4.1B**). These PSMs were mapped to 5,140 phosphopeptides in 4,469 unique modification sites (**Figure 4.1B**). A majority (87%) of those sites belong to amino acid serine, 12% belong to threonine, and the remaining

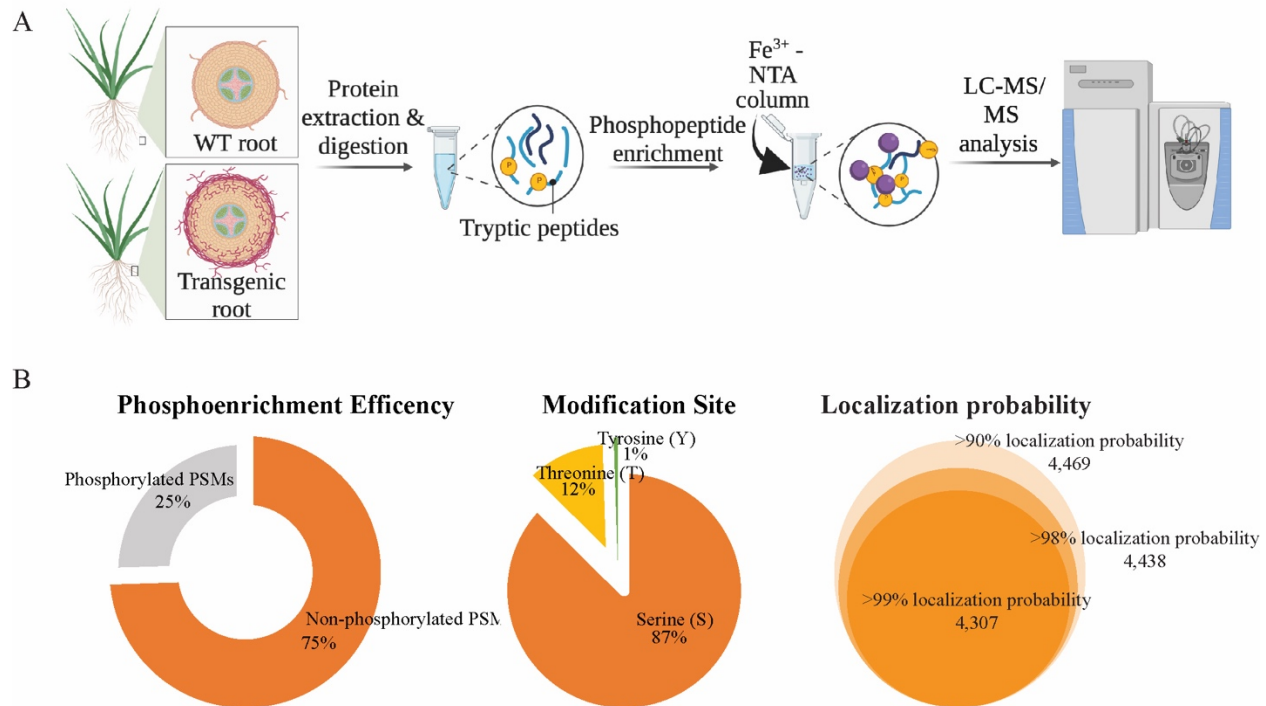


Figure 4.1: Experimental design and qualitative assessment of phosphoproteomics result from the root samples inoculated with *L. bicolor*.

(A) Wild type and Transgenic (expressing PtLecRLK1) roots co-cultured with *L. bicolor* were collected. Proteins were extracted, digested and peptides were enriched using Fe^{3+} NTA columns. Enriched phosphopeptides were measured in a high-resolution LC-MS/MS system. **(B)** Phosphopeptide enrichment efficiency, distribution of phosphosite and localization probability of phosphorylation modification detected in our study. Most PSMs identified were phosphorylated. Most modified amino acid was serine, and the majority of the modified site has localization probability greater than 99%.

portions belong to tyrosine (**Figure 4.1B**). Most modification sites had a localization probability score of >90% (**Figure 4.1B**).

We then aimed to characterize the phosphorylated proteins regulated by PtLecRLK1 during switchgrass-*L. bicolor* interaction. To that end, we utilized Welch's t-test with FDR correction ($q < 0.05$) and absolute $\log_2$ fold change > 1 to identify differentially abundant phosphopeptides in *L. bicolor* inoculated transgenic switchgrass compared to wild type. The quantitative analysis identified 1,260 differentially abundant phosphopeptides between *L. bicolor* inoculated transgenic and wild type (**Figure 4.2A**). Of these, 610 phosphopeptides were significantly upregulated in *L. bicolor* inoculated transgenic line compared to WT and 650 phosphopeptides were significantly downregulated in *L. bicolor* inoculated transgenic line compared to WT (**Figure 4.2A**). These phosphopeptides correspond to 603 and 646 differentially abundant phosphoproteins, respectively. To explore the amino acid pattern in the regions of phosphorylation sites, motif analysis of phosphorylated peptides was performed on the amino acid sequences before and after phosphorylation sites. Motif sequence analysis in upregulated and downregulated phosphopeptides identified proline residue to be frequently occurring in the +1 position and arginine residue in the -3 position of phosphorylated residues (**Figure 4.2B**). This result suggests that these residues in the particular positions may be important for PtLecRLK1-induced phosphorylation modification.

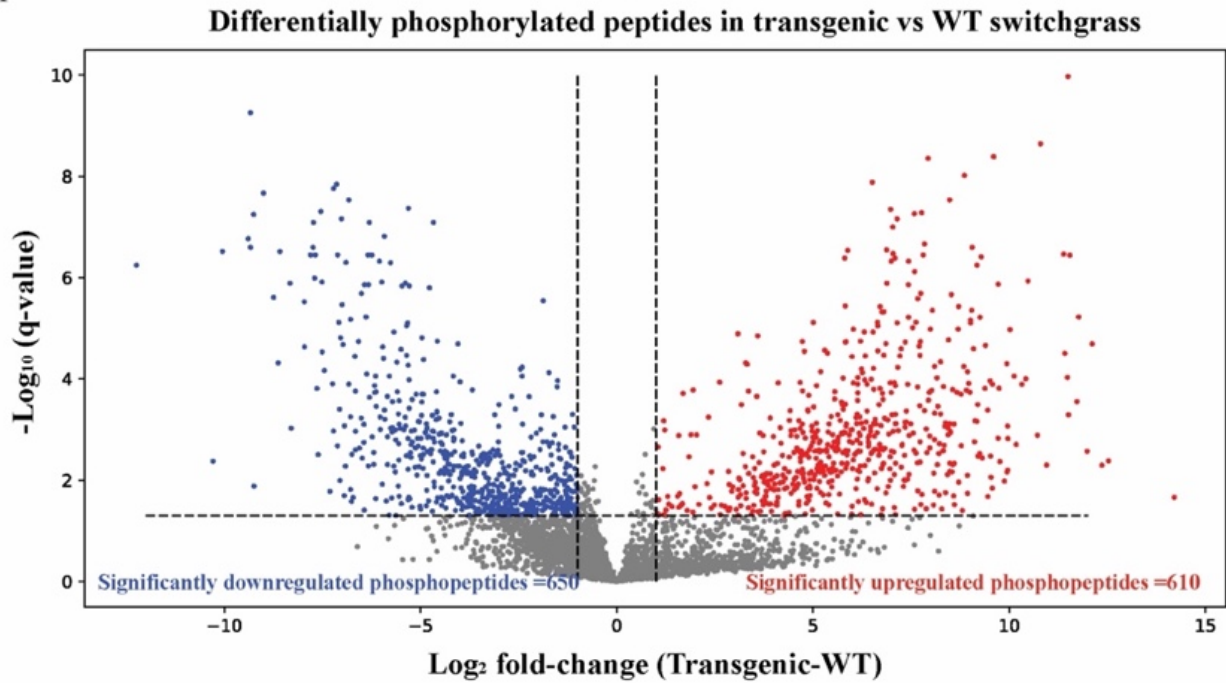
4.3.2 Differentially phosphorylated proteins between fungus inoculated transgenic and WT switchgrass are regulated uniquely at the post-translational level

The interpretation of quantitative phosphoproteomics can be complicated because the differential phosphorylation event could be the result of a change in protein expression and abundance, or it can be due to the change in phosphorylation status. To see if the changes we observed in this study are either due to changes in protein expression or phosphorylation, we compared the differentially regulated proteins identified in proteomics analysis and phosphoproteomic analysis. Our comparison identified only 73 phosphorylated proteins (representing <6%) that are significantly changing in their proteome abundance, suggesting that the majority of these phosphorylated proteins are regulated exclusively at the post-translational level (**Figure 4.2C**).

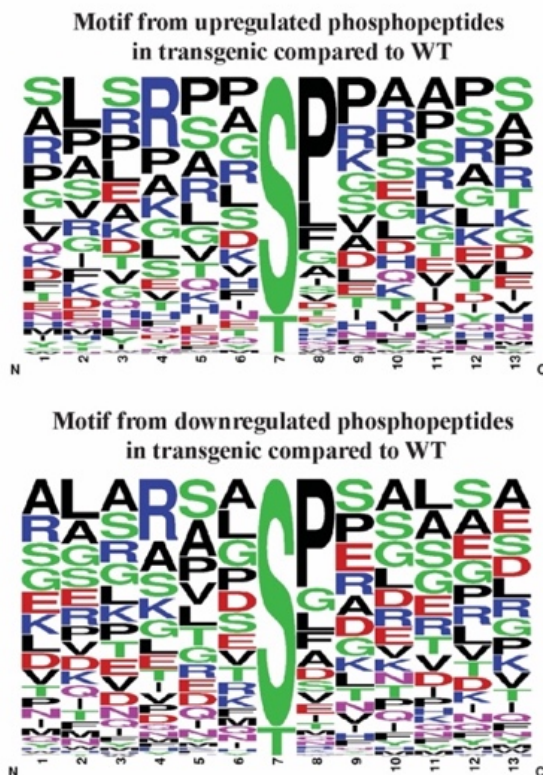
Figure 4.2: Differentially abundant phosphopeptide analysis of transgenic and WT root sample.

(A) Volcano plot showing the differentially abundant phosphopeptide in our comparative phosphoproteomics analysis. Red nodes represent the phosphopeptide that were significantly increased in abundance in transgenic line compared to WT and blue nodes represent the phosphopeptide that were significantly decreased in abundance in transgenic line compared to WT. Significantly changing phosphopeptides are those that passes q-value threshold of 0.05 and $\log_2$ fold change greater than absolute 1. (B) Motif analysis of up/downregulated phosphopeptide in transgenic line compared to WT. Motif analysis was performed using weblogo.berkeley.edu. (C) Comparison of significantly changing phosphopeptide mapped proteins from phosphoproteome analysis to significantly changing proteins from proteome analysis. Red nodes are the proteins significantly upregulated, blue nodes are the proteins significantly downregulated, and yellow are the proteins that are significantly upregulated in one -omics analysis and downregulated in other.

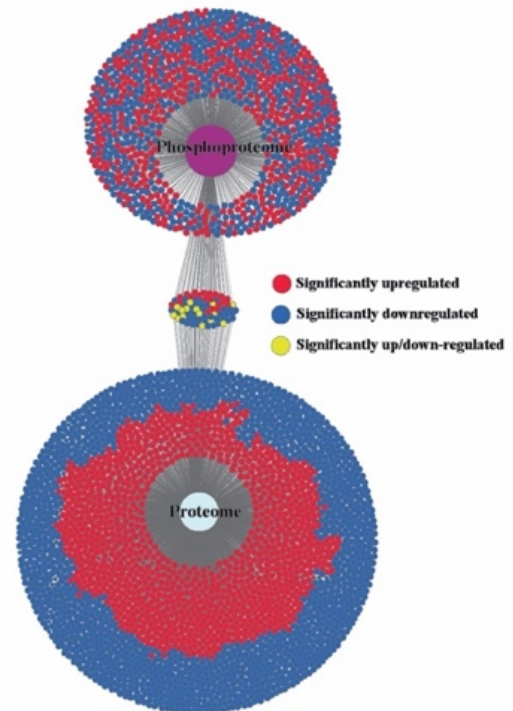
A



B



C



This allows us to dig deeper at the functional level to see the regulation specifically due to phosphorylation.

4.3.3 Functional analysis identified key processes and pathways regulated through phosphorylation by PtLecRLK1 in switchgrass

To understand the biological processes and pathways regulated by the PtLecRLK1 using phosphorylation modification in *L. bicolor* inoculated transgenic plant compared to WT, gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed on differentially abundant phosphoproteins. GO enrichment analysis considered changes in phosphorylation across biological process (BP), molecular function (MF), and cellular component (CC) ontologies. Enriched BP ontology terms spanned a range of different processes, including detection of chemical stimulus, organelle localization, response to carbohydrates, proton transport, and cellular response to nitrogen compounds. Meanwhile, the significantly downregulated phosphoproteins were enriched in biological processes such as organelle assembly, the response of abscisic acid, and negative regulation of gene expression. For molecular function, the significantly upregulated phosphoproteins were mainly enriched in molecular functions such as kinase activity, signaling receptor activity, and GTPase regulator activity. Meanwhile, the significantly downregulated phosphoproteins were enriched in water transmembrane transporter activity. As to the subcellular localization, nuclear speck was enriched in significantly upregulated phosphoproteins and ribonucleoprotein granule was enriched in significantly downregulated phosphoproteins.

KEGG analysis showed that significantly upregulated phosphoproteins in *L. bicolor* inoculated transgenic plant compared to WT were enriched in biological pathways linked to plant-pathogen signaling, spliceosome, and carbon fixation (**Figure 4.3**). Meanwhile, downregulated phosphorylated protein isoforms in *L. bicolor* inoculated transgenic plant were enriched in the SNARE interaction in vesicular transport which is involved in fungal exocytosis (**Figure 4.3**). Pathways like endocytosis, MAPK signaling, phosphatidylinositol signaling, mRNA surveillance pathway, and glycerolipid metabolism were enriched in both significantly up- and downregulated phosphoproteins. Some enriched KEGG pathways such as plant-pathogen signaling, MAPK

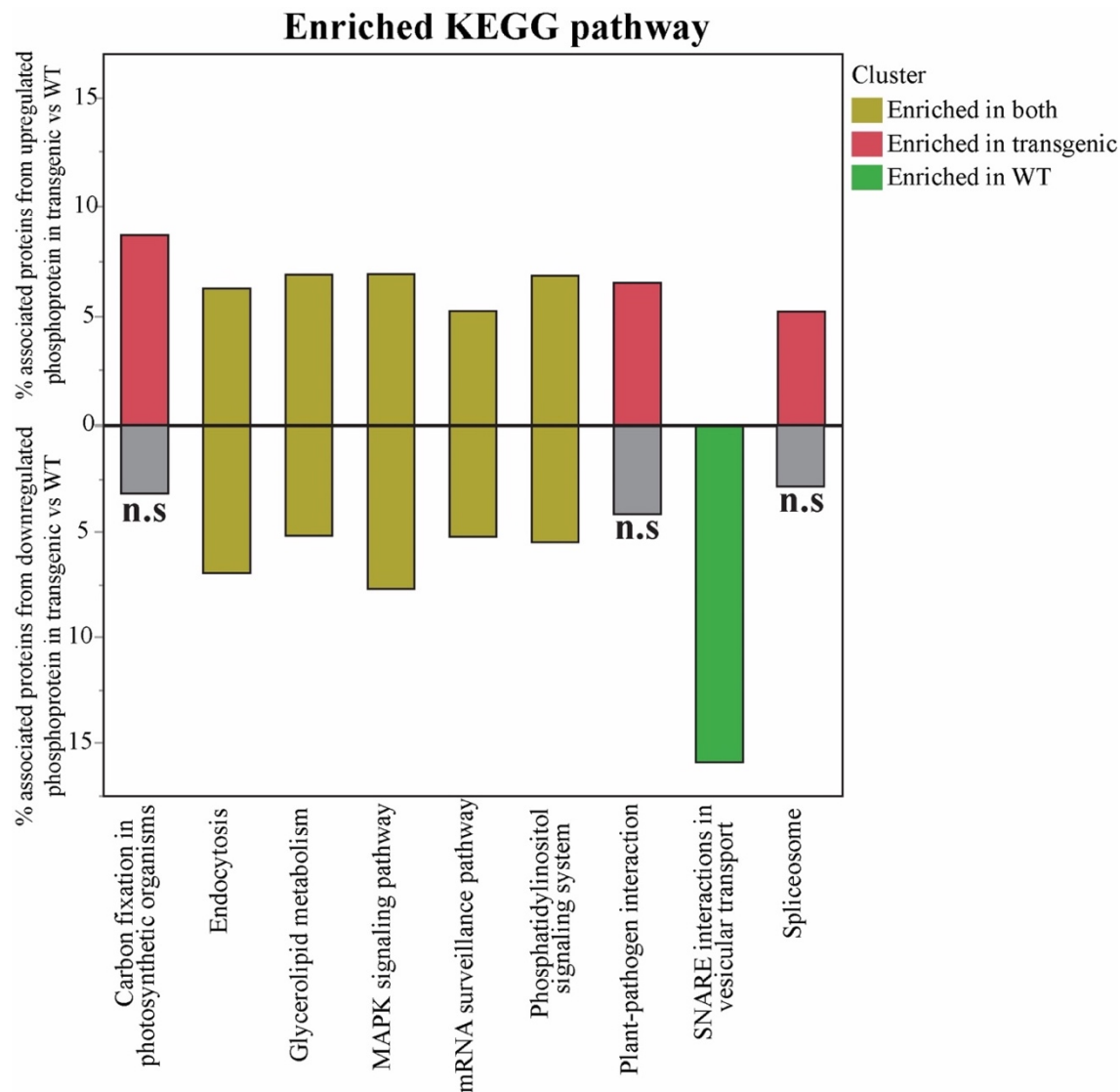


Figure 4.3: Pathway analysis of differentially abundant phosphoprotein isoforms.

Barplot showing the enriched KEGG pathway of differentially abundant phosphopeptide mapped phosphoproteins. Arabidopsis orthologs of switchgrass phosphoproteins were identified and analyzed for enriched KEGG pathway using clueGO software. Red bars are the pathway uniquely enriched among the highly abundant phosphoproteins in transgenic line compared to WT. Green bar is the pathway uniquely enriched among the highly reduced phosphoproteins in transgenic line compared to WT. Yellow bars are the pathway enriched in both conditions. Grey bars are identified but are not significantly enriched.

signaling, and phosphatidylinositol signaling are known to mediate the plant defense response (Hamel et al., 2012, Meng and Zhang, 2013, Lee et al., 2010). The significant enrichment of these pathways suggests that the regulation of plant defense through phosphorylation signaling is important during the plant-fungal interaction. Similarly, pathway like carbon fixation which was found to be enriched in upregulated phosphoproteins could be the plant response as a result of the fungal colonization because plants are known to provide the carbon-derived resources to fungus during the symbiotic interaction (Long, 1989, Zhang et al., 2009, Bais et al., 2006). Therefore, it is plausible that the KEGG enrichment analysis shows diverse biological pathways that are not only important for the establishment of fungal colonization but also for the maintenance of this relationship.

4.3.4 *PtLecRLK1* regulates the phosphorylation status of defense-related transcription factors and kinases

Previous studies on transgenic plants expressing *PtLecRLK1* showed changes in the expression of genes and proteins related to plant defense (Qiao et al., 2021, Labbe et al., 2019). Similarly, the KEGG enrichment analysis identified several enriched pathways such as plant-pathogen signaling, MAPK signaling, and phosphatidylinositol signaling that regulate plant defense response (Hamel et al., 2012, Meng and Zhang, 2013, Lee et al., 2010). This evidence suggests that *PtLecRLK1* regulates the plant defense response. To understand how *PtLecRLK1* regulates the plant defense response, the phosphorylation status of the Plant WRKY DNA-binding transcription factors, a key modulator of plant defense responses, were analyzed. Our analysis identified 4 WRKY proteins (Pavir.1KG087300.1.p, p-site:S350; Pavir.1NG078900.1.p, p-site:S352; Pavir.4KG339400.1.p, p-site:S319; Pavir.6NG193100.1.p, p-site:S49) with significantly lower phosphorylation modification in transgenic plants compared to WT. In contrast, we identified 1 WRKY protein (Pavir.9KG443065.1.p, p-site:T156) with significantly higher phosphorylation modification in transgenic plants compared to WT. Besides this, phosphoproteome measurement showed a significant change in the phosphorylation status of transcriptional regulators called valine-glutamine (VQ) proteins. VQ often interacts with WRKY transcriptional factors (Leon et al., 2020). It can act as a gene activator or repressor and has the ability to modulate gene expression, which depends on phosphorylation status (Leon et al., 2020). Our analysis identified two isoforms

of VQ-21 with four modified sites. All VQ-21 isoforms identified had a lower abundance of phosphorylation in the transgenic line compared to WT. In the similar context, our analysis identified two modification sites in VQ4 both with increased phosphorylation in the transgenic line compared to WT.

In addition to transcriptional factors, protein kinases also play numerous essential roles in the plant including defense response. Among the 1,249 differentially abundant phosphoproteins, 150 of them were annotated as kinases consisting of 22 receptor-like kinases, 7 calcium-dependent protein kinases (CDPKs), and 13 MAP kinases. Notably, a homolog of recently identified and characterized wheat receptor-like cytoplasmic kinase responsible for the susceptibility gene in response to rust fungi effector protein were also identified with increased phosphorylation (Wang et al., 2022). In wheat, this kinase targets the CBF1 proteins that control the downstream expression of susceptibility or immune-related genes depending on phosphorylation status (Wang et al., 2022). Interestingly, a homolog of CBF1 was identified with higher phosphorylation modification in the transgenic plant compared to the WT plant.

4.3.5 PtLecRLK1 regulates nutritional uptake via phosphorylation of transporter proteins

In addition to aiding in uncovering the elaborative signaling mechanisms, a comprehensive phosphoproteomics study can also provide insight into cellular changes as a response to these elaborative signaling mechanisms. Ectomycorrhizal fungi like *L. bicolor* are known to play an integral role in ammonium acquisition and metabolism in trees. To understand how the PtLecRLK1 facilitates the plant ability to regulate the nutritional exchange, we profiled the phosphorylation status of transporter proteins. Among the protein with significantly changing phosphorylation status, 30 were transporter proteins. Out of these, 4 were ammonium transporter (AMT) and all 4 isoforms have significantly decreased phosphorylation abundance (**Figure 4.4**). The AMT can exist in active or inactive states, controlled by the phosphorylation of a conserved threonine residue in the C-terminus (Lanquar and Frommer, 2010). The phosphorylation of conserved threonine levels negatively correlates with root ammonium uptake (Lanquar and Frommer, 2010). The decreased phosphorylated protein abundance of all AMT suggests that transgenic plant with *L. bicolor* inoculation is increasing the uptake of ammonium.

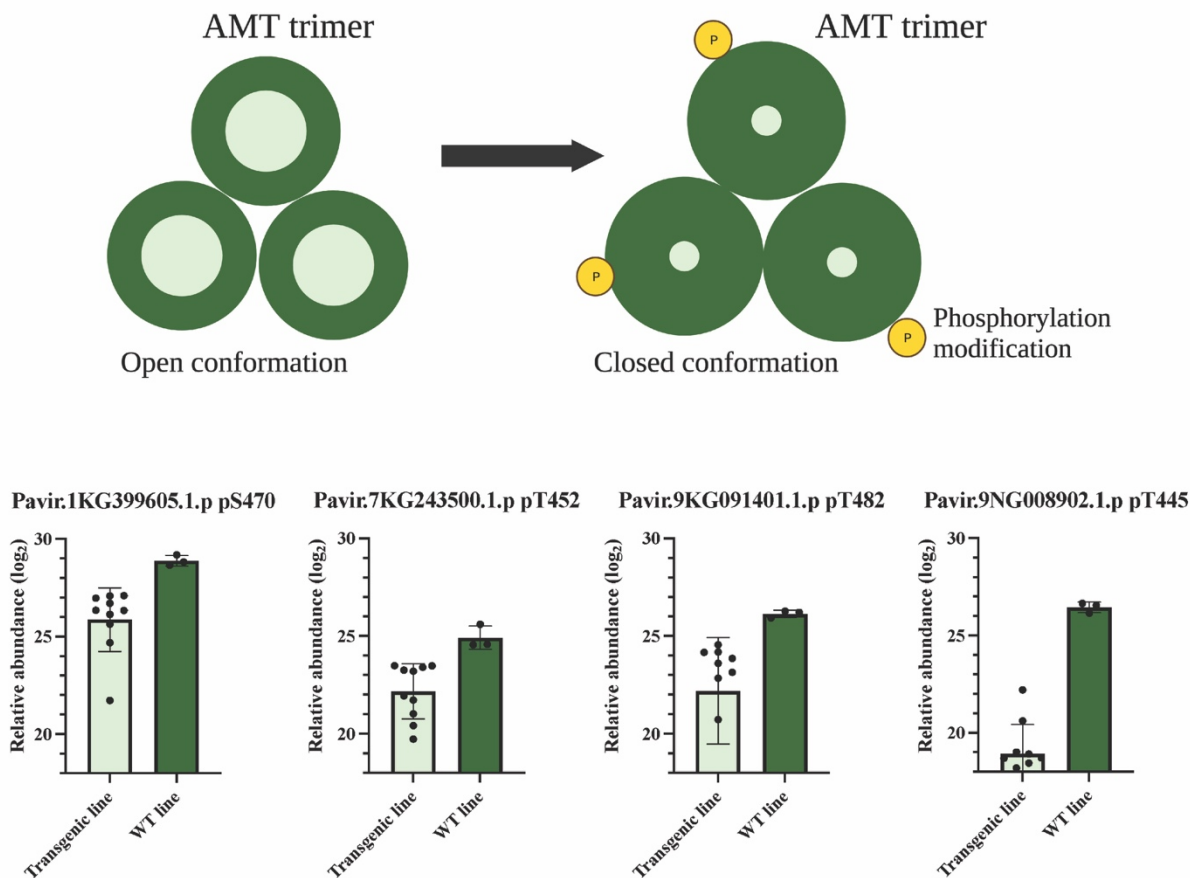


Figure 4.4: Adjustment of ammonium transporter via phosphorylation and phosphorylation profile of ammonium transporter.

Simplified model showing the regulation of ammonia uptake through ammonium transporter via phosphorylation modification. The phosphorylation of conserved threonine level negatively correlates with root ammonium uptake. Barplot below the model figure shows the phosphorylation profile of ammonium transporter in *L. bicolor* inoculated Transgenic and WT plant.

4.3.6 Network analysis and molecular assay identified and validated cGMP kinase as a direct interacting partner of PtLecRLK1

Because phosphorylation can regulate physical or functional interactions of proteins, we next wanted to determine the extent to which the differential abundance phosphoproteins interact with each other. To that end, network analysis using STRING in Cytoscape was performed on differentially phosphorylated protein isoforms (von Mering et al., 2003, Shannon et al., 2003, Szklarczyk et al., 2021). The protein-protein interaction information is derived from experimental, co-expression, co-occurrence, fusion, homology, knowledge, neighborhood, and text mining evidence. The network analysis of differentially abundant phosphoproteins showed extensive interconnectivity. Interestingly, the majority of the differential abundance phosphoproteins with interaction confidence of >0.70 were connected to the node of protein cGMP-dependent protein kinase / PKG II (**Figure 4.5**). To see if the cGMP/PKG II can putatively interact with LecRLK in other organisms, we analyzed the protein-protein interaction in the Arabidopsis homolog of PtLecRLK1. The result show that Arabidopsis homolog of PtLecRLK1 can interact with 9 proteins including cGMP-dependent protein kinase / PKG II. To further explore cGMP/PKG II as a potential interacting partner of PtLecRLK1 that regulate the downstream signaling, we propagated the top 10th percentile nodes from the original network, and proteins from the enriched KEGG pathway were qualitatively searched (**Figure 4.5**). Our result shows that cGMP/PKGII can interact with proteins from enriched pathways like endocytosis, MAPK signaling, phosphatidylinositol signaling, and glycerolipid metabolism, suggesting that PtLecRLK1 might control these pathways with the help of cGMP/PKGII (**Figure 4.5**).

To verify if cGMP-dependent protein kinase / PKG II is a direct interacting partner of PtLecRLK1, bimolecular fluorescence complementation (BiFC) assay was utilized. In the BiFC assay, coding sequence (CDS) of PtLecRLK1 and cGMP-dependent protein kinase / PKG II were cloned into C-terminal fragment of cyan fluorescent protein (CFP^c; pUC119-CD3-1068) and N-terminal fragment of Venus (VENUSⁿ; pUC119-CD3-1076) respectively. These paired plasmids were co-transfected in poplar protoplasts. The interaction of CFP^c -PtLecRLK1 and VENUSⁿ -cGMP-dependent protein kinase / PKG II is indicated by fluorescence due to the formation of an intact fluorescent protein. BiFC assay showed that PtLecRLK1 and cGMP-PK interact with each other

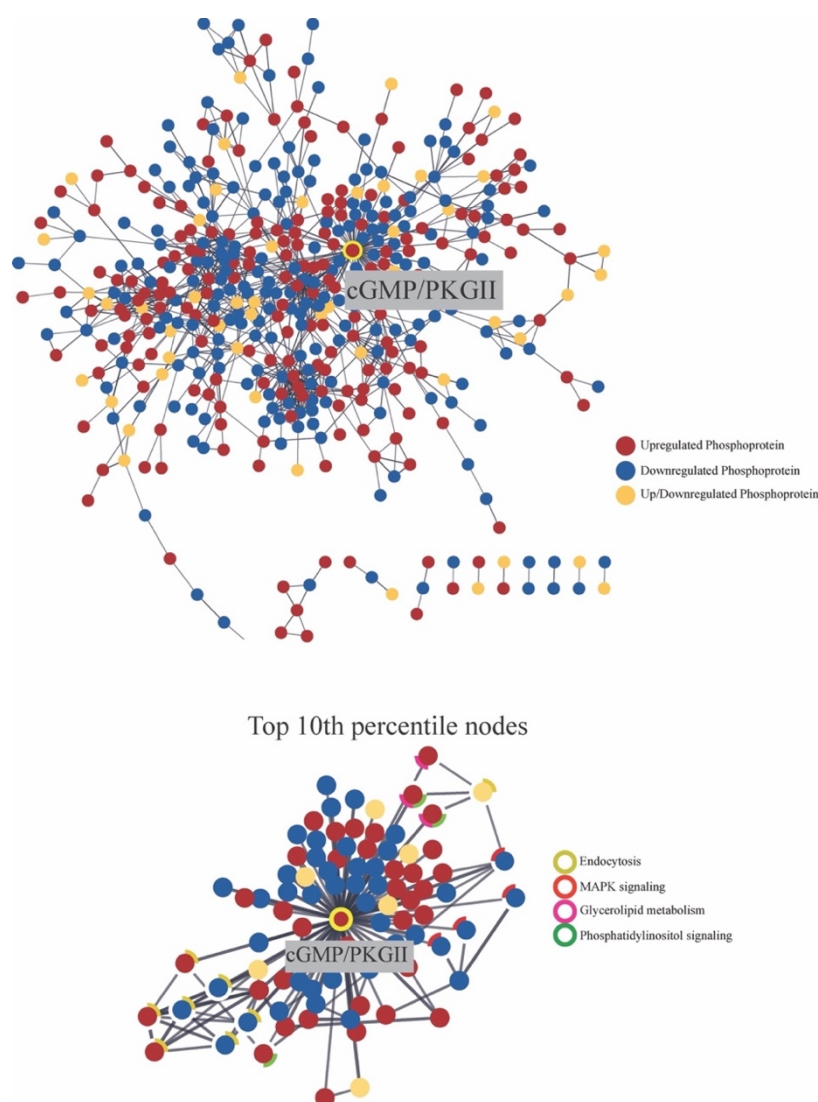


Figure 4.5: Network analysis of differentially abundant phosphoprotein isoforms.

Protein network shows the physical or functional interactions between individual proteins at the interaction confidence score of >0.70 . Red nodes are the significantly upregulated phosphoproteins, blue nodes are the significantly downregulated phosphoproteins and orange nodes are the significantly changing phosphoproteins that contain multiple phosphosite that are significantly up/downregulated. Singletons were filtered out from the network. Network analysis shows the cGMP/PKGII as one of the network hubs with high degree of connectivity with other proteins. The top 10th percentile nodes were propagated from the original network and proteins from the KEGG pathway enriched are color coded.

as they produce green fluorescence similar to the GFP positive control (**Figure 4.6**). The interaction most likely occurs near membrane as the green fluorescence is localized in this region (**Figure 4.6**). Collectively, the finding from network analysis and BiFC assay demonstrate that PtLecRLK1 interact and phosphorylate with cGMP-dependent protein kinase / PKG II. This cellular kinase is a major regulator protein that transduce signals from PtLecRLK1 downstream to the various downstream pathways such as endocytosis, MAPK signaling and phosphatidylinositol signaling.

4.3.7 Cross-species comparison of phosphorylated proteins showed similarity in the signaling mechanism

To explore if the signaling mechanism induced by PtLecRLK1 is conserved across transgenic plants expressing *PtLecRLK1*, transgenic poplar hybrids expressing the *PtLecRLK1* were generated. The transgenic poplar showed phenotypic characteristics seen with ectomycorrhiza (**Figure 4.7A**). Phosphoproteomics analysis was performed in transgenic (host) and wild-type (non-host) poplar and compared with the results of switchgrass. All identified phosphoproteins from poplar samples were sorted based on the fold-change between transgenic poplar treated with *L. bicolor* and WT treated with *L. bicolor*. Switchgrass orthologs of poplar were identified with sequence similarity of >60% using cdhit 2D. A qualitative comparison of poplar and differentially regulating switchgrass phosphoproteins showed a large overlap, suggesting the similarity in the downstream signaling cascade (**Figure 4.7B**). cGMP-dependent protein kinase / PKG II, which was identified as a direct interacting partner of PtLecRLK1 and a key regulator of downstream signaling in network analysis of switchgrass, was also exclusively phosphorylated in transgenic poplar treated with *L. bicolor* (**Figure 4.7B**).

4.4 Discussions

Fungi are crucial decomposers and recyclers of organic matter and a major component of natural ecosystem and modern agriculture due to their nutritional versatility and interactions with plants (Zeilinger et al., 2016). The interactions between plants and fungi are complex and the outcomes can range from symbiotic to pathogenic. Although often mentioned as discrete categories, the type

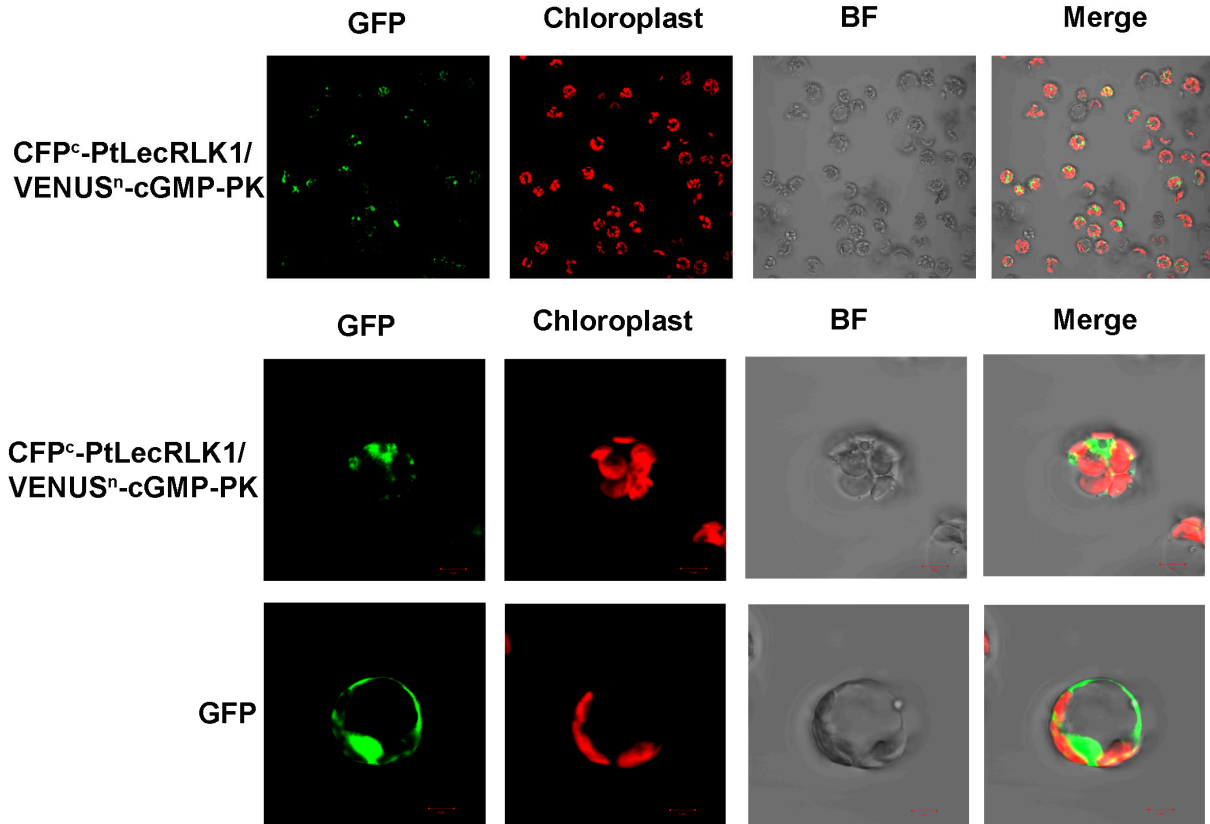
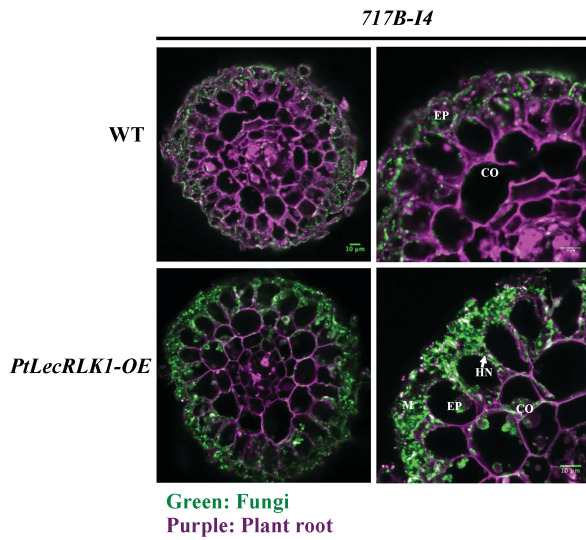


Figure 4.6: Bimolecular fluorescence complementation (BiFC) assay validates the interacting partner of PtLecRLK1.

BiFC assay done in poplar protoplasts showed PtLecRLK1 and cGMP-PK interacts with each other as they produce green fluorescence. The interaction most likely occur near membrane (in cytoplasm) as the green fluorescence is localized in this region.

A



B

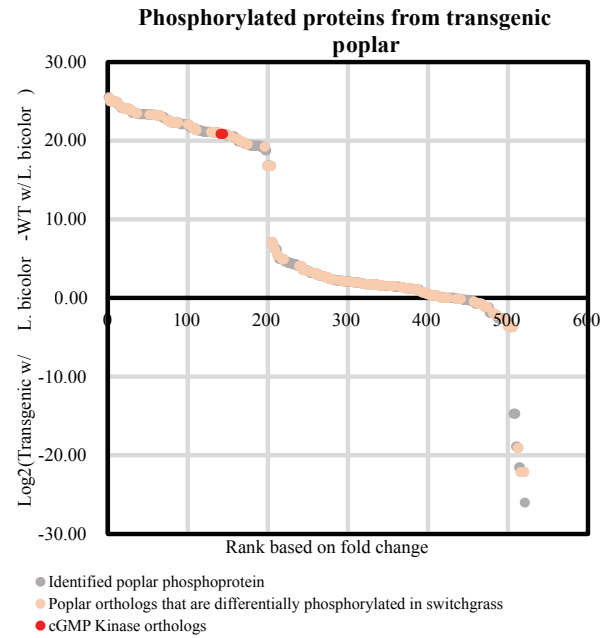


Figure 4.7: Phenotypic characterization of *L. bicolor* colonization in roots of wild-type and transgenic poplar and assessment of phosphoproteins from poplar samples.

(A) Dual fluorescence-stained transverse root section of WT and transgenic poplar. Green color represents fungi cell; Purple color represents root cell. M: Mantle; CO: Cortex; EP: epidermis and HP: Hartig net. **(B)** Assessment of phosphoproteins identified in poplar samples and their comparison with differentially regulated phosphoproteins from switchgrass samples.

of interactions is not discrete but rather a continuum ranging from parasitism to mutualism (Mandyam and Jumpponen, 2014). Moreover, the continuum can change during the lifespan of the interacting organism due to developmental changes and changes in environment. Nonetheless, mutualistic plant-fungal interactions can provide stability to both interacting partners. Beneficial fungus aid plants by promoting plant growth and development, improving plant foraging, and aiding in the nutrients and water acquisition and stress tolerance (Mendes et al., 2011, Berendsen et al., 2012, Lu et al., 2018, Cook et al., 1995). In return, plants provide carbohydrates derived from photosynthesis to the fungus (Buscot et al., 2000). A comprehensive understanding of the molecular mechanism underlying the selection and maintenance of plant-fungal mutualistic interaction can enable the engineering of the mutualistic interaction that can enhance plant productivity and fitness.

A specific bioenergy plant species (*Populus trichocarpa*) forms a beneficial interaction with the fungus *Laccaria bicolor* mediated by a specific lectin receptor-like kinase (PtLecRLK1) (Labbe et al., 2019). Engineering the expression of *PtLecRLK1* in a variety of non-host plant species results in *L. bicolor* root colonization (Qiao et al., 2021, Labbe et al., 2019). However, molecular signaling mediated by PtLecRLK1 during the establishment of beneficial microbes remains unclear. To uncover signaling cascades regulated by *PtLecRLK1* during beneficial plant-fungal interaction, a detailed phosphoproteomics analysis was performed in this study.

Our comprehensive, quantitative phosphoproteome analysis revealed that *L. bicolor* inoculated WT and the transgenic plants are characterized by distinct phosphorylation, which broadly represented the discrete downstream signaling events of PtLecRLK1. Most of the proteins mapped by differentially abundant phosphopeptides were uniquely regulated at the post-translational level, which emphasizes the importance of phosphorylation modification in responding to environmental changes. Differential analysis shows that PtLecRLK1 regulated the phosphorylation status of kinases such as Mitogen-activated protein kinases, calcium-dependent protein kinases, and Phosphatidylinositol 3- and 4-kinases. MAPKs are evolutionarily conserved in plants and fungi and act as a key signal transduction component that uses reversible phosphorylation of proteins to channel information (Hamel et al., 2012). MAPKs upon activation can phosphorylate downstream targets to modulate their biochemical properties, which leads to distinct output responses (Hamel

et al., 2012, Meng and Zhang, 2013). During plant-fungal interactions, plant MAPKs form a signaling cascade, transducing defense signals downstream of multiple plant receptor kinases and regulating plant immunity (Hamel et al., 2012, Meng and Zhang, 2013). Similarly, Phosphatidylinositol 3-kinase is involved in endocytosis, reactive oxygen species (ROS) production, and transcriptional activity (Lee et al., 2010). Phosphatidylinositol 3-kinase regulated phosphatidylinositol signaling pathways have also been shown to involve in signal transduction for the regulation of gene expression, hormone function, cell differentiation, transduction of intracellular signaling molecules, control of cell responses to environmental factors, regulation of ion-channel gating, energy metabolism and rearrangement of the cytoskeleton (Lin et al., 2004, Meijer and Munnik, 2003). The changes in the phosphorylation status of these cellular kinases might be responsible for the fine-tuning in pathways such as plant-microbe interactions, spliceosome, carbon fixation, and SNARE-based vesicular transport observed.

WRKY transcription factors (TFs) are the global modulator of plant defense responses that regulate the transcription of downstream target genes upon being challenged by pathogens (Chakraborty et al., 2019). They interact both physically and functionally to plant defense responses (Chen et al., 2010). Many plant defense genes, including pathogenesis-related (*PR*) genes and the regulatory *NPR1* genes, contain W box sequences in their promoters that are recognized by WRKY proteins (Yu et al., 2001). In plant cells, WRKY can act as a transcriptional activator or repressor. Previous studies have shown that the phosphorylation status of WRKY can protect them from proteasomal degradation or alter the cellular localization (Chakraborty et al., 2019). The decrease in phosphorylation status observed in our study suggest either they are being degraded or have altered cellular localization, which could explain the changes in plant defense pathways observed in previous multi-omics study. Similarly, valine-glutamine (VQ) proteins often interact with WRKY transcriptional factors (Leon et al., 2020) and can act as a gene activator or repressor. Its ability to modulate gene expression depends on phosphorylation status (Leon et al., 2020). The previous study has shown that after infection, MPK4 is activated and phosphorylates VQ21 (Jing and Lin, 2015). This triggered the release of VQ21 along with its interacting partner WRKY (Jing and Lin, 2015). The phosphorylated VQ-WRKY binds to the promoter of defense-responsive genes and activates their expression (Jing and Lin, 2015). The lower abundance of

phosphorylation in VQ-21 protein observed in the transgenic line compared to WT suggests the reduction in expression of defense-responsive genes mediated by the VQ-WRKY complex.

Kinases play numerous essential roles in the plant including defense signaling. The substantial changes in the phosphorylation status of numerous kinases observed in the phosphoproteome study indicate that PtLecRLK1 suppressed defense response is complex and may be controlled via multiple signaling pathways. Some kinases are also known to elevate fungal susceptibility (Wang et al., 2022). Notably, a recently identified and characterized wheat receptor-like cytoplasmic kinase (TaPsIPK1) is responsible for susceptibility against rust fungi. The effector protein from fungus can elevate the phosphorylation of TaPsIPK1, which increases its nucleus localization (Wang et al., 2022). In the nucleus, TaPsIPK1 targets the CBF1 proteins that control the downstream expression of susceptibility or immune-related genes depending on phosphorylation status (Wang et al., 2022). The increased phosphorylation of TaPsIPK1 and CBF1 homolog in the transgenic plant compared to the WT plant suggest that PtLecRLK1 not only regulates the defense-related gene but also elevates the susceptibility-related genes.

Network analysis was able to highlight the complexity of organism-wide cell signaling mediated by PtLecRLK1. Through this, cGMP-PK was identified as a key network hub. cGMP-PK is known to play a key role in perceiving cellular cGMP in diverse physiological processes. In rice, cGMP/PK possesses both protein kinase and phosphatase activities, and cGMP stimulates its kinase activity but inhibits its phosphatase activity. cGMP-PK has been also known to target the phosphorylation status of various transcription factor that regulates phytohormones. This protein was exclusively phosphorylated in transgenic poplar treated with *L. bicolor* and was validated as a direct interacting partner of PtLecRLK1. Moreover, our result shows that cGMP-PK can interact with proteins related to endocytosis, MAPK signaling, phosphatidylinositol signaling, glycerolipid metabolism, and fungal susceptibility regulating kinases. This suggests that PtLecRLK1 might be modulating its function through the interacting partner cGMP-PK which could be conserved across the various species.

Fungi such as *L. bicolor* can secret various molecular signals in the form of volatile organic compounds, hormones and hormone mimics, carbohydrate-based signals, and protein-based

signals. These fungal signals could be potentially perceived by the ligand-binding extracellular domain of *PtLecRLK1*, which then activates the kinase domain (**Figure 4.8**). The activated kinase domain can relay the signal downstream through phosphorylation of cellular kinase cGMP-dependent protein kinase/PKG II (**Figure 4.8**). cGMP/PKG II acts as a regulatory hub that controls the further signaling cascade through other kinases such as Mitogen-activated protein kinases, calcium-dependent protein kinase, Phosphatidylinositol 3- and 4-kinase and cytoplasmic receptor-like kinases (**Figure 4.8**). Through these, *PtLecRLK1* can activate downstream proteins and transcription factors that can aid in nutrient uptake, reduce defense, and increase susceptibility to fungal colonization. All these changes will eventually lead to the alteration of genes necessary for facilitating the symbiotic interaction between *Populus* and *L. bicolor*.

In conclusion, our study utilizes a phosphopeptide enrichment method with high-resolution mass spectrometry to characterize the *PtLecRLK1* regulated signaling pathways in non-host plants i.e., switchgrass and poplar. In the signaling pathway analysis, cGMP/PKG II was identified as a major regulator of downstream signaling complex and a direct interacting partner of *PtLecRLK1*. The interaction between *PtLecRLK1* and cGMP/PKG II was further validated using the BiFC assay. Taken together, our study defined the *PtLecRLK1*-induced signaling cascade that is responsible for *L. bicolor* root colonization and, more broadly, advanced our fundamental understanding of the molecular mechanism underlying the selection and maintenance of plant-fungal mutualistic interaction.

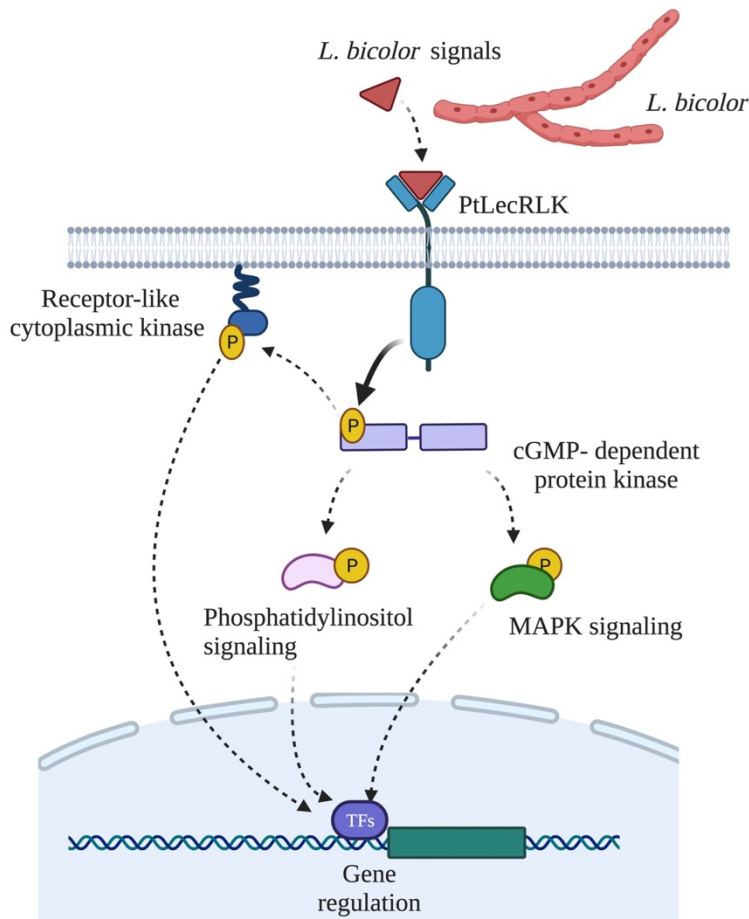


Figure 4.8: Proposed signaling cascade of PtLecRLK1.

L. bicolor produced molecular signals are potentially perceived by the ligand-binding extracellular domain, which then activates the kinase domain. The activated kinase domain can relay the signal downstream through phosphorylation of cellular kinase cGMP-dependent protein kinase/PKG II. cGMP/PKG II act as a regulatory hub that controls the further signaling cascade through MAPK and Phosphatidylinositol signaling. cGMP/PKG II can also activate the other cytoplasmic kinase that are known to be involve in activating fungal susceptibility. All these changes will eventually lead to the alteration of genes necessary for facilitating the symbiotic interaction between *Populus* and *L. bicolor*.

CHAPTER 5: UNDERSTANDING THE PLANT-RESPONSIVE TRANSCRIPTION FACTOR PIPR REGULATED PATHWAYS IN THE POPULUS ENDOPHYTE *PSEUDOMONAS* GM79

This is the collaborative work with researchers from the University of Washington i.e., Dr. Yasu Oda, Dr. Amy Schaefer, and Dr. Carrie Harwood. These collaborators conceptualized this study and carried out microbial growth and sampling for the proteomics measurement.

Previously, poplar derived metabolite that act as an effector to LuxR transcriptional regulator protein in poplar endophyte *Pseudomonas* sp. GM79 was identified. However, the proteins and biological pathways regulated by this metabolite in endophytic microbe hasn't been well characterized. Him K Shrestha is leading the proteomics project to understanding of how poplar-associated bacteria can sense and respond to poplar-derived signaling compound. Him K Shrestha's contributions in this project includes: 1) Proteomics sample processing, 2) Mass spectrometry measurement and data analysis, 3) Manuscript draft writing and figure generation.

5.1 Introduction to orphan LuxR (PipR) and its role in helping microbes perceive plant-derived metabolite

Like plant's response to the plant-microbe interaction (discussed in chapter 4), microbe can also sense and respond uniquely to signals involve in plant-microbe interaction. Quorum sensing is a bacterial process of using small signaling molecules known as autoinducers to regulate cell-to-cell communication and thus regulate their cell density (Nasser and Reverchon, 2007). Many plant-associated bacteria produce quorum-sensing (QS) signaling molecules called *N*-acyl homoserine lactones (AHL) (Subramoni et al., 2011, Schuster et al., 2013). LuxI-family proteins are the main enzymes responsible for AHL synthesis, and the LuxR family of proteins binds to the AHL signaling molecules, which induce conformational change (Subramoni et al., 2011, Nasser and Reverchon, 2007). The change in conformation regulates the interaction of LuxR proteins with a specific site of DNA and modulates the transcription of targeted genes. LuxR proteins are ~ 250 aa long and have two functional domains i.e., AHL-binding N-terminal domain and transcription regulating C-terminal domain (Nasser and Reverchon, 2007).

Several studies have shown that plant-associated proteobacteria possess LuxR-like proteins without the LuxI partner for AHL synthase (Schaefer et al., 2013, Schaefer et al., 2016, Patankar and Gonzalez, 2009, Subramoni and Venturi, 2009). These LuxR proteins are called LuxR orphans or LuxR solos. These orphan LuxR proteins are present on bacteria that contain canonical LuxI-LuxR family pairs as well as on the bacteria that don't have canonical LuxI-LuxR pairs. These LuxR orphan proteins do not react to AHLs but instead respond to plant tissues specific molecules. In the poplar plant endophyte *Pseudomonas* sp. GM79, the orphan LuxR-type transcription factors are referred to as PipR transcriptional regulators (Schaefer et al., 2016, Luo et al., 2021). *PipR* is upstream of *pipA*, a proline iminopeptidase (Pip) gene, and activates *pipA* expression along with other genes in response to plant-derived metabolites (Zhang et al., 2007, Gonzalez and Venturi, 2013, Chatnaparat et al., 2012, Schaefer et al., 2016). Similarly, *pipR* is adjacent to an alanine aminopeptidase (named *aapA*) and operon encoding ATP-binding cassette (ABC)-type transporter (*aapBCDEF*), which is essential for the PipR response to plant metabolites and activation of *pipA* expression (Schaefer et al., 2016).

Recently a plant-derived metabolite has been identified as a potent PipR effector for *Pseudomonas* sp. GM79 (Coutinho et al., 2018). This metabolite is an ethanolamine-derived molecule called N-(2-hydroxyethyl)-2-(2-hydroxyethylamino) acetamide (termed HEHEAA) (Coutinho et al., 2018) found in poplar leaf macerates. Poplar endophyte *Pseudomonas* sp. GM79 has been shown to respond to HEHEAA at picomolar concentrations and activate the gene expression of *pipA* (Coutinho et al., 2018). Currently, it is believed that HEHEAA from poplar leaf binds to AapF periplasmic substrate binding protein (SBP) that delivers it to ABC-type transporter, which helps to bring them inside the cell. Intracellular HEHEAA then binds to PipR, which activates the expression of *pipA* (Luo et al., 2021). However, the system-wide understanding of the proteins and pathways regulated by the PipR-HEHEAA and its broader role in *Pseudomonas* sp. GM79 plant colonization remain unclear. The objective of this study is to advance the understanding of how plant-associated bacteria can sense and respond to their environment by utilizing mass spectrometry-based proteomics to study the system-wide molecular changes in *Pseudomonas* sp. GM79 1-3 hours post HEHEAA addition.

5.2 Materials and Methods

5.2.1 *Pseudomonas* sp. GM79 growth and bottom-up proteomics samples preparation

Pseudomonas sp. GM79 were incubated at 37C until the growth rates (OD_{600nm}) of ~ 0.1 . One micromolar HEHEAA was added, and cell cultures were sampled at times 1, 2, and 3 hours post-HEHEAA addition. These time points were chosen because the PipR-HEHEAA regulated enzyme PipA activity gradually increased till 3 hours, with the highest enzyme activity at 3 hours. The sample was pelleted and frozen using liquid nitrogen. Frozen cell pellets were suspended in 400 μ L lysis buffer (4% sodium dodecyl sulfate, 10 mM dithiothreitol in 100 mM of ammonium bicarbonate buffer). Resuspended samples were boiled for 10 min, sonicated at 30% amplitude, 10 s pulse with 10 s rest, 2 min total pulse time using a Digital Sonifier 250 (Branson Ultrasonics Corporation) and the boiled again for an additional 10 min. The cell lysate was centrifuged, and the supernatant fractions were extracted and then adjusted to 30 mM iodoacetamide to prevent the reformation of disulfide bonds. Proteins were precipitated using chloroform-methanol-water extraction (Jiang et al., 2004) and dried protein pellets were resuspended in denaturing solution (2% sodium deoxycholate (SDC) in 100 mM ammonium bicarbonate). Protein concentrations were estimated using nanodrop (Thermo Fischer Scientific) and then digested and cleaned up in the same way as described in chapter 3.

5.2.2 LC-MS/MS and data analysis

All samples were analyzed on a Q Exactive Plus mass spectrometer (Thermo Fischer Scientific) coupled with an automated Vanquish UHPLC system (Thermo Scientific). Peptides were separated on a MudPIT system made with triphasic precolumn (RP-SCX-RP; 100 μ m inner diameter and 15 cm total length) coupled to an in-house-pulled nanospray emitter of 75 μ m inner diameter packed with 15 cm of 1.7 μ m of Kinetex C18 resin (Phenomenex). For each sample, a 5 μ g injection of peptides was loaded and analyzed from three salt cuts of ammonium acetate (35, 50, and 500 mM) followed by a 180 min split flow organic gradient, wash, and re-equilibration. All other instrument data acquisition settings were kept the same as described in chapter 3.

Raw mass spectrometry data were searched with Proteome Discoverer 2.4 using the *Pseudomonas* sp. GM79 reference proteome downloaded from NCBI website (on 2nd February 2022) to which common laboratory contaminants were appended. The Proteome Discoverer search was processed by the SEQUEST HT database search algorithm (Eng et al., 1994) and the confidence in peptide-to-spectrum matching (PSM) was evaluated by Percolator (Kall et al., 2007). The following parameters were set up for tryptic peptides searches: peptide and fragment ion mass tolerances were set to ± 10 ppm and ± 0.02 Da, respectively. Carbamidomethylation of cysteine was set as static modification and oxidation of methionine was set as a dynamic modification. Peptide and PSMs were considered identified at $q < 0.01$. Protein level identification and quantification were required to have at least one unique peptide sequence. Protein relative abundance values were calculated by summing together peptide extracted ion chromatograms. Proteins with > 1 unique peptide were exported from Proteome Discoverer. Protein data matrix were normalized and processed similarly as described in chapter 3. ANOVA followed by Tukey HSD binary comparison was performed on the normalized and processed data matrix to get the differentially abundant proteins. Proteins that are significantly changing with $p\text{-value} < 0.05$ and greater than 1 absolute $\log_2$ fold change in at least one binary comparison are considered differentially abundant proteins.

5.3 Results and discussions

5.3.1 Proteomics results support the HEHEAA dependent PipR-signaling model of *Pseudomonas* GM79

Poplar leaf macerates contains an ethanolamine-derived metabolite HEHEAA. HEHEAA can act as an effector of LuxR-type protein, PipR in poplar endophyte *Pseudomonas* sp. GM79 (Coutinho et al., 2018). To have a system-wide understanding of how poplar endophyte *Pseudomonas* sp. GM79 can sense and respond to the poplar-derived metabolites, we performed comparative proteomics on the *Pseudomonas* GM79 sampled at 1, 2, and 3 hours post-HEHEAA addition. At a 1% false discovery rate, more than 2300 proteins were identified across experimental conditions, providing sufficient proteome coverage for comprehensive functional interrogation. To investigate if the proteomics data support the current models of HEHEAA-dependent PipR activation of *pipA*

and *aapA* expression, we qualitatively profiled the proteins encoded by adjacent genes upstream and downstream of the PipR regulator (**Figure 5.1**). Our proteomics result shows the increasing abundance of AapF protein, also known as a periplasmic substrate-binding protein (SBP) (**Figure 5.1**). HEHEAA from poplar leaf is initially recognized and bound by SBP, which helps to import this metabolite inside the cell (Luo et al., 2021). Proteomics result did not detect the PipR protein in any conditions, which is most likely due to it being a transcriptional regulator protein that acts at low abundance with low stability. Nonetheless, the known targets of PipR i.e., PipA and AapA, were both detected (Coutinho et al., 2018). PipA abundance was highest at 2 hrs but slightly decreased at 3 hours, while AapA activity increases gradually till 3hrs post-HEHEAA addition (**Figure 5.1**). Overall, the proteomics results support the current model of HEHEAA-dependent PipR activation of known peptidase and provide additional validation that PipR activity is highest at around 2-3 hours post-HEHEAA addition.

5.3.2 Quantitative analysis reveals the differential regulation of pathways relevant to plant colonization

To identify changes in protein abundance resulting due to the addition of HEHEAA, ANOVA followed by Tukey HSD binary comparisons was performed. Proteins that are significantly changing with $p\text{-value} < 0.05$ and greater than 1 absolute $\log_2$ fold change in at least one binary comparison are considered differentially abundant proteins. In total, 714 proteins were differentially abundant, which overall showed six major trends (**Figures 5.2A and 5.2B**). KEGG enrichment analysis showed that these differentially abundant proteins were enriched in pathways involve in starch and sucrose metabolism, phosphatidylinositol signaling, monobactam biosynthesis, chemotaxis, thyroid hormone synthesis and terpenoid biosynthesis pathway (**Figure 5.2C**). The KEGG enrichment results suggest the potential regulation of these enriched pathways by the HEHEAA-dependent PipR activation and their importance in plant colonization. Qualitative assessment on the KEGG enriched pathways filter out the thyroid hormone synthesis and terpenoid biosynthesis pathways because the *Pseudomonas* GM79 doesn't produce them, and their enrichment were due to the ambiguity of KEGG orthologs. Among the remaining enriched pathways, bacterial chemotaxis pathway (18 proteins), starch and sucrose metabolism (9 proteins) and monobactam biosynthesis pathway (6 proteins) were identified with more than 5 differentially

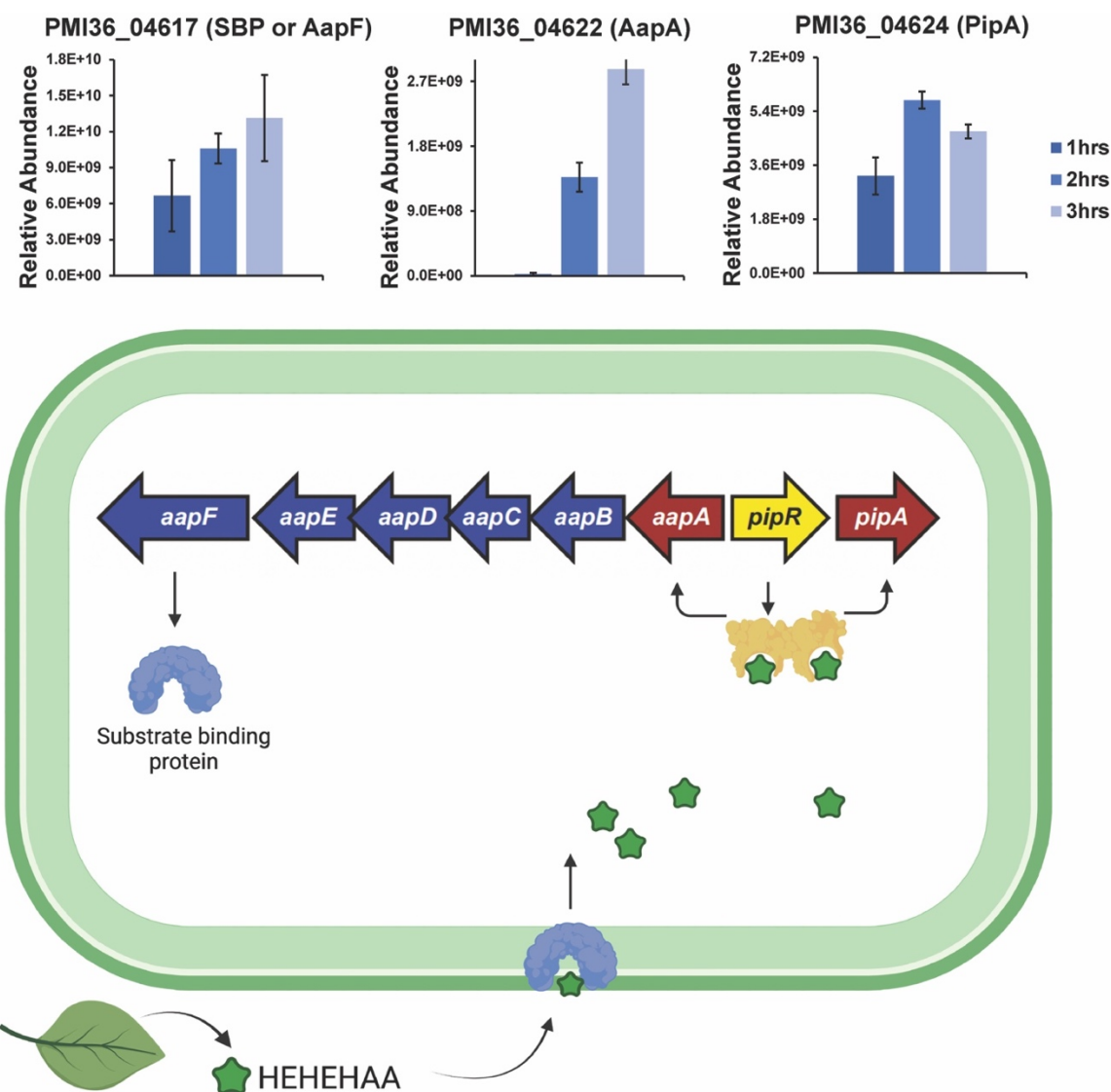


Figure 5.1: Current model of HEHEAA-PipR regulated proteins with the abundance profiles of few proteins identified in proteomics measurement.

The bar chart shows the relative abundance of proteins at 1-3 hours post-HEHEAA treatment. Error bars on the bar chart represent a standard deviation. AapF is a substrate-binding protein (SBP). AapA is an alanyl aminopeptidase and PipA is a prolyl iminopeptidase.

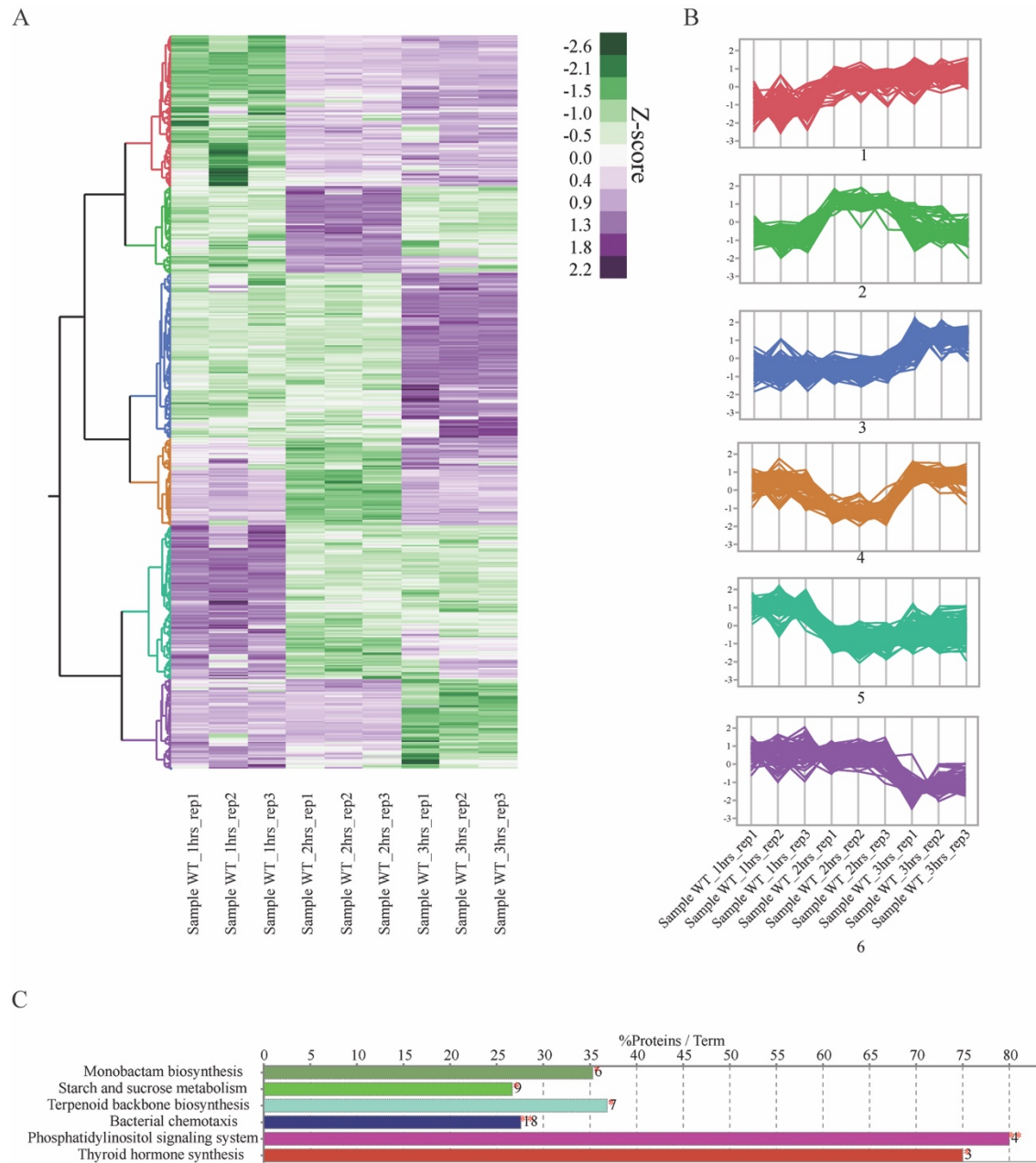


Figure 5.2: Differentially abundant proteins and enriched KEGG pathways.

A) Heatmap showing the differentially abundant proteins. Heatmap showed 6 major clusters B) Six major trends were among the differentially abundant proteins. C) Enriched KEGG pathway across the differentially abundant proteins.

abundant proteins. Chemotaxis and monobactam (antibiotics) biosynthesis pathways were particularly interesting because they have been previously linked with plant colonization by endophytes (Kandel et al., 2017, Martinez-Klimova et al., 2017, Tsai et al., 2020). The monobactam biosynthesis pathway is used to synthesize monobactam, a member of beta-lactam antibiotics, that inhibits peptidoglycan synthesis which is a crucial process for the maintenance of bacterial cell wall integrity resulting in bacterial cell death (Kong et al., 2010). Endophytes often confer resistance to the host plant from various biotic stressors and this HEHEAA-regulated pathway could be one of the phytopathogen-defending traits of endophyte *Pseudomonas* sp. GM79 during plant colonization (Kandel et al., 2017, Rosenblueth and Martínez-Romero, 2006).

5.3.3 HEHEAA influences chemotaxis and motility in *Pseudomonas* GM79

Chemotaxis was one of the enriched pathways identified in differentially abundant proteins. At the protein level, several chemotaxis-related proteins were identified as being upregulated that mapped to clusters 1 and 3 (**Figure 5.2B**). Two methyl-accepting chemotaxis proteins (MCPs) (PMI36_02740 and PMI36_01971) increased in abundance by greater than 6 log₂ fold in 3 hrs compared to 1hrs post-HEHEAA addition (**Figure 5.3**). Methyl-accepting chemotaxis proteins are the dominant member of chemoreceptors in bacteria that sense intracellular or environmental signals and relay them downstream to get the precise response (Salah Ud-Din and Roujeinikova, 2017, Berleman and Bauer, 2005). MCPs contain a ligand-binding domain, transmembrane domain, and a cytoplasmic signaling domain that can interact with various downstream proteins (Wadhams and Armitage, 2004, Bi and Lai, 2015). Some of the interacting downstream proteins include receptor coupling protein CheW, histidine kinase CheA, methyltransferase CheR, and methylesterase CheB (**Figure 5.3**) (Wadhams and Armitage, 2004, Bi and Lai, 2015). The methylation or demethylation of MCPs determines the adaptation of bacterial chemotaxis (Engstrom and Hazelbauer, 1980). Although the methylation or demethylation status of MCPs protein is yet to be measured, the abundance of proteins controlling the methylation/demethylation, i.e., CheR and CheB, have been identified in our proteomics measurement. These proteomics results showed a significant increase of CheR (PMI36_02783) and CheB proteins (PMI36_02793) of more than 4 log₂ fold and 10 log₂ fold respectively in 3 hrs compared to 1 hr post-HEHEAA addition (**Figure 5.3**). CheR catalyzed methylation of helix subdomain of MCPs, while CheB is

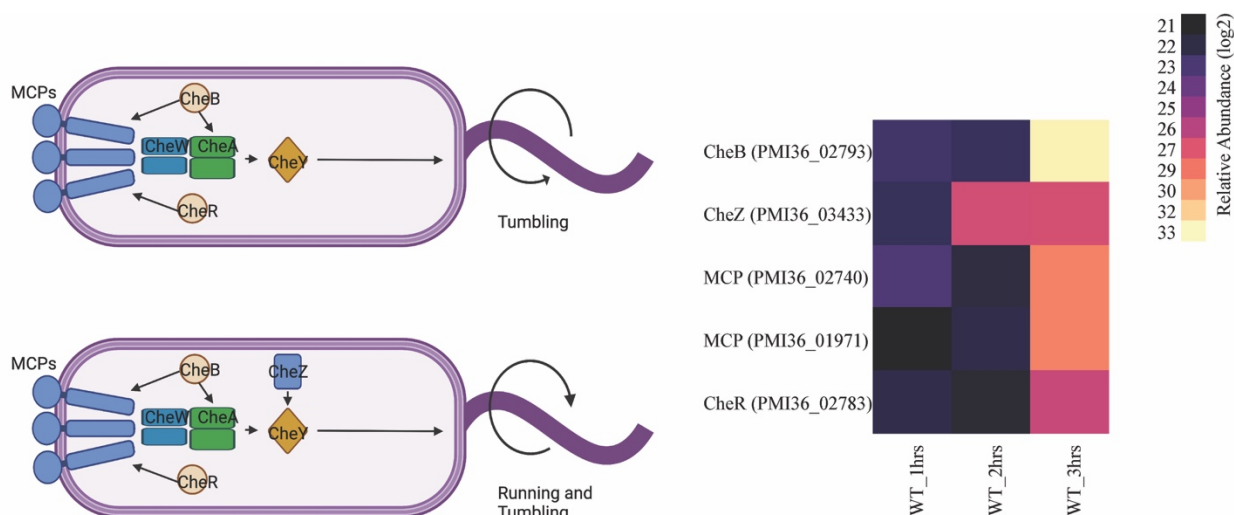


Figure 5.3: Model representing the types of motilities and abundance profile of chemotaxis-related proteins identified.

Bacteria exhibit two types of motions i.e., running and tumbling and tumbling. Running and tumbling are due to the counterclockwise rotation of flagella resulting in directed motility toward the attractant or away from the repellent. Tumbling is due to the clockwise rotation of flagella. Heatmap next to the model shows the abundance profile of significantly changing chemotaxis-related proteins across the three hours.

responsible for demethylation. The methylation and demethylation of MCPs by CheR and CheB coordinates cell to navigate in gradients of signals ranging in diverse concentrations (Endres and Wingreen, 2006, Bi and Lai, 2015). CheY is the response regulator protein of MCP that control the direction of flagella rotation depending on phosphorylation status (Bi and Lai, 2015). CheZ is a phosphatase that removes the phosphate from the CheY (Salah Ud-Din and Roujeinikova, 2017). The phosphorylated CheY favors the tumble motilities whereas dephosphorylated CheY favors the run and tumble type of motilities. Our result shows a significant increase in CheZ (PMI36_03433) by greater than 4 log₂ fold in 3 hrs compared to 1 hr (**Figure 5.3**). Our result also shows that the increase in abundance occurs between time 2 to 1 hrs post-HEHEAA addition, suggesting a rapid response. The increased abundance of CheZ also suggests the running and tumbling motion of *Pseudomonas* GM79 in later time stages. Running and tumbling motion is associated with counterclockwise movement of flagella motor that causes motility in a directed manner either towards an attractant or away from a repellent. Overall, our result shows the HEHEAA increases the directed motility or chemotaxis in *Pseudomonas* GM79, which could induce movement towards the plant-based signal during plant colonization.

5.3.4 HEHEAA increases the abundance of the Type VI secretion system in *Pseudomonas* GM79

Endophytes can promote plant growth by suppressing the biotic stressor (Kandel et al., 2017, Rosenblueth and Martínez-Romero, 2006). To investigate if the endophyte *Pseudomonas* GM79 demonstrates those traits, we looked at the proteins relevant to inter-bacterial competition and bacterial defense. Our analysis identified several significant proteins from the Type VI secretion system (T6SS) that were upregulated in *Pseudomonas* GM79 post-HEHEAA addition. T6SS contains thirteen core proteins that are conserved across T6SS (Filloux et al., 2008, Shalom et al., 2007, Cascales, 2008). Among the 13 core proteins, our analysis identified membrane associated complex TssM (PMI36_04024) which was increased in abundance by more than 8 log₂ fold, T6SS contractile sheath subunit (PMI36_04041) which was increased in abundance by more than 2.8 log₂ fold and baseplate protein TssA (PMI36_04059) which was increased in abundance by more than 1.3 log₂ fold in 3 hrs compared to 1 hrs post-HEHEAA addition. T6SS is a versatile bacterial nanomachine that is widely used as a protein translocating machinery involved in inter-bacterial

competition and bacterial defense (Jani and Cotter, 2010, Alcoforado Diniz et al., 2015). The increase in T6SS-related proteins suggests that plant-derived signal, HEHEAA can promote the endophyte *Pseudomonas* GM79's pathogen-defending ability.

5.3.4 HEHEAA increases the siderophore synthesis and transport in *Pseudomonas* GM79

Some endophytes are known to secrete siderophores (Gaiero et al., 2013) that promote plant growth through protection against microbes or by scavenging iron (Anilkumar et al., 2017). The protein level analysis identified several proteins that were involved in siderophore synthesis and transport after the HEHEAA addition. Siderophore synthetase of IucA/IucC family (PMI36_00407) that is involved in the biosynthesis of the siderophore aerobactin was increased in abundance by more than 2.8 log₂ fold in 3 hrs compared to 1hrs post-HEHEAA addition. Similarly, the siderophore receptor (PMI36_04879) that is involved in the active transport of siderophores across the membrane and energy-dependent siderophore transporter (PMI36_04007) were increased in abundance by more than 3 and 8.6 log₂ fold respectively in 3 hrs compared to 1 hr post-HEHEAA addition. This result suggests that HEHEAA can induce *Pseudomonas* GM79's ability to produce and transport siderophores. Siderophores are low molecular weight metabolites that not only contribute to iron acquisition through iron chelation but are also involved in protection against oxidative stress, interspecies interactions, and biofilm formation (Kramer et al., 2020, Li et al., 2021). A higher amount of microbial siderophore can also help plant growth through iron uptake and can act as a biocontrol agent against phytopathogens (Anilkumar et al., 2017).

5.4 Conclusions

In this study, bottom-up proteomics was utilized to gain a system-wide understanding of the proteins and pathways regulated by the PipR-HEHEAA and its broader role in *Pseudomonas* sp. GM79 plant colonization. Overall, the finding within this study supports the existing model of HEHEAA dependent PipR activity. Besides this, proteomics provides a unique perspective of how the plant endophyte *Pseudomonas* sp. GM79 responds to the HEHEAA. This study shows that upon HEHEAA addition, *Pseudomonas* sp. GM79 mainly modulates biological processes that fall under three categories i.e., chemotactic traits, phytopathogen defending traits, and plant growth

promoting traits through the regulation of protein abundance related to chemotaxis, siderophore synthesis and transport, antibiotic synthesis, and T6SS (**Figure 5.4**). This finding helps to further advance our understanding of how plant-associated bacteria can sense and respond to their environment. While we observed the HEHEAA-dependent PipR activation as expected, there still is a need for further work to verify if the system-wide response we observed is dependent on the PipR binding of HEHEAA or through a different mechanism. Moreover, how much of the system-wide response we observed is influenced by the microbial growth during sampling time points is still unknown. Nonetheless, future experiments on mutant *pipR* and *pipA* will help further clarify these limitations.

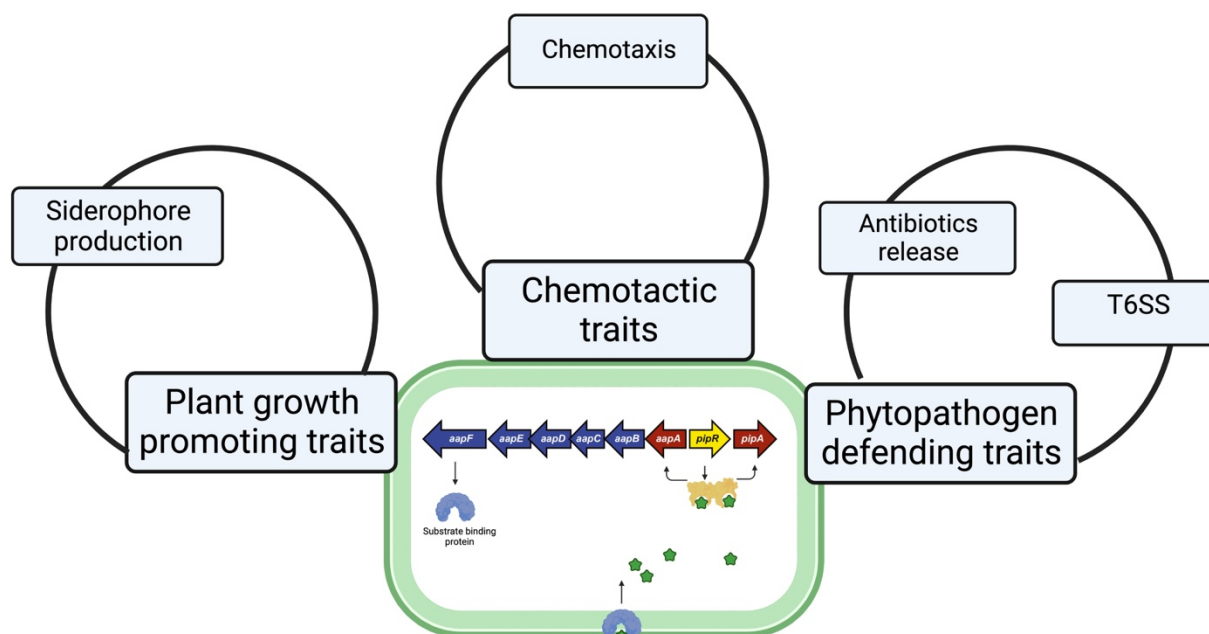


Figure 5.4: Impact of HEHEAA in *Pseudomonas* sp. GM79.

HEHEAA addition mainly modulates three types of traits in *Pseudomonas* sp. GM79 i.e., chemotactic traits, phytopathogen defending traits, and plant growth-promoting traits. It controls these traits through the regulation of chemotaxis, siderophore synthesis and transport, antibiotic synthesis, and T6SS regulation.

CHAPTER 6: TEMPORAL DYNAMICS OF PROTEIN AND POST-TRANSLATIONAL MODIFICATION ABUNDANCES IN POPULUS LEAF ACROSS A DIURNAL PERIOD

A version of this chapter was originally published by Him K. Shrestha:

Him K. Shrestha, Dr. Manuel I. Villalobos Solis, Sara S. Jawdy, Dr. Gerald A. Tuskan, Dr. Xiaohan Yang, Dr. Paul E. Abraham (2021). Temporal dynamics of protein and post-translational modification abundances in *Populus* leaf across a diurnal period. *Proteomics* 21.20: 2100127.

This was the collaborative work with researchers from The Center for Bioenergy Innovation (CBI) group at Oak Ridge National Laboratory. Sample were grown and provided by Sara S. Jawdy, Dr. Gerald A. Tuskan, Dr. Xiaohan Yang.

The proteins and PTMs level regulation of diurnal cycle in poplar plant hasn't been characterized. Him K. Shrestha led this project to characterization this phenomenon. Him K. Shrestha's contribution to this work included: 1) Biological data analysis and figure generation 2) Writing, editing, and revising manuscript for publication.

Notes: Supplemental table mentioned in text are available at DOI: [10.1002/pmic.202100127](https://doi.org/10.1002/pmic.202100127)

6.1 Introduction to the importance of the diurnal cycle

Similar to plant's biotic factors discussed in chapter 3-5, abiotic factors can also influence the plants growth and productivity. *Populus* spp. (poplar) represent a series of dedicated woody biomass feedstocks for advanced biofuels and bioproducts (Reddy et al., 2000, Stanton, 2020). Proper growth and fitness of poplar as a sustainable feedstock depends on timely perception and response to environmental signals (e.g., light, temperature, water). Poplar leaves, like other plants, have evolved oscillating or circadian rhythms that play important roles in synchronizing biological processes with external cues (Reddy et al., 2000, Mayrhofer et al., 2005). These rhythmic patterns play a major role in a plant's phenotypic plasticity, which allows for new phenotypes under changing environments, and synchronize plant behavior across cellular and molecular levels. The characterization of this phenomenon at the proteome level under natural conditions is currently lacking for a model woody plant species. Herein, we employed bottom-up proteomics using high-resolution mass spectrometry and *de novo*-assisted database searching to identify abundance

changes in proteins and post-translational modifications in poplar leaf tissue sampled across a 12/12-hour light/dark diurnal period. The objective of this study was to broadly characterize the proteome and PTMs regulation across the diurnal cycle to generate a rich dataset for the scientific community to utilize as a resource. The goal of the study was to extract overall trends but not a detailed evaluation.

6.2 Proteomics and PTM approach to characterize this phenomenon at a molecular level

P. deltoides WV94 cuttings were rooted and established in the greenhouse, moved to a growth chamber under a 12/12-hour light/dark cycle, with photon flux density of 30 $\mu\text{mol m}^{-2} \text{s}^{-1}$ and day/night temperature 21°C/17°C, respectively. After one week of growth, mature leaf tissue (leaf plastochron index 4-6) was harvested every 2 hours over a 24-hour period in biological triplicate (T00, T02, T04, T06, T08, T10, T12, T14, T16, T18, T20, and T22 hrs). The light was turned on after collecting timepoint 6 and turned off after collecting timepoint 18. After the mid-rib was removed, and leaves were ground under liquid nitrogen using a mortar and pestle, ground samples were suspended in lysis buffer (4% sodium dodecyl sulfate in 100 mM of Tris-HCl, pH 8.0), then boiled for 5 min, sonically disrupted using a Digital Sonifier 250 (Branson Ultrasonics Corporation; 30% amplitude, 10 s pulse with 10 s rest, 2 min total pulse time) and boiled for an additional 5 min. Crude protein extracts were pre-cleared via centrifugation, adjusted to 10 mM dithiothreitol and incubated at room temperature for 30 min, then adjusted to 30 mM iodoacetamide to prevent reformation of disulfide bonds. Proteins were precipitated using chloroform-methanol-water extraction (Jiang et al., 2004) and dried protein pellets were resuspended in denaturing solution (2% sodium deoxycholate (SDC) in 100 mM ammonium bicarbonate). Protein amounts were estimated by performing a bicinchoninic acid (BCA) assay. For each sample, an aliquot of 250 μg protein was digested with trypsin (Promega, 1:75 [wt/wt]) overnight at 37°C under constant shaking and then digested a second time using the same protein to trypsin ratio as before, but this time for 3 hrs at 37°C under constant shaking. Resulting peptide mixtures were adjusted to 1% formic acid to precipitate SDC. Hydrated ethyl acetate was added to each sample at a 1:1 (vol/vol) ratio three times to effectively remove the detergent. Samples were then placed in a SpeedVac concentrator (Thermo Fisher Scientific) to remove remaining ethyl acetate. Dried peptides samples were additionally cleaned/concentrated on Pierce peptide desalting

spin columns (Thermo Scientific) as per the manufacturer's instructions. Peptide concentrations were obtained by the BCA assay and placed in autosampler vials.

All samples were analyzed on a Q Exactive Plus mass spectrometer (Thermo Fischer Scientific) coupled with a Proxeon EASY-nLC 1200 liquid chromatography (LC) pump (Thermo Fisher Scientific). Peptides were separated on a 75 μm inner diameter microcapillary column packed with 40 cm of Kinetex C18 resin (1.7 μm , 100 Å, Phenomenex) that was heated to 60°C in a Phoenix S&T NanoLC column heater. For each sample, 2 μg aliquots of peptides were loaded in buffer A (0.1% formic acid, 2% acetonitrile) and eluted with a linear 240 min linear gradient from 2 – 20% buffer B (0.1% formic acid, 80% acetonitrile), followed by an increase in buffer B to 30% for 10 min, and concluding with a 10 min equilibration to 98% buffer A. The flow rate was kept at 200 nL/min. MS data was acquired with the Thermo Xcalibur software v4.27.19 using the topN method where N could be up to 10. Other spectral data acquisition parameters have been reported before (Villalobos Solis et al., 2020).

Tandem mass spectrometry data was searched with *de novo*-assisted database searching against the *P. trichocarpa* v3.1 reference proteome (Goodstein et al., 2012) accompanied with common contaminant proteins using PEAKS DB, PEAKS PTM and PEAKS SPIDER in PEAKS X Pro Studio (Bioinformatics Solutions, Waterloo, Canada). The peptide and fragment ion mass tolerances were set to ± 10 ppm and ± 0.02 Da, respectively. Trypsin was set as the enzyme parameter. Features associated with chimera scan were enabled. *De novo* ALC score was set at $>90\%$. For PEAKS DB, carbamidomethylation of cysteine was set as fixed modification and oxidation of methionine was set as a variable modification. The PEAKS PTM algorithm was used to identify other types of modifications by allowing the search against all possible modifications from the Unimod database (Creasy and Cottrell, 2004). Similarly, PEAKS SPIDER algorithm was used to detect possible homology peptide mutations. Protein functions were annotated with OmicsBox (<https://www.biobam.com/omicsbox/>). All peptides (with or without modification) were required to pass a false discovery rate (FDR) threshold of $< 1\%$.

6.3 Results

6.3.1 Proteome-wide changes across the diurnal period

At a 1% FDR, we identified 52,343 peptide sequences across the experiments that provided insight into the qualitative state of 12,983 functional proteins. The numbers of proteins identified in a sampling time point ranged from approximately 6,300 to 9,900 (**Figure 6.1A**). Principal component analysis (PCA) of diurnal proteome data showed grouping between dark and light conditions rather than individual timepoints (**Figure 6.1B**). To quantify protein abundance changes across the 24-hour period, identified proteomes were filtered by requiring proteins to have at least 1 unique peptide. Protein abundances were log₂-transformed, LOESS normalized between the biological replicates, and mean-centered across the entire dataset using the InfernoRDN software (Polpitiya et al., 2008). Proteome datasets were filtered to improve the robustness of the downstream analysis by removing proteins that were not quantified in at least two out of three biological replicates of at least one time point. The abundance for proteins with missing abundance values were imputed with random values drawn from the normal distribution (width 0.3, downshift 2.8) of proteome datasets from the previous step using the Perseus software v.1.6.12.0 (Tyanova et al., 2016).

An analysis of variance (ANOVA) test was performed to identify changes in protein abundance across the 24-hour time series. In total, 8,686 proteins were used for quantification, out of which 1,032 showed the significant diurnal changes using ANOVA (p-value < 0.05). The abundance values for the significantly changing proteins were standardized (z-score) and clustered using the fast ward method from Euclidian distance metric. Clustering identified 5 major patterns of protein abundance changes across the 24-hour period (**Figure 6.2**): Cluster 1 proteins experience an initial increase in abundance at the beginning for the photoperiod and a significance decrease in abundance in the middle of the dark period; Cluster 2 proteins increase in relative abundance throughout the light period; Cluster 3 protein abundances show substantial changes in relative abundance during day/night transitions; Cluster 4 protein abundances are abundant in the dark, then decrease throughout the light period; Cluster 5 proteins experience a significant decrease in

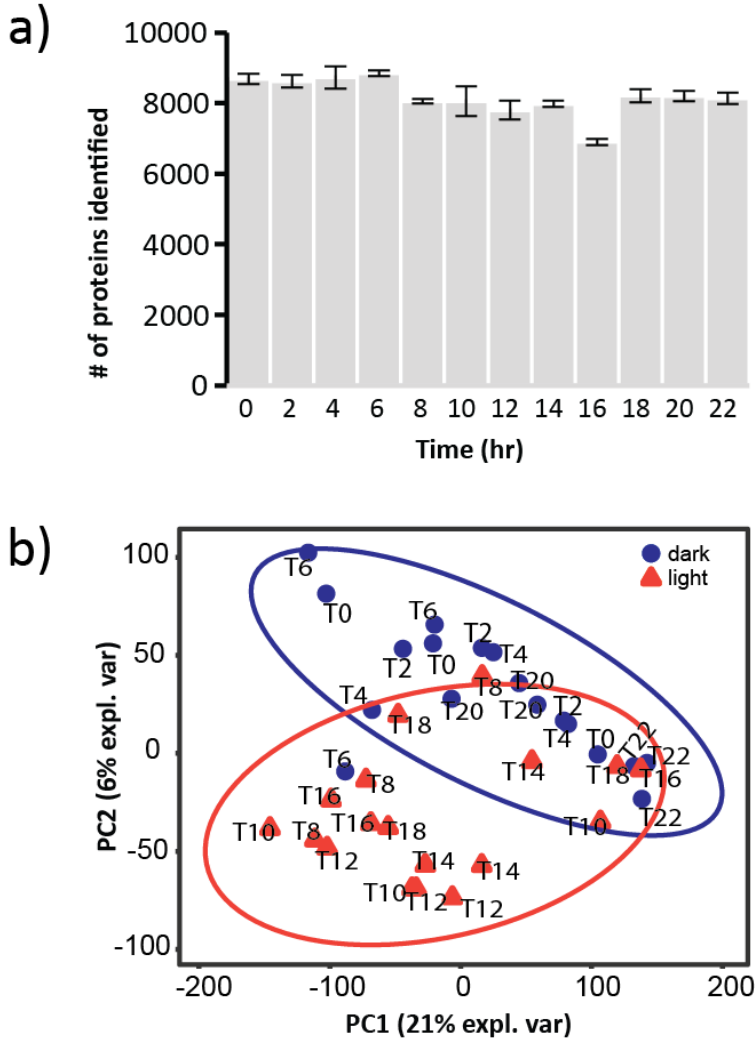


Figure 6.1: Qualitative assessment of poplar leaf proteome across a 12/12-hour light/dark period.

(A) The number of proteins identified across each 2-hour sampling time point. The number of proteins identified ranges from around 6,300 to 9,900. The error bars represent the standard error between biological triplicates. (B) Principal component analysis (PCA) for the entire dataset (36 samples). The ellipse represents 90% confidence interval. Each circle (blue) represents a proteome sampled from the dark period (T0-T6 and T20-T22) and each triangle (red) represents a proteome sampled from the light period (T8-T18).

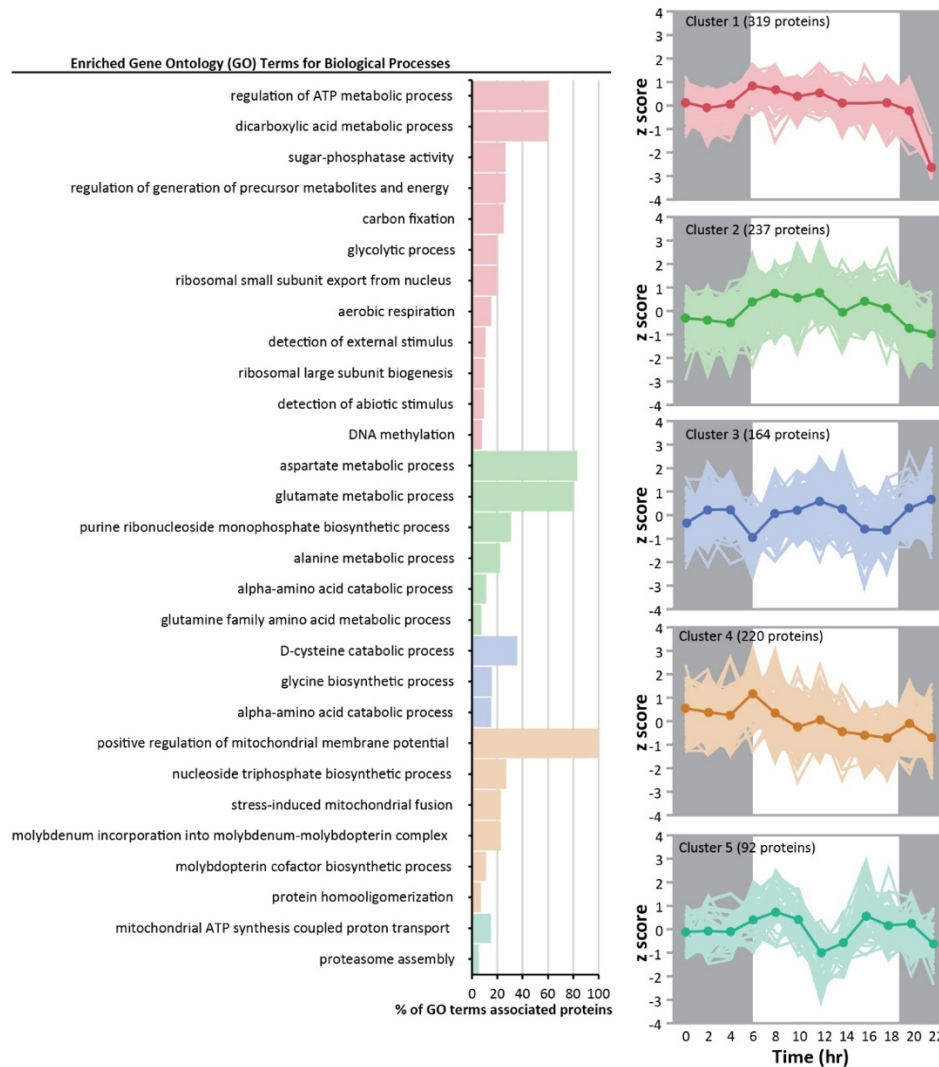


Figure 6.2: Quantitative assessment of poplar leaf proteome abundance profile changes across a diurnal period.

Differentially expressed proteins were identified using ANOVA with a significant threshold set at p -values < 0.05 . Overall, 1,032 proteins significantly changed across 12 timepoint measured. Five major trends were apparent among the significantly changing proteins. Enriched biological processes were assessed across the major protein trends using the software ClueGO and setting a significance threshold to an adjusted p -value < 0.05 , a minimum GO level 0 and maximum GO level 20, and a kappa score of 0.4. From the GO groups obtained from ClueGO, a single representative was selected to reduce term redundancies.

abundance in the middle of the light period. Proteins representing each trend were extracted and analyzed to identify enriched biological processes (**Figure 6.2**) among the observed abundance trends. The enrichment analysis was performed by the ClueGO software and over-represented gene ontology (GO) terms were identified using Benjamini-Hochberg corrected p-value thresholds of 0.05 (Bindea et al., 2009).

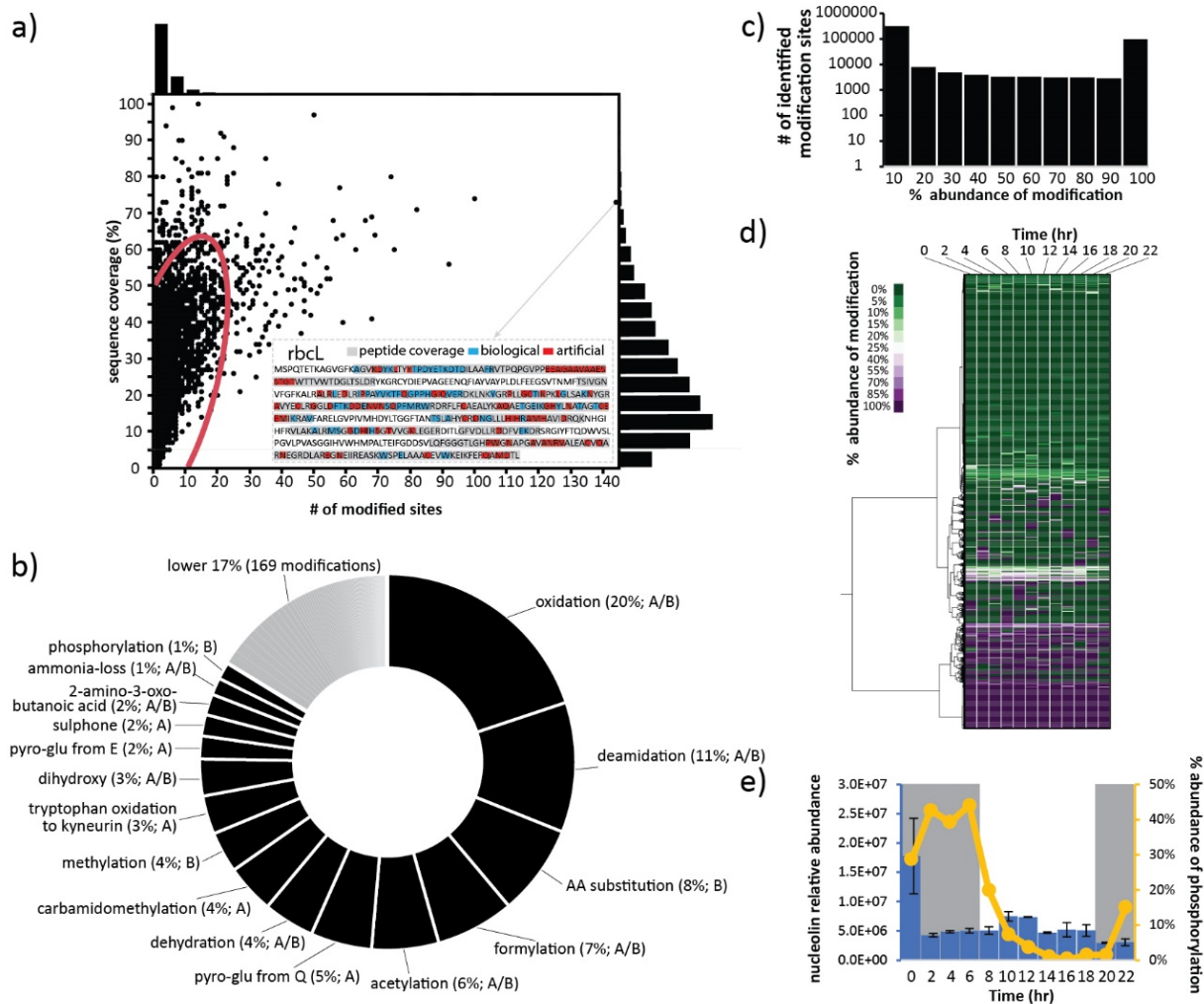
6.3.2 PTM analysis provides the unique regulation throughout the diurnal cycle

In addition to protein abundance changes, the PEAKS X Pro software profiles post-translational modifications (PTMs) and amino acid mutations in all identified proteins. **Figure 6.3A** shows the relationship between protein sequence coverage and the number of PTM sites observed per protein across the entire experiment. It is important to note that each of the reported number of identified sites and modifications for a particular protein represents the average sampling of all its natural, biologically relevant modifications (e.g., phosphorylation) as well as artificial modifications (e.g., cysteine carbamidomethylation) introduced during sample preparation. For any given encoded protein, a central challenge is regarding determination of how many protein-level variations and PTMs occur (Han et al., 2011). The number of distinct proteins in a cell (i.e., proteoforms) can be difficult to assess because bottom-up proteomics measurements sample the entire population of proteoforms, and the result is an averaged perspective (Aebersold et al., 2018). For proteins that are relatively abundant, there is a greater opportunity to sample deeper into the proteoform population. For example, we identified 144 unique modification sites for the chloroplast-encoded Rubisco large subunit (rbcL), which has been shown to be heavily decorated with various PTMs including methylations, deformylations and acetylations that remain functionally undefined until this day (Houtz et al., 2008). Nevertheless, by assessing the types of modifications and the relative abundance difference between modified and unmodified versions, a semi-quantitative estimate of a given proteoform can be obtained.

Across the entire dataset, we observed putative identifications for 144 modifications and 253 amino acid substitutions. In general, oxidation was the most frequently occurring modification across the entire experiment (**Figure 6.3B**). While this modification can be introduced during sample preparation, it is interesting to note that a test for cellular compartment enrichment revealed

Figure 6.3: Overview of poplar leaf protein modifications observed across a diurnal period.

(A) The sequence coverage and number of identified modification sites were plotted against each other for each protein identified across the entire dataset. A histogram is provided juxtaposition to each axis to illustrate the distribution of proteins identified in the context of sequence coverage or number of modification sites. The red line represents 95% of the data. As one example of protein with extensive sequence coverage and PTMs identification, the figure inset shows coverage for peptides and modifications for the chloroplast-encoded rubisco large subunit (rbcL). (B) Across the entire dataset, 80% of the site modifications observed mapped to 16 different PTM classes and the remaining 20% of identified sites observed were not delineated. Letters A and B in parentheses represents artificial and biological PTMs, respectively, when specific amino acid localization determines if the PTMs is artificial vs biological. (C) For each modification site identified in the entire dataset, the relative difference between modified and unmodified abundances was determined per timepoint, which equals to the abundance of modified isoform divided by the summed abundance of modified and unmodified isoforms. These percentages were then binned into deciles for the entire experiment. (D) One-way hierarchical clustering (Fast ward method) was performed on the percentage of modified versus unmodified isoforms for each site (rows) across each time point (columns, right to left from T00 to T22). (E) An example of diurnal changes in protein abundance (blue; summed total of all peptide abundances observed) and phosphorylation abundance (yellow; summed abundance of peptides covering residue S136) for nucleolin (accession Potri.012G090200). Error bar for protein abundance represents standard errors.



many oxidized proteins predicted to be localized to the chloroplast and mitochondria, which harbor extensive reduction-oxidation (redox) reactions (Foyer, 2018). To assess the relative abundance of PTM modifications, the percentage abundance ratio between modified and unmodified residue sites were calculated (**Figure 6.3C**). Interestingly, there was almost an equal number of sites with modifications that were relatively low in abundance (< 10% of total site abundance) or quite prevalent (100% of total site abundance). Moreover, many modification abundances were quite dynamic across the diurnal period while others were static (i.e., amino acid residues that are always modified) (**Figure 6.3D**).

6.3.3 An instance of dynamic regulation at both the protein and PTM level and its biological implications

While it is beyond the scope of this dataset to describe in detail the implications of these findings across the entire proteome, we provide an example to illustrate dynamic regulation at both the protein and PTM-level. Protein and phosphorylation abundances were observed to be diurnally regulated for the protein nucleolin (accession Potri.012G090200) (**Figure 6.3E**), which is a ribosome biogenesis protein that regulates rRNA transcription and expression. It has been shown to affect plant development and is required for DNA methylation states in *Arabidopsis thaliana* (Petricka and Nelson, 2007). Interestingly, studies using human tumor cells have demonstrated that nucleolin phosphorylation regulates the mRNA decay during cellular stress (Zhang et al., 2018).

6.4 Conclusions

The ability to study temporal regulation of both protein and PTM abundances in parallel provides a more complete understanding of how plants synchronize molecular mechanisms with environmental signals such as light to control the timing of physiological and metabolic processes. As such, information from this dataset serves as a valuable resource and foundational dataset for future biofeedstock improvements in yield and growth (Yang, 2020) as diurnal regulation of proteins is known to control responses to drought, UV light, and temperature. For example, the information from this study could be used to identify biological parts for building stress-inducible

gene circuits to increase biomass yield and cold, drought and salinity stress tolerance, as inspired by previous studies (Nakamichi et al., 2016).

More broadly, this dataset can be leveraged for valuable cross-species comparisons that study how evolution has shaped the diurnal regulation of proteins and PTMs (Abraham et al., 2016). By comparing this poplar dataset with recently published diurnal proteomics experiments performed in *Arabidopsis* (Krahmer, 2021, Uhrig et al., 2019), we identified several PTMs that have evidence for abundance changes across a diurnal period in both organisms and thus can be considered candidates for functionally important regulatory events underpinning phenotypic plasticity. For example, in both organisms we observed a diurnally regulated phosphorylation event for the enzyme pyruvate orthodiphosphate dikinase that plays an important role in carbon flow in photosynthetic plants (Chastain et al., 2008, Taylor et al., 2010).

Overall, this proteomics dataset revealed dynamic changes in the poplar leaf proteome across a diurnal period. Because the rhythmic patterns and oscillators play an important role in regulating numerous biological processes of agricultural and ecological importance (e.g., photosynthesis, plant development, and tolerance to various abiotic and biotic stresses) (Dakhiya et al., 2017, Srivastava, 2019), this dataset brief provides new opportunities to dissect phenotypic plasticity at the molecular level.

CHAPTER 7: MULTI-OMIC CHARACTERIZATION OF BIFUNCTIONAL PEROXIDASE 4-COUMARATE 3-HYDROXYLASE KNOCKDOWN IN *BRACHYPODIUM DISTACHYON* PROVIDES INSIGHTS INTO LIGNIN MODIFICATION-ASSOCIATED PLEIOTROPIC EFFECTS

A version of this chapter was originally published by Him K. Shrestha:

Him K. Shrestha, Dr. Yosef Fichman, Nancy L. Engle, Dr. Timothy J. Tschaplinski, Dr. Ron Mittler, Dr. Richard A. Dixon, Dr. Robert L. Hettich, Dr. Jaime Barros and Dr. Paul E. Abraham (2022). Multi-omic characterization of bifunctional peroxidase 4-coumarate 3-hydroxylase knockdown in *Brachypodium distachyon* provides insights into lignin modification-associated pleiotropic effects. *Frontiers in Plant Science*. Volume 13 <https://doi.org/10.3389/fpls.2022.908649>.

This was the collaborative work with researchers from The Center for Bioenergy Innovation (CBI) group at Oak Ridge National Laboratory/University of North Texas. Sample were grown and provided by Dr. Jaime Barros and Dr. Richard A. Dixon. Originally, science questions were also driven by Dr. Jaime Barros and Dr. Richard A. Dixon. Nancy L. Engle and Dr. Timothy J. Tschaplinski measured and analyzed metabolomics results. Dr. Yosef Fichman and Dr. Ron Mittler from the University of Missouri, Columbia assisted with ROS imaging and H₂O₂ measurement.

Lignin modification in plants often introduces unwanted phenotypic effects. Him K. Shrestha led the project to understand the underlying mechanism of how lignin biosynthesis enzyme C3H/APX knockdown induce pleiotropic effects in *Brachypodium*. Him K. Shrestha's contributions in this project were: 1) Proteomics sample preparation and measurement 2) Biological data analysis and figure generation, 3) Writing, editing, and revising manuscript.

Notes: Supplemental figures and tables mentioned in text are available at:

<https://www.frontiersin.org/articles/10.3389/fpls.2022.908649/full#supplementary-material>

7.1 Introduction

Lignin is a natural, branched plant phenolic heteropolymer generated by radical coupling of monomers called monolignols, mainly coniferyl, sinapyl, and *p*-coumaryl alcohols, which creates a gluing matrix for cellulose fibers in higher plant cell walls (Boerjan et al., 2003, Ralph et al.,

2004). In general, lignin is a crucial secondary plant cell walls component, providing structural and physical integrity to support plant growth and development, and also functions in plant defense (Boerjan et al., 2003, Whetten and Sederoff, 1995, Zhao and Dixon, 2014). Additionally, lignification benefits plants by preserving cell wall integrity against a range of environmental stressors (Barros et al., 2015, Hilal et al., 2004, Fan et al., 2006, Delessert et al., 2004, Malinovsky et al., 2014, Lee et al., 2019); however, this makes it challenging to access the available cellulosic biomass for the production of biofuels and other bioproducts (Velvizhi et al., 2022, Beckham et al., 2016, Ragauskas et al., 2014).

A more complete understanding of the molecular processes underlying lignification would benefit future technological developments aiming to deconstruct plant cell walls. Yet, efforts to reduce lignin content and/or alter lignin substructures to alleviate the polymer's recalcitrant properties often introduce several pleiotropic effects such as stunted plant growth and associated yield penalty (Ha et al., 2021, Muro-Villanueva et al., 2019, Vanholme et al., 2019). Unfortunately, the precise molecular mechanisms underlying the negative effect on plant growth and development in low lignin mutants and transgenics are not well understood. Prevalent hypotheses include restricted nutrient transport through collapsed xylem vessels (Man Ha et al., 2019, Muro-Villanueva et al., 2019), overaccumulation/overproduction of bioactive molecules (El Houari et al., 2021, Bonawitz and Chapple, 2013), deficiency of metabolites critical for normal plant growth (Bonawitz and Chapple, 2013, Muro-Villanueva et al., 2019), and cell wall integrity sensing and stress signaling (Ha et al., 2021, Xie et al., 2018, Gallego-Giraldo et al., 2020) leading to growth defects. Because of the availability of more robust genetic resources, most of these studies have thus far been focused on model dicot plants, such as *Arabidopsis* and alfalfa, but our knowledge on lignin modification-induced dwarfism in monocotyledonous grass species remains largely unexplored.

Lignin metabolism in grasses differs substantially from that in eudicots (Dixon and Barros, 2019). One difference is that the phenylpropanoid pathway intermediate 4-coumarate is formed entirely from deamination of phenylalanine followed by 4-hydroxylation in dicots, whereas it derives partially from direct deamination of tyrosine in grasses (Barros et al., 2016). In addition, a central lignin gene/enzyme in dicots, caffeoyl-shikimate esterase (CSE), is absent in many grass species (Ha et al., 2016, Vanholme et al., 2013). Recent findings uncovered the existence of a bifunctional

peroxidase enzyme, coumarate 3-hydroxylase (C3H/APX), which provides a parallel route to catalyze the direct conversion of 4-coumarate to caffeate in the early steps of lignin biosynthesis in *Brachypodium distachyon* (Barros et al., 2019). C3H/APX is the only soluble non-membrane-bound hydroxylase in the lignin pathway reported to date and uses molecular oxygen and ascorbate as a reducing agent to oxidize phenolic substrates like 4-coumarate. To characterize the role of C3H/APX, the knockdown of C3H/APX (C3H/APX-KD) was generated by T-DNA insertion in *Brachypodium* (Hsia et al., 2017). Along with a markedly reduced lignin content, like many other lignin-deficient plants, the C3H/APX-KD exhibited a negative growth phenotype along with higher percentage of non-viable seeds and delayed senescence, providing a suitable biological system to study growth defects associated with lignin depletion in grasses (Barros et al., 2019).

Grass stems contain significantly more lignified cell wall material as compared to leaves and seeds, providing us a rationale to focus the present study in stem tissue (Sarath et al., 2008). Similarly, high-resolution proteomics allows the large-scale identification and quantification of *Brachypodium* proteins for a system level interrogation of altered biological processes, which will help to connect molecular change to the observed phenotype (Aebersold and Mann, 2016). The objectives of this study were to i) characterize the changes in phenylpropanoid biosynthetic enzymes resulting from knockdown of C3H/APX protein, and ii) gain a deeper understanding of how knockdown of C3H/APX protein leads to negative growth phenotype in the model grass *Brachypodium*. Therefore, in this study, bottom-up proteomics was utilized to characterize the functional importance of C3H/APX in mature stem tissues of *Brachypodium*. In support of these measurements, transcriptomics data and metabolomic measurements were obtained to provide additional lines of evidence. From this work, we hypothesize that C3H/APX links the cytosolic route of monolignol precursor hydroxylation to reactive oxygen species metabolism and knockdown of C3H/APX to induce ROS associated growth-defense trade-offs.

7.2 Materials and methods

7.2.1 Plant material and growth conditions

Brachypodium knockdown lines were obtained from the JGI Brachypodium T-DNA collection (Hsia et al., 2017), and selected as reported previously (Barros et al., 2019). The Brachypodium C3H/APX knockdown line (JJ25124) was identified as an activation tagging silencing mutant transformed with the pJJ2LBA vector and with the T-DNA inserted into the last intron of the *c3h/apx* gene Bradi1g65820 (I1H6P1). Similarly, Brachypodium APX3 knockdown line (JJ22251) transformed with the same pJJ2LBA vector has the T-DNA inserted near the stop codon of the microsomal ascorbate peroxidase 3 (APX3, Bradi3g42340, I1I9A3) and was used as a T-DNA control. The parental line of the T-DNA mutant population (accession Bd21-3) was used as non-transgenic WT control. Briefly, Brachypodium seeds were sown in the soil, vernalized in the dark at 4 °C for 3 days and then moved to a growth chamber with under 70% humidity, and 16-h-light:8-h-dark photoperiod at 24 °C, and light intensity of 100 $\mu\text{Em}^{-2} \text{ s}^{-1}$. Plantlets were covered with a dome and kept well-watered for 1 week. After this time, plantlets were transferred to half-gallon pots containing Metro-Mix 360 (Sun Gro Horticulture) and grown in the greenhouse for 12 weeks with 70% humidity and 14-h light at 25 °C, 10-h dark at 22 °C, supplemented with photosynthetically active radiation (PAR) lights when the range was out of 40–120 $\mu\text{Em}^{-2} \text{ s}^{-1}$ during the day. The growth and lignin phenotypes of C3H/APX-KD Brachypodium lines were assessed and were consistent with the phenotypes of the previously characterized and published C3H mutant in multiple alleles of *A. thaliana* (Barros et al., 2019). Mature stem tissues from 12-week-old plants were collected for the proteomics measurement.

7.2.2 Proteomics sample preparation

Ground Brachypodium stem tissue samples were suspended in lysis buffer (4% sodium dodecyl sulfate + 10 mM dithiothreitol in 100 mM ammonium bicarbonate (ABC)). Samples were boiled for 5 min at 90°C, sonicated (Branson 450 Digital Sonifier with 30% amplitude, 10 s pulse with 10 s rest, 1 min total pulse time) and boiled for an additional 5 min at 90°C. Samples were centrifuged and supernatants were collected, which were then alkylated by incubating with 30 mM

iodoacetamide for 15 min in the dark at room temperature. After alkylating the samples, proteins were then extracted using a chloroform-methanol extraction protocol (Jiang et al., 2004). The extracted protein layer was washed using 100% methanol, air-dried, and reconstituted in 2% sodium deoxycholate (SDC) solution made in 100 mM ABC. Protein concentration was measured using a NanoDrop OneC spectrophotometer (Thermo Scientific). Two hundred and fifty ug of total protein for each sample were digested with two aliquots of sequencing grade trypsin (Promega, 1:75 [wt/wt]) at two different sample dilutions, 2% SDC (3 h) and subsequently 1% SDC (overnight). After digestion, SDC was removed by precipitating with 1% formic acid followed by an ethyl acetate wash. Samples were then lyophilized in a SpeedVac concentrator (Thermo Fischer Scientific). Peptide samples were desalted on Pierce peptide desalting spin columns (Thermo Scientific) as per the manufacturer's instructions and then resuspended in 0.1% formic acid solution.

7.2.3 Proteomics LC-MS/MS analysis

Peptides were analyzed by liquid chromatography and tandem mass spectrometry (LC-MS/MS) as previously described (Villalobos Solis et al., 2020) using a Proxeon EASY-nLC 1200 liquid chromatography pump (Thermo Fisher Scientific) coupled to a Q Exactive Plus mass spectrometer (Thermo Fischer Scientific). Briefly, peptides were separated on an in-house-pulled nanospray emitter of 75 μm inner diameter packed with 30 cm of Kinetex C18 resin (1.7 μm , 100 $\text{\AA}$, Phenomenex). For each sample, 2 μg aliquots of peptides were loaded in buffer A (0.1% formic acid, 2% acetonitrile) and separated with a linear 210 min organic gradient, followed by a wash and re-equilibration step: 0% to 2% solvent B over 27 min, 2% to 25% solvent B over 148 min, 25% to 50% solvent B over 10 min, 50% to 0% solvent B over 10 min, hold at 0% solvent B for 15 min. The flow rate was kept at 250 nl/min. MS data were acquired with the Thermo Xcalibur software v4.27.19 using the top 10 data-dependent acquisition method. Other spectral data acquisition parameters have been reported before (Villalobos Solis et al., 2020).

7.2.4 Protein identification and quantification

Raw files were analyzed via Proteome Discoverer v2.5 using MS Amanda 2.0 (Dorfer et al., 2014) and confidence in peptide-to-spectrum (PSM) matching was evaluated by Percolator (Kall et al., 2007). Spectral data were searched against the *Brachypodium distachyon* reference proteome database (downloaded on 09/12/2019) from UniProt to which common laboratory contaminants were appended. The following parameters were set up in MS Amanda to derive fully tryptic peptides: MS1 tolerance = 5 ppm; MS2 tolerance = 0.02 Da; missed cleavages = 2; Carbamidomethyl (C, + 57.021 Da) as static modification; and oxidation (M, + 15.995 Da) as dynamic modifications. The percolator false discovery rate (FDR) threshold was set to 1% at the PSM and peptide levels. FDR-controlled peptides were mapped to their respective proteins and protein-level abundance were calculated by summing together peptide extracted ion chromatograms. For quantitative analyses, proteins were required to have at least 1 unique peptide. Protein abundances were LOESS normalized between the biological replicates and mean-centered across the entire dataset on log₂-transformed values using the InfernoRDN software (Polpitiya et al., 2008). Proteome datasets were further filtered to improve the robustness of the downstream analysis by removing proteins that were not quantified in at least two biological replicates in at least one condition. The missing abundance values on proteins were then imputed with random values drawn from the normal distribution (width 0.3, downshift 2.8) using the Perseus software v.1.6.12.0 (Tyanova et al., 2016). Protein abundances were tested for significant changes across samples by performing an analysis of variance (ANOVA) test followed by Tukey HSD binary comparisons using an in-house written python code using scipy and statsmodels. A protein was considered significantly changing in abundance in C3H/APX-KD line only if it: i) passes a significant threshold of 0.05 in ANOVA, ii) passes a significant threshold of 0.05 in binary comparison of C3H/APX-KD line to WT as well as T-DNA control, and iii) doesn't pass the significant threshold of 0.05 in binary comparison of WT and T-DNA control.

7.2.5 Functional enrichment analysis

The differentially abundant proteins data matrix was divided into two groups to contain upregulated and downregulated proteins by hierarchical clustering using the fast Ward's method.

These two groups were used for analyzing enriched gene ontology (GO) terms using the g:Profiler tool (<https://biit.cs.ut.ee/gprofiler/gost>) (Raudvere et al., 2019). In g:Profiler, *Brachypodium distachyon* was used for the organism, Benjamini-Hochberg FDR of 0.05 was used for significance threshold to correct for multiple hypothesis testing. All other parameters were kept to default. Enriched GO terms were further filtered for redundancy using REVIGO (Supek et al., 2011).

Protein class overrepresentation test was performed using PANTHER v16.0 (Mi et al., 2021). All proteins in the *Brachypodium distachyon* database were used for the background, and a statistical test was done using Fisher's exact test followed by a PANTHER specific FDR correction with a significance threshold of 0.05. All proteins in the *Brachypodium distachyon* database instead of only proteins detected in our search were utilized as background because all the proteins from the database had a chance to be identified in our database search approach. Nonetheless, we acknowledge that utilizing sets of proteins detected in MS run as a background could potentially avoid developmental stage or tissue specific enrichment biasness.

7.2.6 Untargeted metabolic profiling

Lyophilized and ground *Brachypodium* stem samples (~10 mg) were weighed and extracted using 2 ml of 80% ethanol and 50 μ l of internal standard (sorbitol 1 mg ml⁻¹) as previously described (Barros et al., 2019). One microgram samples were injected into a GC-MS Agilent 5975C inert XL operated in electron impact (EI; 70 eV) ionization mode as described previously (Tschapinski et al., 2012). Metabolite peaks were extracted using a characteristic mass-to-charge (m/z) ratio to minimize integration of co-eluting metabolites. The extracted peaks were scaled back to the total ion current (TIC) using predetermined scaling factors. Peaks were quantified by area integration and normalized to the quantity of internal standard recovered, amount of sample extracted, derivatized, and injected. A user-created database (>2400 spectra) of EI fragmentation patterns of TMS-derivatized compounds and the Wiley Registry 10th Edition/NIST 2014 Mass Spectral Library were used to identify the metabolites in the samples. Unidentified metabolites were designated by their retention time and key m/z ratios.

7.2.7 RNA-seq data analysis

Brachypodium and Arabidopsis C3H/APX-KD and WT stem samples were used for the RNA-seq analysis. Fragments Per Kilobase of transcript per Million mapped reads (FPKM) output data was processed using the tag count comparison (TCC) R package through TCC-Graphical User Interface (TCC-GUI v.2021.11.13) (Su et al., 2019). The Trimmed Mean of M-values (TMM) normalization method with the Bioconductor package edgeR were used to filter differentially expressed genes (DEGs) (Su et al., 2019). Genes with count lower than 15 FPKM were filtered. A p-value < 0.05 and log₂FC of greater than 1 or less than -1 were considered significant. The log₂FC values were based on “M Value” column (provides log-ratios = log₂ (Group#2/Group#1) from TCC-GUI. Genes differentially expressed were analyzed for Gene Ontology (GO) enrichment for biological process using AgriGO under Poaceae group (Du et al., 2010, Tian et al., 2017). *Brachypodium distachyon* (Phytozome v11.0) was used for the reference database and all other parameters were set at default (Du et al., 2010, Tian et al., 2017).

7.2.8 Reactive oxygen species (ROS) and H₂O₂ measurement

Brachypodium mutant and control plants were germinated and grown for 4-weeks in Jiffy peat pellets under controlled conditions of 25°C and 10-h light/14-h dark cycle, 50 µmol photons s⁻¹ m⁻². ROS measurement of 4-week-old Brachypodium mutant and control plants didn't demonstrate significant changes therefore, these plants were subjected to oxidative stress to see if they have differences in handling oxidative stress. To impose oxidative stress, plants were fumigated for 30 min with 1 µM paraquat and 50 µM H₂DCFDA in 50 mM phosphate buffer (pH 7.4) and 0.001% Silwet L-77 using a nebulizer in a glass container as previously described (Fichman et al., 2019). After fumigation, plants were exposed to high light stress (1,100 µmol photons s⁻¹ m⁻¹) using an array of LED lights (BESTVA) for 30 min. Plants were immediately placed in a fluorescence imager (IVIS Lumina S5) and imaged for ROS detection (excitation/emission: 480 nm/520 nm). The analysis of the images was conducted using the Living Image software (PerkinElmer). Following the imaging, the aerial parts of the plants were fast-frozen in liquid nitrogen and ground to a fine powder for H₂O₂ measurement using Amplex Red (Fichman et al., 2022). Briefly, two hundred microliters of 0.1% trichloroacetic acid was added to

ground tissues and samples were centrifuged at 12,000 g for 12 min at 4°C. Following the centrifugation, 100 µl of supernatant was mixed with 400 µl of 100 mM sodium phosphate buffer (pH 7.4) for pH balance. One hundred microliters of supernatant were transferred into fresh tubes and incubated with 100 µl of phosphate buffer containing 50 µM Amplex Red and 0.05 U ml⁻¹ horseradish peroxidase for 30 min at room temperature in the dark. The concentration of H₂O₂ in each sample was quantified by fluorescence spectrophotometry using a Qubit Fluorometer (Invitrogen) from a standard curve from 0-25 µM of H₂O₂. Following the measurement, tissue samples were completely dried using a speed vacuum concentrator at 60°C for 2h min and H₂O₂ accumulation per mg dry weight was calculated.

7.3 Results

7.3.1 C3H/APX knockdown leads to substantial alterations in the *Brachypodium* stem proteome

To better understand the molecular implications of the C3H/APX knockdown (**Figure 7.1A and 1B**), we analyzed the *Brachypodium* C3H/APX-KD, T-DNA controls, and wild-type controls (WT) using comparative proteomics. Proteomics measurement identified C3H/APX protein with 56% coverage (**Supplement Figure 7.1A**) and as expected, the C3H/APX protein showed ~60% and 65% lower abundance in the *Brachypodium* C3H/APX-KD when compared to the WT and TDNA controls, respectively (**Figure 7.1C**). Besides the decrease in C3H/APX protein abundance, *Brachypodium* C3H/APX-KD also showed a significant reduction of microsomal APX (mAPX), stromal APX (sAPX), and thylakoid membrane bound APX (tAPX) protein abundances (**Supplement Figure 7.1B**).

At a 1% false discovery rate, 12,164 proteins were identified across genotypes, with the highest number (~10,300) of proteins being observed in the C3H/APX-KD, followed by the WT (~9,900) and T-DNA control (~9,700) (**Figure 7.2A and Supplement Table 7.1**). Approximately 70% of the proteins identified were present in all sampling conditions with 4.6%, 3.7%, and 8.0% proteins identified uniquely in C3H/APX-KD lines, WT, and T-DNA controls, respectively (**Figure 7.2B**). Principal component analysis (PCA) of proteome data showed discrete grouping between the

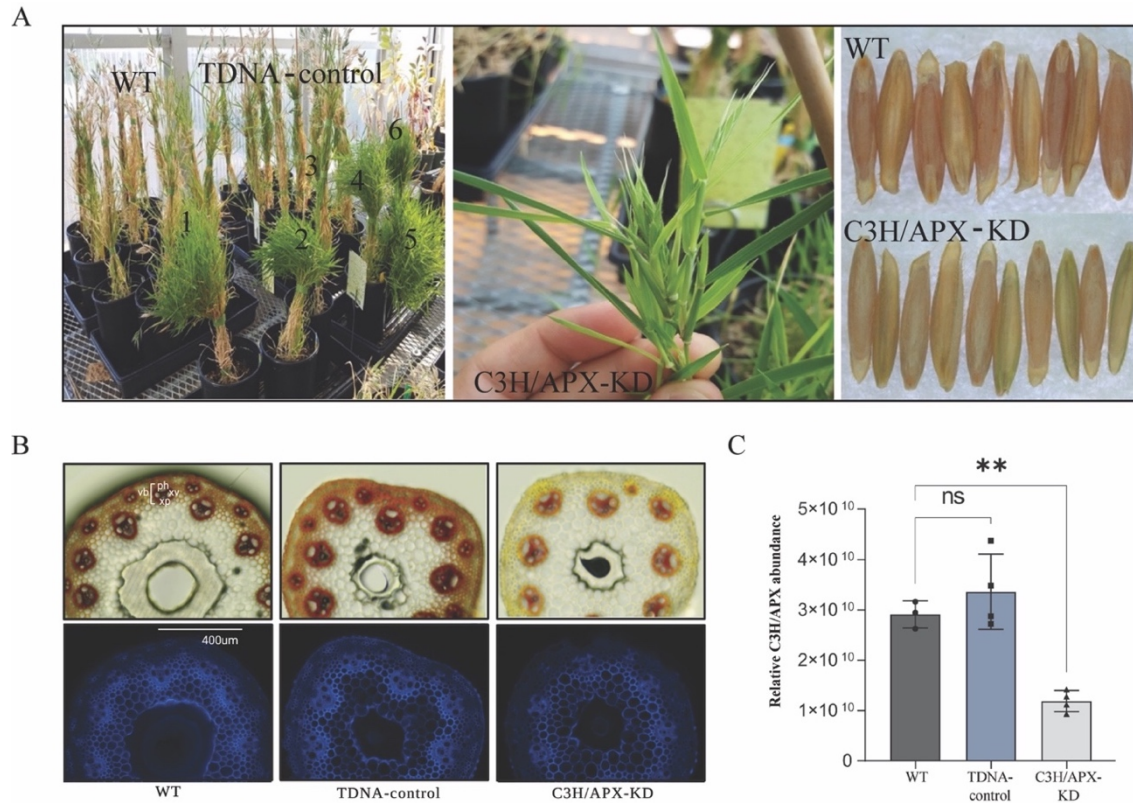


Figure 7.1: Phenotypic characterization and Brachypodium C3H/APX protein abundance profile across experimental conditions.

(A) C3H/APX-KD lines (numbered 1-6 in panel 1) exhibit stunted growth, delayed leaf senescence (stay-green phenotype) and aborted seeds. (B) Transverse sections (phloroglucinol-HCl staining and UV autofluorescence) of Brachypodium stems from wild-type (WT), -DNA controls and C3H/APX-KD lines. In the phloroglucinol-HCL image, G-lignin is red stained; S lignin is peach orange stained. 'ph' represents phloem; 'xv' represents xylem vessels; 'xp' represents xylem parenchyma and 'vb' represents vascular bundle. UV autofluorescence images were captured with an EVOS FL Cell Imaging System (Thermo Fisher) equipped with a DAPI led light cube (excitation: 357/44 nm; emission: 447/60 nm). Overall, C3H/APX-KD does not appear to have collapsed vessel walls. (C) Bar chart showing the targeted knockdown of C3H/APX (I1H6P1) resulted in substantially lower protein abundance (~60% lower abundance when compared to WT). 'ns' represent not significant and '*' represent significant differences (adj p-value < 0.05) among genotypes.

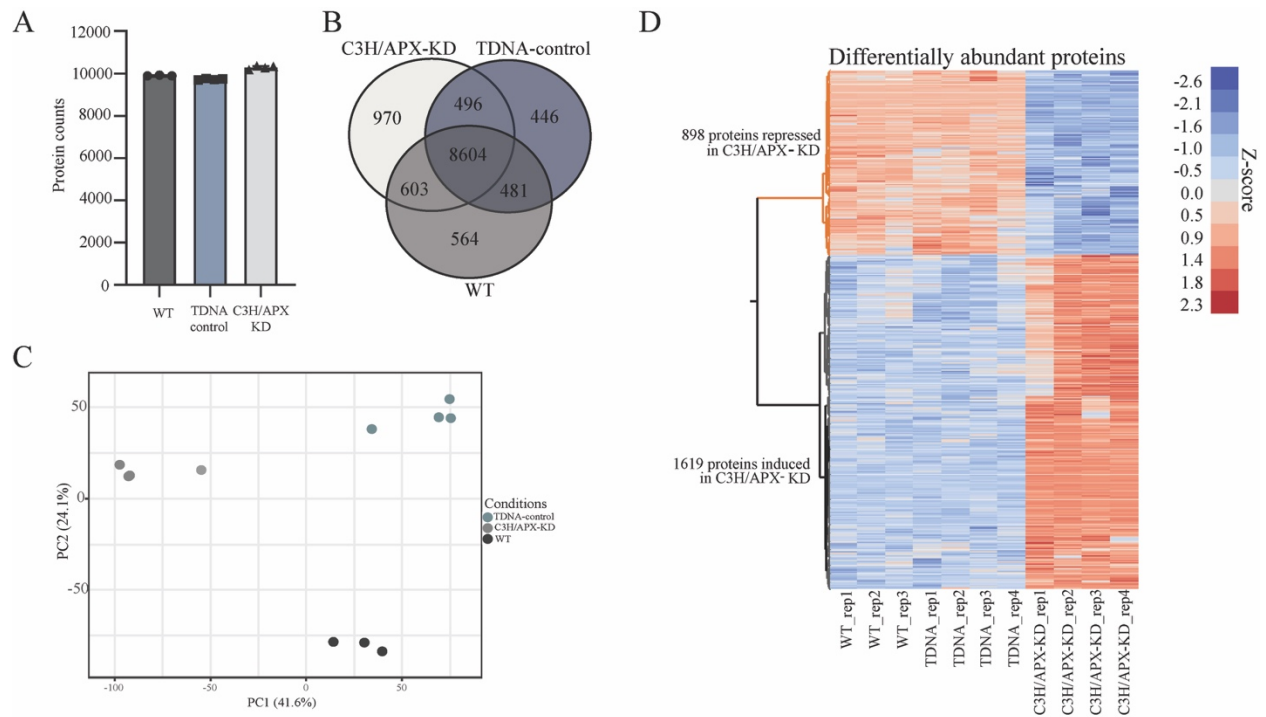


Figure 7.2: Global assessment of proteomics measurements.

(A) The number of proteins identified across experimental conditions. The number of proteins identified ranges from around 9,600 to 10,400. Each experimental condition has at least biological triplicates. (B) Venn diagram of the identified proteins in each experimental condition. Majority of the proteins were identified in all conditions with ~ 4-8% uniquely identified in individual treatment condition. (C) Principal component analysis (PCA) for the experimental samples. PCA of proteome data showed discrete grouping between the different experimental conditions. (D) Heatmap of proteins that are significantly changing in abundance between controls (WT and TDNA control) and the C3H/APX-KD.

different treatment lines (**Figure 7.2C**). To identify changes in protein abundance resulting due to knockdown of C3H/APX, ANOVA followed by Tukey HSD binary comparisons was performed. In total, 2,517 proteins (21.2% of the total identified proteome) were significantly different in abundance when comparing C3H/APX-KD lines to both controls. Among the proteins with altered abundances in C3H/APX-KD lines, 898 proteins had significantly lower abundances, and 1,619 proteins had significantly higher abundances in C3H/APX-KD when compared to both controls (**Figure 7.2D and Supplement Table 7.1**).

7.3.2 C3H/APX deficiency impacts enzymes associated with the phenylpropanoid biosynthesis pathway

As a bifunctional enzyme associated with phenylpropanoid metabolism, it remains to be determined how reduced C3H/APX function impacts the abundance of lignin biosynthesis-related proteins. Of the 87 potential candidate lignin biosynthetic proteins in *Brachypodium* (Barros et al., 2022), 62 were identified in the current study, and 14 were significantly altered in protein abundances as per our criteria (**Supplement Table 7.2**). The proteomics analysis showed that the reduction of C3H/APX abundance in stem tissues led to increased levels of cinnamoyl-CoA reductase (CCR; I1I7E4; Bradi3g36887), which catalyzes the NADPH-dependent conversions of *p*-coumaroyl-CoA, feruloyl-CoA, and sinapoyl-CoA into the corresponding aldehydes. Similarly, *p*-coumarate monolignol transferase (PMT; I1HM65; Bradi2g36910), a grass-specific acyltransferase enzyme that acylates monolignols with *p*-coumarate, also increased in abundance when compared to controls. Besides CCR and PMT, a small increase in abundance of cinnamic acid 4-hydroxylase (C4H1; I1HSV5; Bradi2G53470) and laccase-12 (LAC12; I1GXW0; Bradi1G37620) was observed in C3H/APX-KD as compared to controls. Apart from these, we observed 10 proteins with significantly lower abundances in C3H/APX-KD compared to controls. Among these reduced-abundance proteins were two laccase-7 proteins (I1HTF1; Bradi2G55050 and I1ITB5; Bradi4G39330), two CCR-like proteins (I1GX75; Bradi1g35730 and A0A0Q3H468; Bradi1g35736), two C3H/APX proteins (I1H6P1 & I1H6P2; Bradi1g65820), two ascorbate peroxidases APX5 (I1IWB9; Bradi5g03640 and I1IWC0; Bradi3g42340), aldehyde dehydrogenase (ALDH) known as reduced epidermal fluorescence REF1 (I1HP11; BRADI2g42360) and the early flavonoid pathway enzyme chalcone synthase (I1ILF3;

Bradi4g17230) (**Figure 7.3**). We also profiled the protein abundance of other peroxidases previously reported to be implicated in lignin polymerization (Cosio et al., 2017, Tokunaga et al., 2009, Fernandez-Perez et al., 2015, Hoffmann et al., 2020) including PRX17 (Hoffmann et al., 2020, Cosio et al., 2017), PRX47 (Tokunaga et al., 2009) and PRX52 (Fernandez-Perez et al., 2015) all with decreased abundance in C3H/APX-KD compared to controls (**Figure 7.3 and Supplement Table 7.2**).

Given that phenylpropanoid metabolism is tightly linked to the shikimate pathway, with shikimate acting as an upstream precursor of 4-coumaroyl-CoA as well as a substrate of hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase/ coumaroyl shikimate 3-hydroxylase (HCT/C3'H) mediated 3-hydroxylation (Qian et al., 2019, Hoffmann et al., 2004), we extended the focus onto how the enzymes of the shikimate pathway may be changing in the C3H/APX-KD line. Interestingly, the proteomics analysis identified the majority of the proteins involved in shikimate and aromatic amino acid biosynthesis pathways, with several proteins such as shikimate kinase, arogenate dehydratase, and arogenate dehydrogenase being substantially upregulated in C3H/APX-KD when compared to both controls (**Supplement figure 7.2**). We then shifted the focus to the flavonoid biosynthetic pathway which is also linked to the lignin biosynthesis pathway. Proteomics analysis showed that early flavonoid biosynthetic proteins like naringenin-chalcone synthase and flavanone 3-dioxygenase were significantly downregulated, whereas late flavonoid biosynthetic proteins such as anthocyanidin 3-*O*-glucosyltransferase were significantly upregulated in C3H/APX-KD compared to controls (**Supplement Table 7.2**).

7.3.3 C3H/APX knockdown altered abundance of transcriptional regulators involved in the phenylpropanoid biosynthesis pathway

Transcription factors tightly modulate the phenylpropanoid biosynthesis pathway (Zhao and Dixon, 2011, Muro-Villanueva et al., 2019, Bonawitz et al., 2014, Bonawitz et al., 2012, Zhang et al., 2020a). To assess how the knockdown of an enzyme involved in the phenylpropanoid biosynthesis pathway affects transcriptional regulation, we profiled the abundance of transcriptional factors across experimental conditions.

Figure 7.3: Impact of C3H/APX knockdown in lignin biosynthesis pathway.

(A) scheme of the lignin biosynthesis pathway is overlaid with significantly changing proteins and metabolites in C3H/APX-KD compared to WT. The heat blocks represent the log₂ fold change obtain from the binary comparison of C3H/APX-KD with WT. All proteins showed passed significance threshold. CCR isoforms shown are I1I7E4 (CCR1), I1GX75 (CCR-like3) and A0A0Q3H468 (CCR-like5). LAC isoforms shown are I1GXW0 (LAC12), I1HTF1 (LAC7-1) and I1ITB5 (LAC7-2). PRX isoforms shown are A0A2K2CRW2 (PRX17), A0A2K2CWW9 (PRX47), I1GYF3 (PRX52-1) and A0A0Q3KU45 (PRX52-2). Detailed information about these proteins is provided in Supplemental Table 2. The red or blue outlines in metabolites represent significantly up or down-regulated metabolites in C3H/APX-KD compared to WT. (B) Bar plots below the pathway show relative abundance of the significantly changing proteins in all experimental conditions. A flavonoid pathway marker, CHS were significant but are not shown in this pathway. PAL: L-phenylalanine ammonia-lyase, C4H: cinnamate 4-hydroxylase, COMT: caffeate/5-hydroxyferulate 3-*O*-methyltransferase, F5H: ferulate 5-hydroxylase/coniferaldehyde 5-hydroxylase, 4CL: 4-hydroxycinnamate:CoA ligase, HCT: 4-hydroxycinnamoyl CoA:shikimate/quinate hydroxycinnamoyltransferase, C3'H: 4-coumaroyl shikimate/quinate 3'-hydroxylase, CSE: caffeoyl shikimate esterase, CCoAOMT: caffeoyl CoA 3-*O*-methyltransferase, CCR: cinnamoyl CoA reductase, CAD: cinnamyl alcohol dehydrogenase, PMT: *p*-coumarate monolignol transferase, LAC: laccase, PRX: Peroxidase, REF: reduced epidermal fluorescence, PTAL: bifunctional L-phenylalanine/L-tyrosine ammonia-lyase, C3H/APX: 4-coumarate 3-hydroxylase, APX: Ascorbate peroxidase, CHS: chalcone synthase.

Metabolic pathway diagram illustrating the biosynthesis of lignin precursors from Tyrosine. The pathway involves several enzymes (C4H, PTAL, 4CL, HCT, C3H, CSE, COMT, F5H, CCR, REF1, CAD, PRX, LAC, PMT) and metabolites (Cinnamic acid, Phenylalanine, p-coumarate, p-coumaroyl CoA, p-coumaroyl shikimate, Caffeoyl shikimate, Caffeoyl-CoA, Ferulic acid, Feruloyl-CoA, 5H Ferulic acid, 5H Feruloyl CoA, Sinapic acid, Sinapoyl CoA, Sinapaldehyde, Sinapyl alcohol, Coniferyl alcohol, 5H Coniferyl alcohol, Monolignol p-coumarate, p-coumaraldehyde, p-coumaroyl alcohol, Coniferyl alcohol, 5H Coniferyl alcohol, Sinapyl alcohol, p-coumaroyl shikimate, Caffeoyl shikimate, Caffeoyl-CoA, Ferulic acid, Feruloyl-CoA, 5H Ferulic acid, 5H Feruloyl CoA, Sinapic acid, Sinapoyl CoA, Sinapaldehyde, Sinapyl alcohol, Coniferyl alcohol, 5H Coniferyl alcohol, Sinapyl alcohol, Monolignol p-coumarate, p-coumaraldehyde, p-coumaroyl alcohol, Coniferyl alcohol, 5H Coniferyl alcohol, Sinapyl alcohol, p-coumaroyl shikimate, Caffeoyl shikimate, Caffeoyl-CoA, Ferulic acid, Feruloyl-CoA, 5H Ferulic acid, 5H Feruloyl CoA, Sinapic acid, Sinapoyl CoA, Sinapaldehyde, Sinapyl alcohol, Coniferyl alcohol, 5H Coniferyl alcohol, Sinapyl alcohol, Monolignol p-coumarate, p-coumaraldehyde, p-coumaroyl alcohol, Coniferyl alcohol, 5H Coniferyl alcohol, Sinapyl alcohol).

Log₂ fold change C3H/APX-KD vs WT values for various metabolites:

- p-coumarate: 0.37 (upregulated)
- p-coumaroyl CoA: 1 (upregulated)
- p-coumaroyl shikimate: -0.86 (downregulated)
- p-coumaraldehyde: -0.91 (downregulated)
- p-coumaroyl alcohol: 0.47 (upregulated)
- Monolignol p-coumarate: -0.65 (downregulated)
- Coniferyl alcohol: 0.47 (upregulated)
- 5H Coniferyl alcohol: -0.65 (downregulated)
- Sinapyl alcohol: 0.47 (upregulated)
- 5H Sinapyl alcohol: -0.65 (downregulated)

Enzyme names and their corresponding Log₂ fold change values are provided in the diagram. The color scale indicates the Log₂ fold change, ranging from -10 (blue) to 10 (red).

Overall, we observed an increased abundance of transcriptional regulators such as Mediator complex, MYB, and WRKY (**Figure 7.4A and Supplement Table 7.3**). Proteomics analysis identified nine significantly changing subunits of the Mediator complex (including MED9, MED26C, MED23, and MED36B). All Mediator proteins identified were upregulated in the C3H/APX-KD line compared to controls, with 6 of them $>1 \log_2$ fold higher in abundance than in both controls (**Figure 7.4B and Supplement Table 7.3**). Similarly, our analysis showed that 23 MYB transcription factors were substantially upregulated, with at least a $2.5 \log_2$ fold change in the C3H/APX-KD line compared to both controls (**Figure 7.4B and Supplement Table 7.3**).

A recent study identified an importin-beta-like protein that mediates the transport of the MYB4 transcriptional repressor from the cytosol into the nucleus (Panda et al., 2020, Zhao et al., 2007). In Arabidopsis, importin-beta-like protein and MYB4 have been shown to mediate lignin modification-induced dwarfism, and suppression of these genes alleviates the dwarf phenotype (Panda et al., 2020). Interrogation of Brachypodium orthologs revealed 3 importin-beta-like proteins upregulated by at least $1 \log_2$ fold in C3H/APX-KD when compared to controls (**Supplement Table 7.3**). Likewise, MYB4 (A0A0Q3J6J9; Bradi2g40620) was $7.5 \log_2$ fold higher in abundance in C3H/APX-KD when compared to controls. Further experimental validation is needed to see if the MYB4 in Brachypodium has similar role as it has in Arabidopsis. In addition to these, 4 WRKY transcription factors were substantially upregulated with at least $5.2 \log_2$ fold in the C3H/APX-KD line when compared to controls (**Figure 7.4B**).

7.3.4 C3H/APX suppression alters the abundance of the proteins associated with phytohormone signaling and senescence

Brachypodium C3H/APX-KD lines exhibit a dwarf and stay green (delayed senescence) phenotype (**Figure 7.1A**). Phytohormones have been known to impact the progression and regulation of the senescence process (Guo, 2021, Joshi et al., 2019). Additionally, alteration in phytohormones can also result in reduced-growth and dwarf phenotypes (Thomas and Sun, 2004, Fujioka and Yokota, 2003) and previous studies indicate modifying lignin can alter phytohormone levels (Gallego-Giraldo et al., 2011, Besseau et al., 2007). To understand how phytohormones are regulated in C3H/APX-KD Brachypodium, proteins associated with phytohormone biosynthesis

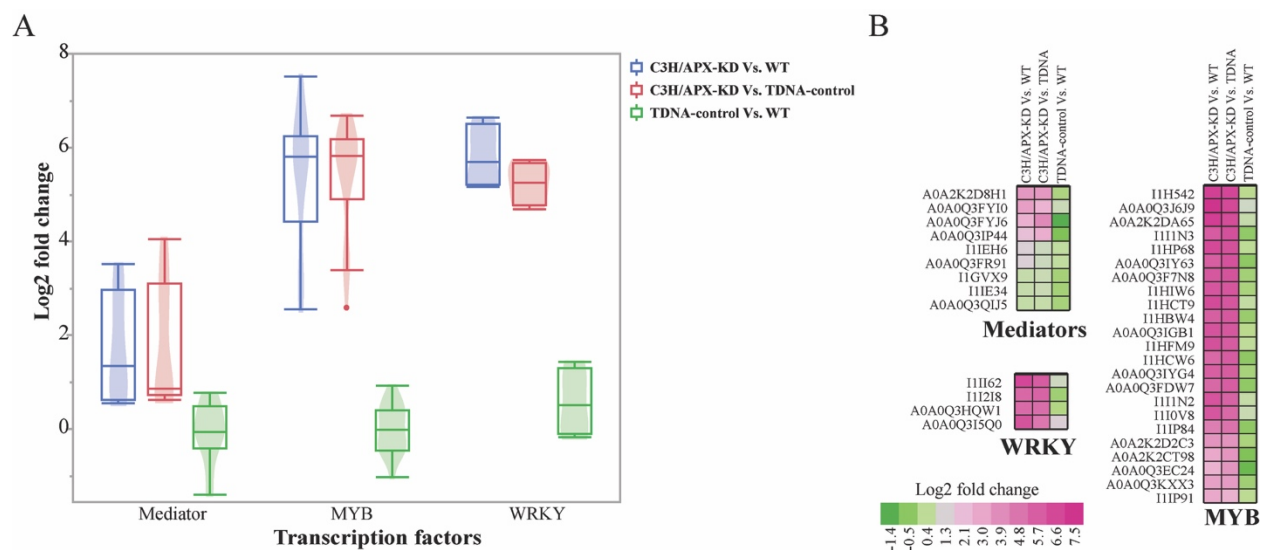


Figure 7.4: Protein abundance profile of transcriptional regulators.

(A) Boxplot shows the overall abundance change of three type of transcription factors. The violin plot inside the boxplot shows the distribution of data. (B) Cell plot shows the log2 fold change for mediators, MYB, WRKY. All the proteins represented passes significant threshold (adj p-value < 0.05). Color gradients represent the log₂ fold change.

and perception were interrogated. Overall, significantly changing abundances were observed for proteins related to all eight phytohormones (**Supplement Figure 7.3A and Supplement Table 7.4**). Among these, proteins associated with abscisic acid (ABA) and auxin were the most represented and showed increased abundances in the C3H/APX-KD line (**Supplement Figure 7.3A and Supplement Table 7.4**).

For ABA, we observed a significant increase in ABA regulating proteins such as ABA receptor PYL8, ABA-regulated RNA-binding protein (APR1), THO complex subunit 6 (THOC6) in the C3H/APX-KD line compared to both controls (**Supplement Table 7.4**). Meanwhile, a significant decrease in ABA biosynthesis proteins such as Zeaxanthin epoxidase (ZEP), HVA22-like protein, and serine/threonine-protein kinase SRK2 were observed in the C3H/APX-KD line compared to controls (**Supplement Table 7.4**). SRK2 plays a key role in abscisic acid (ABA) signaling and helps to regulate seed germination and seedling growth (Kim et al., 2012). It interacts with bZIP transcription factors in rice which were also found to be reduced in C3H/APX-KD by 7.17 log₂ fold of WT.

For auxin, our analysis identified substantial decreases in proteins involved in auxin response such as auxin-induced in root cultures (AIR) protein in the C3H/APX-KD line compared to both controls (**Supplement Table 7.4**). The WAT1 related protein (A0A2K2DKI4; Bradi1g21860) was one of the most significantly decreased proteins (~ 9.94 log₂ fold lower than WT) in the C3H/APX-KD line (**Supplement Table 7.4**). WAT1 is a transmembrane protein that is involved in growth and secondary cell wall development in Arabidopsis (Ranocha et al., 2010, Ranocha et al., 2013). It has also been reported to be involved in auxin transport and homeostasis (Ranocha et al., 2013) and may suggest why the Brachypodium C3H/APX-KD line has reduced growth and development. Proteins negatively regulated by auxins such as dormancy-associated protein homologs (DRMHs) were substantially increased in abundance (**Supplement Table 7.4**) (Rae, 2013).

Besides phytohormones, the senescence process is also highly coordinated through senescence-associated genes (SAGs). Proteomics analysis identified two significantly changing SAGs which were both substantially downregulated in C3H/APX-KD compared to controls (**Supplement Figure 7.3B**). Meanwhile, senescence-associated e3 ubiquitin ligase 1 (I1IFT9; Bradi3g60810)

which prevents premature senescence was significantly upregulated in C3H/APX-KD compared to controls (**Supplement Figure 7.3B**).

7.3.5 C3H/APX deficient *Brachypodium* exhibit disrupted redox homeostasis, photosynthesis, and DNA repair

C3H/APX encodes a cytosolic ascorbate peroxidase and therefore it is anticipated that its reduced expression will impact biological processes influenced by redox homeostasis (Agrawal et al., 2003, Davletova et al., 2005). For the proteins having reduced abundance in C3H/APX-KD lines, a gene ontology (GO) enrichment test showed a significant overrepresentation of GO terms related to ROS metabolic processes, cellular oxidant detoxification and hydrogen peroxide catabolic processes (**Supplement Figure 7.4A, Supplement Table 7.5**). Among proteins related to oxidative stress responses, peroxidase, sulfiredoxin, ABC1-like kinase, class I heat shock protein, and peroxiredoxin were substantially reduced in abundance in C3H/APX-KD compared to controls (**Supplement Table 7.5**).

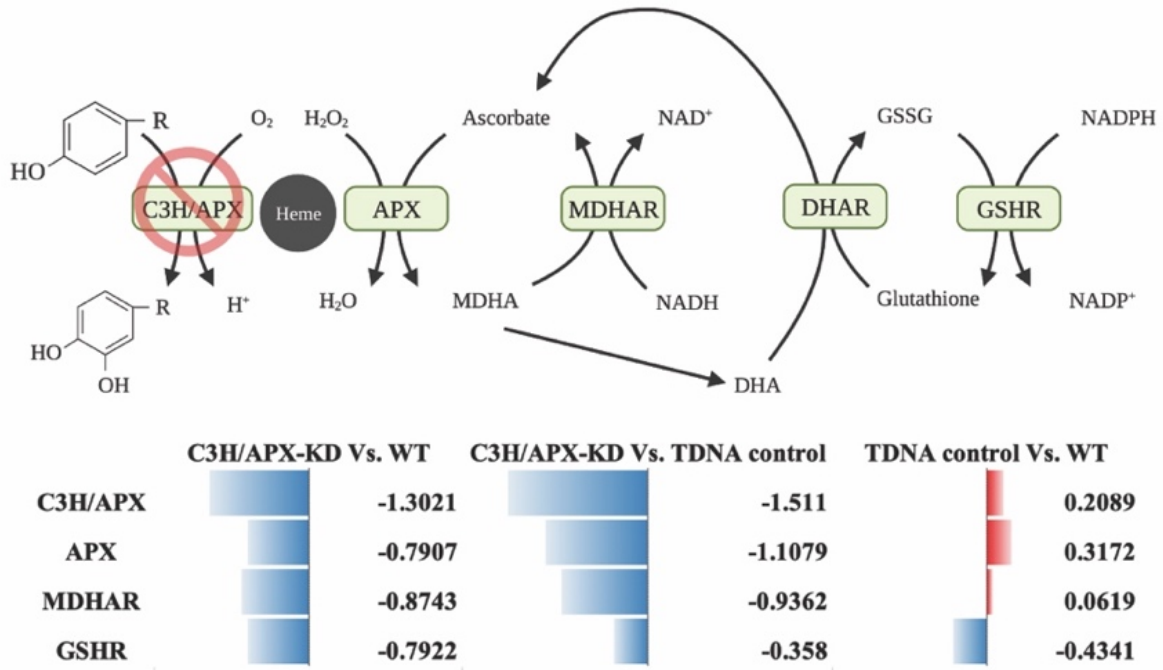
Cytosolic ascorbate peroxidase is also involved in the recycling of ascorbate (Asc), the most abundant water-soluble antioxidant, via the ascorbate-glutathione cycle (**Figure 7.5A**). Through this cycle, cellular Asc content and redox state are maintained, which is critical under conditions with increased reactive oxygen species (ROS) production (Gallie, 2013). When compared to controls, proteome expression changes in the C3H/APX-KD line showed a significant reduction in protein abundances associated with ascorbate recycling including monodehydroascorbate reductase (MDHAR) and glutathione reductase (GSHR) (**Figure 7.5A**).

Previous studies have demonstrated that reduced cytosolic ascorbate peroxidase function leads to disrupted oxidative protection of chloroplasts that results in lower photosynthetic rates and slower growth under normal growth conditions (Miller et al., 2007, Pnueli et al., 2003, Bonifacio et al., 2011, Ribeiro et al., 2012). Corroborating previous studies, our proteome-wide functional analysis showed that biological processes related to photosynthesis, including light harvesting and light reaction functions, were significantly enriched among the proteins repressed in the C3H/APX-KD lines when compared to controls (**Supplement Figure 7.4 and Supplement Table 7.5**). In

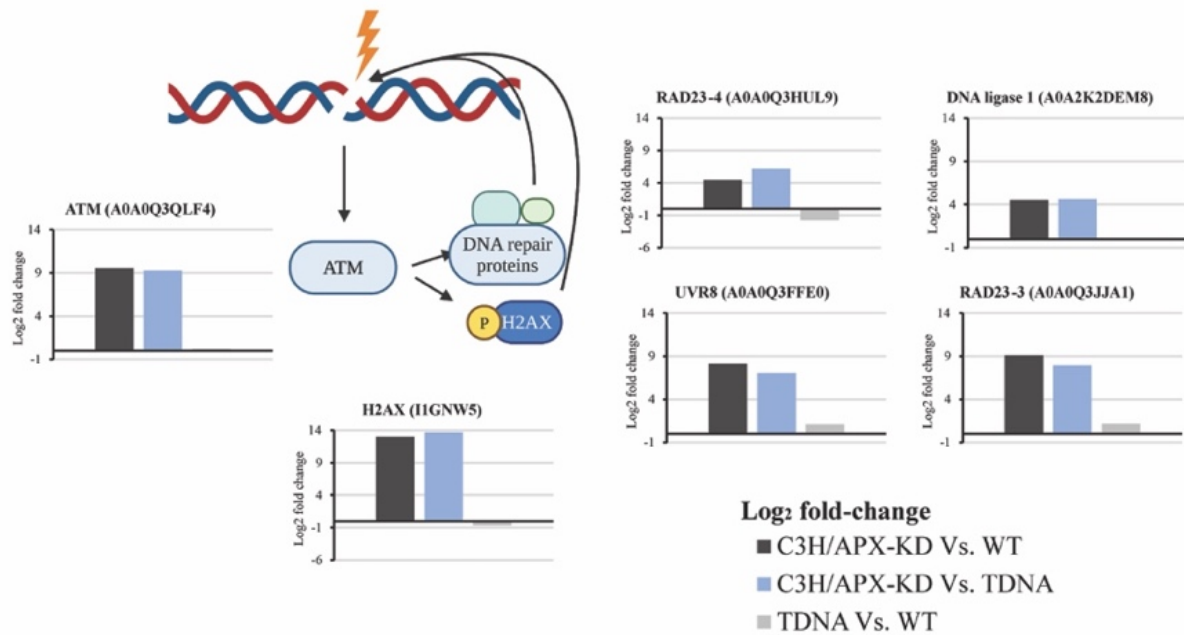
Figure 7.5: C3H/APX-deficient *Brachypodium* exhibit altered redox homeostasis and DNA damage response.

(A) Schematic overview of ascorbate recycling pathway involved in plant redox homeostasis. Multiple enzymes involved in ascorbate recycling were significantly decreased in abundance in C3H/APX-KD compared to controls. Numbers next to data bars are the $\log_2$ fold change from binary comparison. DHAR was identified but did not pass significant threshold (adj p-value<0.05). APX: L-ascorbate peroxidase; MDHAR: monodehydroascorbate reductase; DHAR: dehydroascorbate reductase; GSHR: glutathione reductase. B) Simplified schematic diagram showing the plant DNA damage response upon cellular stress overlaid with bar plot showing the $\log_2$ fold changes in each binary comparison for individual proteins involve in the process. All the proteins represented passed the significance threshold of adj p-value<0.05 and number of replicates used were at least 3. ATM: Ataxia telangiectasia mutated; H2AX: Histone H2A; RAD: DNA repair protein; UVR: ultraviolet-B receptor.

A



B



particular, the relative abundances of chloroplast proteins such as NAD(P)H-quinone oxidoreductase, serine/threonine-protein kinase STN8, and chlorophyll a-b binding protein were significantly lower in the C3H/APX-KD line compared to controls (**Supplement Table 7.5**).

Among the proteins with significantly increased abundances in the *Brachypodium* C3H/APX-KD line when compared to controls, the PANTHER protein class overrepresentation test identified damaged DNA-binding proteins to be overrepresented (**Supplement Figure 7.4B**). Specifically, substantial upregulation of a protein ataxia-telangiectasia mutated (ATM) homolog (A0A0Q3QLF4; Bradi2g00610, log₂ fold >9.56) was identified in the C3H/APX-KD line compared to controls (**Figure 7.5B**). ATM is involved in regulating DNA damage responses; it phosphorylates the histone variant H2AX at double-strand breaks (Bakkenist and Kastan, 2003, Podhorecka et al., 2010). Although the phosphorylation profile was not measured in this study, H2AX (I1GNW5; Bradi1g10390) was the most abundant differential protein between the C3H/APX-KD line and WT with a log₂ fold change of 13.0 (**Figure 7.5B**). ATM is also involved in regulating other proteins involved in DNA damage responses, such as the DNA repair protein RAD, RCC1 domain-containing protein, and DNA ligase 1, which were all significantly upregulated in C3H/APX-KD compared to controls (**Figure 7.5B**). Moreover, we also observed the concomitant upregulation of organelle DNA repair proteins (I1HS48; Bradi2g51390 and A0A0Q3JCJ2; Bradi2g51390) that repair organelle-specific DNA double-stranded breaks by microhomology-mediated end-joining (Garcia-Medel et al., 2019).

7.3.6 C3H/APX deficiency impacts cellular kinases implicated in plant defense

Non-receptor serine/threonine kinases are intracellular cytoplasmic proteins that relay intracellular signals (Siveen et al., 2018) and were found to be differentially regulated in the C3H/APX-KD line compared to controls (**Supplement Figure 7.4B**). Significantly changing proteins mapping to this protein class were extracted and qualitatively analyzed based on ranking fold changes between the C3H/APX-KD line and WT (**Supplement Table 7.5**). In general, most upregulated non-receptor kinases are implicated in defense signaling such as EDR (Hiruma et al., 2011), and cellular development and expansion like M3KE1 (Chaiwongsar et al., 2012) or NEK5 (Sakai et al., 2008) (**Supplement Table 7.5**). On the other hand, down regulated non-receptor kinases have been

associated with regulation of stress tolerance such as SRK2 (Umezawa et al., 2004), phosphorylation of acidic proteins such as casein, calcium signal transducer like CDPK3 (Mori et al., 2006) (**Supplement Table 7.5**).

Several studies have shown that defense responses frequently come at the cost of reduced growth in plants (Karasov et al., 2017, Stamp, 2003, Huot et al., 2014, Ha et al., 2021). Because the C3H/APX-KD line displayed abnormal growth phenotype and the proteome-wide analysis shows defense signaling kinases to be differentially regulated, proteins that have been associated with disease resistance were interrogated qualitatively using keywords. Overall, several proteins associated with disease resistance (25 proteins with $\log_2$ fold >1 vs 10 proteins with $\log_2$ fold < -1) were differentially abundant between the C3H/APX-KD line when compared to controls (**Figure 7.6 and Supplement Table 7.5**). In addition, pathogen-associated molecular patterns-induced protein (A0A0Q3K2Z8; Bradi2g18237) was increased by 3.56 $\log_2$ fold in the C3H/AP-KD compared to WT.

7.3.7 Transcriptomics, metabolomics, and ROS levels support proteomics findings

To obtain information on the transcript level changes in C3H/APX-KD, RNA-seq was performed on C3H/APX-KD and WT *Brachypodium*. Unfortunately, one of the replicates from WT condition was lost during sample preparation which limited our statistical power. Nonetheless, transcript level analysis identified 470 genes that were differentially expressed. Among these DEGs, 293 genes were significantly upregulated, and 177 genes were significantly downregulated in C3H/APX-KD compared to WT (**Supplement Figure 7.5A**). When comparing DEGs to significantly changing proteins, we identified 54 significantly changing proteins whose transcript levels were also differentially expressed. Correlation analyses showed a positive correlation ($R^2 = 0.61$) between transcript and protein levels (**Supplement Figure 7.5B**). Overall, the DEGs were enriched in biological processes associated with regulation of transcription and cellular responses to stress (**Supplement Figure 7.5C**). At the individual gene level, we observed the increased expression of pathogen-associated molecular patterns-induced gene (Bradi2g18237, $\log_2$ FC=2.5) in C3H/APX-KD compared to WT, comparable to the results observed at the proteome levels. Similarly, we found decreased expression levels of the lignin biosynthetic gene

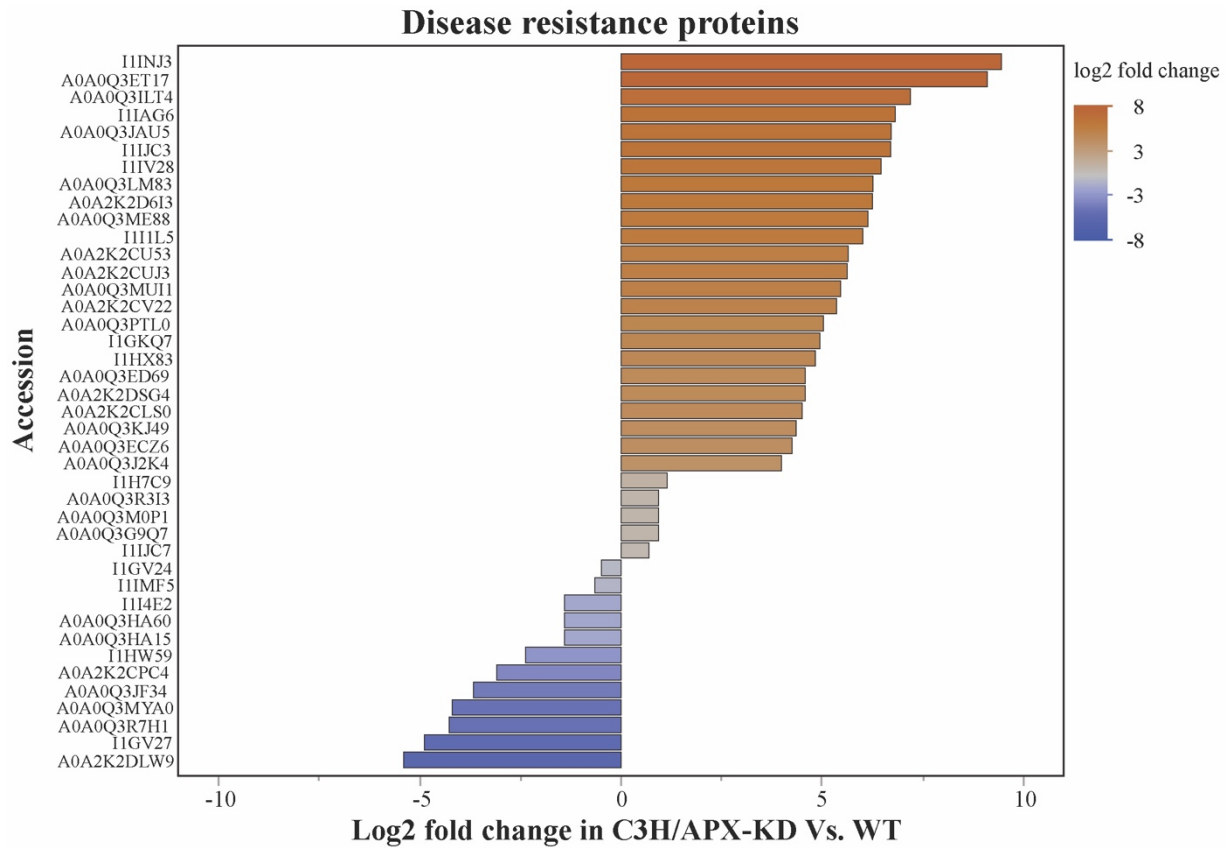


Figure 7.6: C3H/APX-deficient *Brachypodium* exhibit elevated levels of disease resistance proteins.

Bar plot showing the log₂ fold change of disease resistance protein in C3H/APX-KD compared to WT. All the protein represented passes the ANOVA significance followed by Tukey HSD binary comparisons significance (adj p<0.05). Blue bars represent the significantly downregulated proteins in C3H/APX-KD compared to controls and orange bars represent the significantly upregulated proteins in C3H/APX-KD compared to controls. Color gradients represent the log₂ fold gradient. Further information about these proteins are provided in Supplement Table 5.

CCR2 (Bradi1g35736, log₂FC= -1.3), senescence related gene SAG12 (Bradi4g07130, log₂FC= -1.3), and peroxidases such as peroxidase 12 (Bradi5g27150, log₂FC= -2.6) and peroxidase 52 (Bradi1g41115, log₂FC= -1.7) (**Supplement Table 7.6**). Like *Brachypodium*, *Arabidopsis c3h/apx* mutants exhibited similar molecular changes with increased expression of genes related to oxidative response, defense response and decreased expression of gene related to phytohormone response (**Supplement Table 7.6**). In general, the transcriptome-level characterization of the C3H/APX-KD line was correlated and consistent with the proteome findings.

Metabolomics was performed to identify changes in metabolites because of C3H/APX downregulation. Overall, 91 metabolites were identified, of which 12 metabolites were uniquely upregulated and 8 metabolites were uniquely downregulated in C3H/APX-KD compared to control plants (**Supplement Table 7.7**). Metabolomics measurement showed significant upregulation of *p*-coumarate, the substrate of C3H/APX, and downregulated levels of G-lignin and syringyl lignan glycoside in C3H/APX-KDs compared to WT and T-DNA controls (**Figure 7.3 and Supplement Table 7.7**). The increased levels of C4H protein and 4-coumarate (product of the C4H reaction) in the C3H/APX-KD lines suggest that at least part of the 4-coumarate levels may derive from higher C4H activity rather than from only downregulation of the target C3H reaction. In addition to metabolites specific to lignin biosynthesis, compounds involved in ROS signaling and abiotic stress response such as linoleic acid and aconitic acid were significantly upregulated in the C3H/APX-KD lines (**Supplement Table 7.7**). In contrast, the levels of a conjugate of ascorbic acid and gentisic acid-5-*O*-glucoside potentially involved in alleviating oxidative stress in plants were substantially decreased in C3H/APX-KD compared to WT (**Supplement Table 7.7**).

To validate a potential misregulation of the redox state supported by the multiple omic datasets, we measured the levels of ROS and hydrogen peroxide in C3H/APX-KD and control plants (**Figure 7.7**). To avoid free radical damage in the stored plant material, plants were grown again under controlled conditions for 4-weeks (**Figure 7.7A**), fumigated with the herbicide paraquat and exposed to high-light to induce oxidative stress, and immediately tested for ROS and H₂O₂ accumulation. Fluorescence imaging detected higher levels of ROS in the stems of C3H/APX-KDs compared to controls (**Figure 7.7B-C**). Moreover, the total concentration of H₂O₂ in whole C3H/APX-KDs plants detected by fluorescence spectrophotometry was significantly elevated

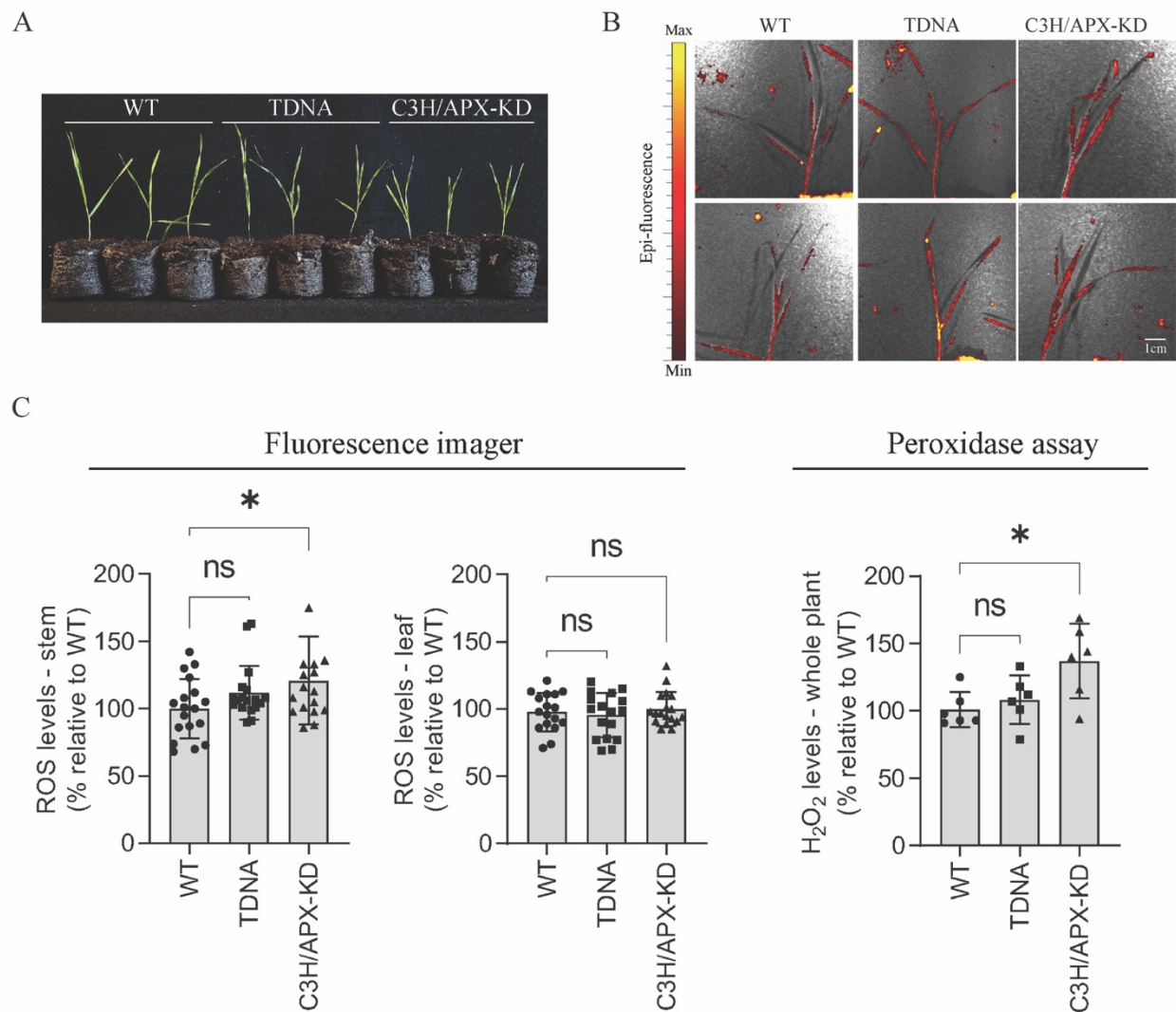


Figure 7.7: Reactive oxygen species (ROS) and H₂O₂ levels in *Brachypodium* C3H/APX-KD lines.

(A) C3H/APX-KD lines and controls grown in Jiffy peat pellets for 4-weeks. (B) Imaging of ROS in whole plants fumigated with paraquat and stressed under high-light conditions for 30 min each. (C) Measurements of ROS levels in the stem and leaf tissues and total H₂O₂ accumulation in *Brachypodium* C3H/APX-KD lines and control plants. Student's t-test, error bar indicates SD, *P < 0.05, n > 6. Scale bar represents 1 cm.

when compared to both TDNA and WT controls (29% and 37% increase, respectively) (**Figure 7.7C**).

7.4 Discussion

Although lignin provides essential mechanical support for plant cell walls, it is a major contributor to biomass recalcitrance of biofuel crops (Boerjan et al., 2003, Velvizhi et al., 2022). Thus, there has been a special interest in modifying the content and/or composition of this important plant biopolymer to facilitate bioprocessing. However, attempts to modify lignin result in pleiotropic effects with several unfavorable traits, including reduced growth. Despite several hypotheses, the precise molecular mechanisms underlying the negative growth of lignin-modified plant remains largely unanswered. Moreover, most of the studies have been primarily focused on dicot plants, and lignin metabolism in grasses, which differs substantially from that in dicots, is understudied.

Recent research in grass species has identified a monolignol pathway enzyme, C3H/APX, which provides a parallel route to the shikimate shunt pathway (Barros et al., 2019). As in many other lignin-modified plants, knockdown of C3H/APX reduces plants lignin content but results in negative growth phenotypes (Barros et al., 2019). In lignin knockdown lines, collapse of xylem vessels which impair water and solute conductance has been hypothesized as one of the likely causes of perturbed growth; however, C3H/APX-KD knockdown plants do not exhibit abnormal vessel walls (**Figure 7.1**). The precise molecular alterations underlying the negative growth effect observed due to C3H/APX knockdown remain to be explored. In this study, we performed a comparative proteomics analysis in C3H/APX-KD, TDNA control and WT *Brachypodium* stem tissue to investigate underlying mechanism of C3H/APX-KD associated negative growth and gain a deeper understanding of how grasses alter levels of lignin pathway enzymes and other biological processes following C3H/APX knockdown.

Monolignols are synthesized in the cytoplasm and transported to the cell wall, where they are oxidized by laccases and/or peroxidases prior to their incorporation into the polymer (Bonawitz and Chapple, 2010, Dixon and Barros, 2019). In *Arabidopsis*, mutation of laccase proteins has been shown to reduce the lignin content (Liang et al., 2006, Zhao et al., 2013). The substantial

decrease in abundance of laccase proteins observed from proteomics measurement in C3H/APX-KD compared to controls suggests that the C3H/APX-KD may have reduced ability to oxidize monolignols, which might contribute to their low total lignin. Besides lignin quantity, lignin composition can also be controlled at the polymerization level (Zhuo et al., 2022, Tobimatsu and Schuetz, 2019). Our proteomics and transcriptomics analyses identified several peroxidases co-regulated with C3H/APX. One of the significantly downregulated peroxidases, PRX52, is a basic peroxidase that has been linked to lignification. A previous study in *Arabidopsis* has shown that suppression of AtPRX52 causes a reduction in S-lignin content (Fernandez-Perez et al., 2015). Our results suggest that the markedly reduced levels of several lignin polymerizing enzymes, and particularly PRX52, could be associated with the reduction of S-lignin observed in the C3H/APX-KD lines. CCR proteins are involved in the initial step of monolignol biosynthesis (Lauvergeat et al., 2001) and catalyze the NADPH-dependent conversions of *p*-coumaroyl-CoA, feruloyl-CoA and sinapoyl-CoA into the corresponding aldehydes (Davin et al., 2008, Vanholme et al., 2010). Previous work on tobacco has shown that CCR downregulation reduce plants' lignin content and increases the S/G ratio (Piquemal, 1998, Jones et al., 2001). Our proteomic analysis identified decreased abundance of two CCR-like proteins with unknown roles, and increased abundance of one CCR homolog of *Panicum virgatum* CCR1 with substrate preference for feruloyl CoA (Escamilla-Trevino et al., 2010, Zhou et al., 2010) in the C3H/APX-KD lines, which suggests an efficient conversion of feruloyl-CoA to coniferaldehyde for lignin production in these plants. Coniferaldehyde can be converted to coniferyl alcohol (G-lignin precursor) directly via CAD. Therefore, increased abundance of CCR1 could also be partially responsible for reduced S/G ratios seen in the C3H/APX-KD lines. Lastly, proteomics measurement identified PMT, a grass-specific acyltransferase enzyme, to be increased in abundance in C3H/APX-KD plants compared to controls. This protein acylates monolignols with *p*-coumarate prior to their incorporation into lignin (Withers et al., 2012, Petrik et al., 2014). The monolignol conjugates produced by PMT do not enter the radical coupling (polymerization) reactions of lignification (Withers et al., 2012). Brachypodium plants overexpressing PMT have shown to contain less lignin likely due to excessive BdPMT activity drawing *p*-coumaroyl-CoA away from monolignol biosynthesis, thereby lowering the amounts of monolignols available for polymerization (Petrik et al., 2014). Therefore, upregulation of PMT seen in C3H/APX-KD plants compared to controls could have

contributed to the lower level of lignin observed in the C3H/APX-KD lines. Changes in other phenylpropanoids, particularly flavonoids, can impact the cellular oxidative stress, auxin movement and plant development (Agati et al., 2012, Brunetti et al., 2013, Jacobs and Rubery, 1988, Taylor and Grotewold, 2005). C3H/APX-KD has been shown to alter the protein abundance of flavonoid biosynthesis pathway which could have a broader impact on plant development and fitness.

Mediator is a large, multi-subunit conserved transcriptional coregulatory complex that is increasingly recognized as an essential element of both basal and regulated eukaryotic transcription (Kornberg, 2007, Kim et al., 1994). In plants, mediators are involved in regulating complex biological processes, including pathogen resistance (Kidd et al., 2009, Wathugala et al., 2012), reproductive development (Imura et al., 2012), and phytochrome signaling (Cerdan and Chory, 2003). Recent studies suggest that Mediator also plays a key role in regulation of phenylpropanoid homeostasis in Arabidopsis (Bonawitz et al., 2012, Bonawitz et al., 2014). In Arabidopsis, the *p*-coumaroyl shikimate 3'-hydroxylase (C3'H) mutation associated lignin deficiency and negative growth phenotype can be overcome by the disruption of the Mediator complex subunits MED5a and MED5b (Muro-Villanueva et al., 2019, Bonawitz et al., 2014). Although MED5a and MED5b were not identified in our dataset, our proteomics analysis identified nine other significantly upregulated subunits of the Mediator complex in C3H/APX-KD compared to controls. In a similar vein, transcription factor families like MYB and WRKY also play crucial roles in regulating various cellular processes including phenylpropanoid synthesis (Zhang et al., 2020a, Muro-Villanueva et al., 2019, Yang et al., 2017). A recent study identified an importin-beta-like protein that mediates the transport of the MYB4 transcriptional repressor from the cytosol into the nucleus (Panda et al., 2020, Zhao et al., 2007). In Arabidopsis, importin-beta-like protein and MYB4 have been shown to mediate lignin modification-induced dwarfism, and suppression of these genes alleviates the dwarf phenotype in C3'H mutant but not in CCR1 mutant (Panda et al., 2020). Like Mediator subunits, our proteomics analysis identified several MYB and WRKY transcription factors including MYB4 and importin-beta-like protein, which were all significantly upregulated in C3H/APX-KD compared to controls. Overall, the proteomics analysis of the low lignin *Brachypodium* C3H/APX-KD indicates that there are complex and subtle

transcriptional control mechanisms linking phenylpropanoid biosynthesis, growth and development, and pathogen resistance. Moreover, some of the transcriptional regulators identified in this study can be targets for future studies to rescue the dwarf and non-viable seed phenotype observed in C3H/APX-KD plants.

Phytohormones are biochemical molecules that are known to impact various biological processes in plants including the progression and regulation of the senescence process, and plant growth and development (Guo, 2021, Joshi et al., 2019, Thomas and Sun, 2004, Fujioka and Yokota, 2003). Phytohormones can act either close to or remote from their sites of synthesis to regulate responses to environmental stimuli (Fahad et al., 2015). As a response to environmental stress, they can reduce plant growth to focus its resources on withstanding the stress (Fahad et al., 2015). Phytohormones like ethylene, jasmonic acid (JA), salicylic acid (SA) and abscisic acid (ABA) are known to promote leaf senescence; in contrast, cytokinins (CKs), gibberellic acid (GA), and auxin are known to delay leaf senescence (Guo, 2021, Zhang et al., 2013, Li et al., 2013, Lim et al., 2007, Gan and Amasino, 1997, Gan and Amasino, 1995). Our proteomics results identified significantly changing proteins involved in biosynthesis and/or perception of all major phytohormones. Therefore, the altered hormonal balance might also contribute towards the negative growth phenotype observed with C3H/APX-KD. Besides phytohormones, the senescence process is also highly coordinated through senescence-associated genes (SAGs). Proteomics and transcriptomics analyses identified SAG proteins/gene which were highly downregulated in C3H/APX-KD compared to controls, and senescence-associated ϵ 3 ubiquitin, which prevents premature senescence, as significantly upregulated in C3H/APX-KD compared to controls. The delayed senescence observed in C3H/APX-KD might therefore be influenced by the SAGs in addition to phytohormone abundance changes.

C3H/APX is a bifunctional cytosolic ascorbate peroxidase that is known to play crucial roles in regulating the level of cytosolic peroxides via ascorbate-glutathione cycle (Davletova et al., 2005). Through this process, cellular ascorbate content and redox state are maintained (Gallie, 2013). Thus, it is plausible that the deficiency of this peroxidase would lead to disruption in cellular redox homeostasis. Proteome results showed the decreased abundance of several proteins involved in the ascorbate-glutathione cycle, supporting the potential disruption in redox balance. In a similar

fashion, the proteome results also showed substantial reduction of photosynthesis related proteins such as NAD(P)H-quinone oxidoreductase, serine/threonine-protein kinase STN8, and chlorophyll a-b binding protein in C3H/APX-KD compared to controls. NAD(P)H-quinone oxidoreductase is involved in shuttling electrons from NAD(P)H:plastoquinone to quinones in the photosynthetic electron transport chain (Melo et al., 2004). Serine/threonine-protein kinase STN8 is a light-dependent kinase that specifically phosphorylates N-terminal threonine residues in components of the core antenna complex of photosystem II like psbA/D1, psbD/D2, psbC/CP43 and psbH (Vainonen et al., 2005). Chlorophyll a-b binding protein functions as a light receptor to capture and deliver excitation energy to photosystems (Allen, 1992, Aro and Ohad, 2003). Under normal photosynthetic conditions, cells produce a steady rate of reactive oxygen species which are usually scavenged to prevent cellular oxidative damage (Pnueli et al., 2003). In Arabidopsis, the absence of the cytosolic C3H/APX was shown to collapse the entire chloroplastic H₂O₂-scavenging system (Davletova et al., 2005). Thus, it is reasonable that the C3H/APX knockdown of *Brachypodium* with disrupted redox homeostasis has damaged organelle H₂O₂-scavenging systems, which could lead to compromised photosynthetic ability. Moreover, proteomics analysis also showed decreased abundance of stromal and thylakoid membrane bound APX in C3H/APX-KD compared to controls, suggesting reduced ascorbate recycling in the chloroplast that would also compromise plant's photosynthetic ability. Since the photosynthesis and growth are positively correlated (Kirschbaum, 2011), the potentially compromised photosynthetic ability of *Brachypodium* C3H/APX-KD might be one of the causes of the negative growth phenotype observed.

Exposure of DNA to reactive species can cause strand breakage, nucleic acid-protein crosslinking, and nucleotide base modifications (Marnett, 2000, Berg et al., 2004). The accumulation of DNA damage could lead to undesirable effects on the growth and yield of plants by blocking critical cellular processes such as transcription or replication and causing cell death (Kimura and Sakaguchi, 2006). Plant cells can activate DNA damage responses to repair damaged DNA; however, this process also contributes to reduction in plant growth (Nisa et al., 2019). Our proteomics results identified a higher abundance of proteins involved in DNA damage response. This implies that C3H/APX-KD has elevated oxidative DNA damage contributing to the negative growth phenotype observed in the C3H/APX-KD knockdown line. Similarly, ROS regulates plant

growth through phytohormone signaling (Considine and Foyer, 2014). They can also act as signaling molecules that are perceived by various cellular kinases to trigger the downstream defense signaling pathways (Jalmi and Sinha, 2015). In plants, trade-offs between growth and defense exist and the plant's optimal fitness depends on growth-defense balance (Ha et al., 2021). Proteome results suggest that the C3H/APX-KD plants, while having reduced growth, has a higher abundance of disease resistance proteins regulated via intracellular kinases which could potentially be activated by ROS.

Proteomics result from C3H/APX-KD *Brachypodium* is further supported by metabolomic measurements of C3H/APX-KD *Brachypodium* (Barros et al., 2019). In particular, linoleic acid was upregulated by more than 2-fold in C3H/APX-KD compared to WT (**Supplement Table 7.7**). Linoleic acid has been shown to regulate the expression of genes involved in the abiotic stress response, particularly those mediated by ROS signaling (Mata-Perez et al., 2015). Aconitic acid is another metabolite that was significantly upregulated by more than 2-fold in C3H/APX-KD compared to WT (**Supplement Table 7.7**). This metabolite has been previously demonstrated to strongly inhibit the soybean (*Glycine max*) plant's growth and its photosynthetic ability while enhancing the cellular H₂O₂ content (Bortolo et al., 2018). Metabolomics also detected a substantial decrease of an ascorbate conjugate metabolite which could be related to ascorbate recycling (**Supplement Table 7.7**). Moreover, gentisic acid-5-*O*-glucoside was also substantially decreased. This metabolite and its derivatives are known to act as antioxidants and have been associated with plant defense and salicylic acid metabolism (Joshi et al., 2012, Boreham and Martin, 1969, Todd et al., 1971, Campbell et al., 1984).

Overall, our findings suggest that C3H/APX is involved in specific crosstalk between lignin biosynthesis and abiotic stress through ROS metabolism, and thus knockdown of C3H/APX results in a negative growth phenotype through the growth-defense trade-offs (**Figure 7.8**). The deficiency of C3H/APX likely results in the ROS imbalance, which elevates cellular oxidative stress, thus disrupting the cellular and organelle redox balance. Additionally, the redox imbalance can directly elevate DNA damage, compromise photosynthesis and trigger stress-related signal cascades through intracellular kinases or phytohormone-based signaling to regulate growth/defense transcriptional regulators. Our previous study of lignin-modified plants with

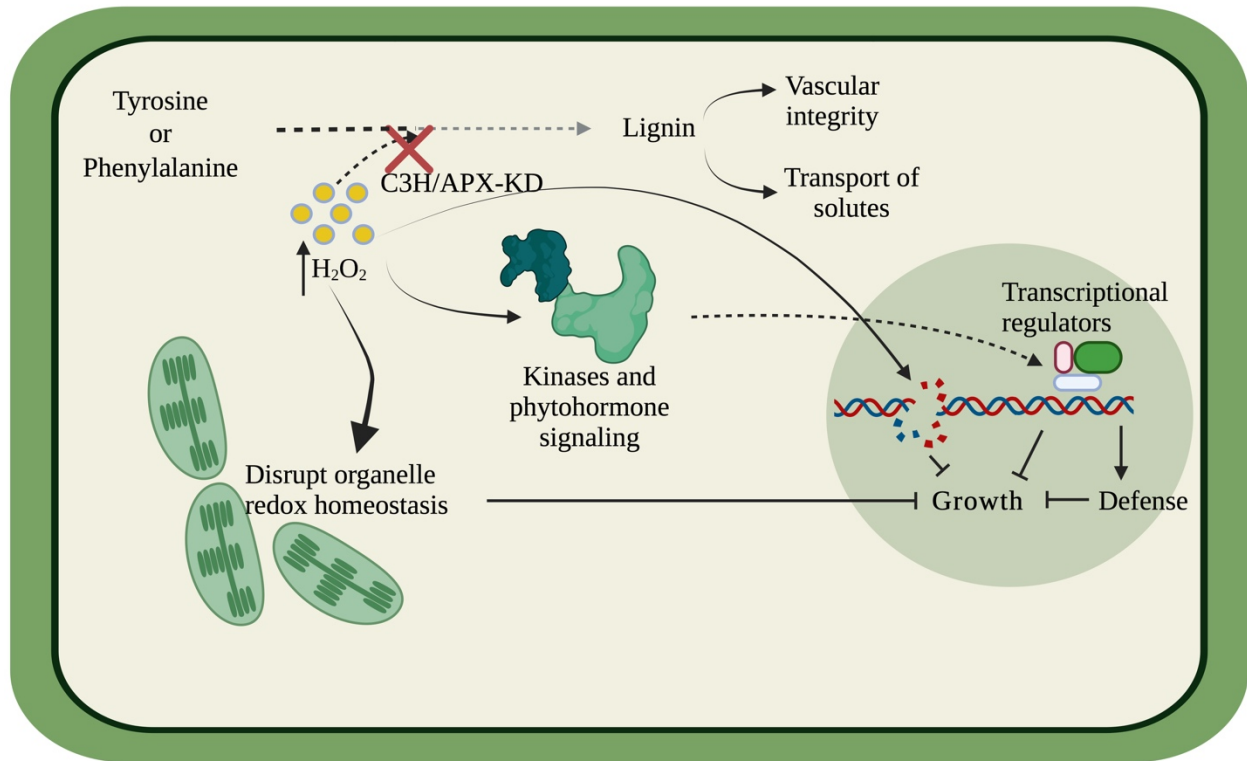


Figure 7.8: Simplified model linking C3H/APX and the lignin biosynthesis pathway to other biological processes.

C3H/APX utilizes hydrogen peroxide (H₂O₂) as a co-factor in the lignin biosynthetic pathway. The deficiency of C3H/APX elevated cellular oxidative stress thus disrupting the cellular and organelle redox balance contributing to reduced photosynthetic ability. The redox imbalance can directly elevate DNA damage. Moreover, the disrupted redox homeostasis can transduce stress signal downstream through intracellular kinases or phytohormone-based signaling to regulated growth/defense transcription and translation with the help of transcriptional regulators. The compromised photosynthetic ability, oxidative DNA damage, elevated defense response in tandem could contribute to the negative growth phenotype of the C3H/APX-KD lines.

growth defects has shown alterations in the carbon-nitrogen balance (Barros et al., 2022). Meanwhile, this study has shown ROS and/or redox imbalance correlate with these growth inhibitory effects. Further studies are needed to determine if aspects of the trade-offs between carbon and nitrogen allocation or ROS imbalance or both are associated with the growth inhibition observed in lignin-modified plants.

CHAPTER 8: OUTLOOK AND CONCLUSIONS

8.1 Overview of this dissertation

Our ability to comprehensively identify and quantify the proteome of an organism using mass spectrometry has proven to provide important information regarding molecular processes and the cellular function of biological systems. This ability has made mass spectrometry one of the indispensable analytical tools for systems biology. Like in many other areas of systems biology, mass spectrometry-based comprehensive characterization of proteins and their post-translational modifications has emerged as a promising technique for a system-wide understanding of bioenergy plants and their dynamic environment and represents a crucial aspect of bioenergy research.

The main goal of bioenergy research is to utilize perennial grasses and trees to get a sustainable high yield of bioenergy so that global concerns of energy security, climate change, and environmental pollution can be mitigated (Karp and Shield, 2008). This goal of bioenergy research depends upon two main factors. The first factor is improvement in biomass feedstock yield and production. The higher yield of biomass directly correlates with a higher yield of biofuel and bioproducts. The second factor is the efficiency in conversion of biomass feedstock containing cellulose, hemicellulose, and lignin structure to biofuel and high valued bioproducts. The goal of this dissertation was to utilize mass spectrometry-based proteomics and post-translational modifications analysis to gain system-wide insight into the dynamic environment of bioenergy plants as well as to get a detailed molecular understanding of lignin-modified bioenergy crops. This advancement from this dissertation can directly benefit the overall goal of bioenergy research because the biotic and abiotic environments are known to influence the health and fitness of bioenergy plants, and lignin is known to dictate the efficiency of biomass conversion. For a successful outcome towards this overall goal, the dissertation focused on addressing the following key aims i.e., i) integrate a reductionist approach that allows to generate the tractable and controllable defined microbial community representing the natural poplar rhizosphere microbiome with high-resolution metaproteomics to study microbe-microbe interactions during the assembly of a poplar-associated defined microbial community (chapter 3), ii) advance comprehensive understanding of proteins and pathways regulated by the plant and microbe individually during

plant-microbe interaction to get insights into how the plant and microbe each respond to the plant-microbe interaction (chapter 4 and 5), iii) characterize the proteins and PTMs that show temporal regulation across the diurnal cycle to get a more complete understanding of how plants respond to environmental cues (chapter 6), and iv) gain a deeper molecular understanding of the pleiotropic effects in lignin modified bioenergy crops (chapter 7).

The first step towards our goal is to understand the current state of the field, knowledge gaps, and challenges. Currently, the challenges with bioenergy research are not about being able to produce biofuel but about how we can maintain sustainable growth and a high yield of bioenergy. We realized there are two factors that could influence this outcome which are introduced and discussed in chapter 1. Further, the basics of mass spectrometry and what makes it a viable tool for bioenergy research are described in chapter 1. The second step is to know what are the mass spectrometry techniques that could be utilized in this process and what advancements in these methods could allow us to gather more information that didn't exist before. The mass spectrometry-related experimental, technical, analytical, and computational advancements are discussed in chapter 2.

One factor that benefits the goal of bioenergy research is through a better understanding of the biotic environment that is known to influence the health and fitness of bioenergy plants. The interactions and regulations between plants and their biotic environment can be divided into two categories i.e., microbe-microbe and plant-microbe interactions. Inter-microbial interactions within or outside plant tissues are intimately connected to plants' essential processes including water and nutrient acquisition, stress response, and host defenses (Pii, 2015, De Zelicourt, 2013, Berendsen et al., 2012, Pineda et al., 2015). In chapter 3, a reductionist approach was utilized that allowed the generation of the tractable and controllable defined microbial community representing the natural poplar rhizosphere microbiome. The reductionist approach along with metaproteomics and PTM analysis was utilized for a 10-member defined community of poplar rhizosphere to uncover to what extent inter-microbial interactions shape microbial assemblages and what molecular mechanisms are used by microbes to cooperate, compete, and persist in complex microbial consortia. In brief, this study found that during community assembly, *Pseudomonas* sp. GM17 dominates the microbial community by creating unfavorable conditions for other microbes. *Pantoea* sp. YR343 copes with the stressful environment by increasing carotenoid biosynthesis,

motility, and antibiotics resistance, and *Bacillus* sporulates as an adaptive response. Other microbes that cannot cope or adapt eventually die out. To the best of this author's knowledge, this study was the first metaproteomic study conducted on plant-associated microbes for studying the inter-microbial interactions. This has led the foundation for several other metaproteomic studies to follow on complex systems for higher-order interactions.

While chapter 3 discussed the inter-microbial regulations, chapter 4 focused on the plant-microbe interaction from the perspective of the plant. A specific G-type lectin receptor-like kinase (*PtLecRLK1*) has been identified as a crucial gene found in specific poplar species (*Populus trichocarpa*) that regulates the poplar root colonization by the beneficial fungus *L. bicolor* (Labbe et al., 2019, Qiao et al., 2021). *PtLecRLK1* being a kinase protein, in chapter 04, phosphoproteomics was utilized to understand the *PtLecRLK1*-induced phosphorylation signaling cascade in transgenic and wild type non-host species i.e., switchgrass and poplar hybrid co-inoculated with *L. bicolor*. A notable result of this study was that *PtLecRLK1* regulates the phosphorylation status of cellular protein kinase called cGMP kinase which is responsible for the downstream signaling cascade necessary for plant-fungal symbiosis and these features are shared between transgenic plants expressing *PtLecRLK1*.

In a similar context, the plant-microbe interaction from the perspective of the microbe is investigated in chapter 5. The phytochemicals secreted by the plants are crucial for plant-microbe interactions as well as to shape the microbial community. HEHEAA is a poplar signaling compound that is perceived by *Pseudomonas* GM79. In chapter 5, proteomics was utilized to reveal the mechanism by which the plant-based signaling compounds regulate various biological processes in the microbe and how it regulates traits important for plant colonization. Interestingly, the study revealed that upon HEHEAA addition, *Pseudomonas* sp. GM79 mainly alters biological processes that fall under three categories i.e., chemotactic traits, phytopathogen defending traits, and plant growth promoting traits by regulating the abundance of proteins with functions in chemotaxis, siderophore synthesis and transport, antibiotic synthesis, and T6SS.

Similar to biotic environments, abiotic environments are crucial factors influencing the plants' phenotype. The day-night rhythmic patterns (diurnal cycle) play a major role in a plant's

phenotypic plasticity, which allows for new phenotypes under changing environments and synchronizes plant behavior across cellular and molecular levels. In chapter 6, bottom-up proteomics using high-resolution mass spectrometry and *de novo*-assisted database searching was employed to identify abundance changes in proteins and post-translational modifications in poplar leaf tissue across 12-hr light/12-hr dark photoperiod. This study characterized that around 12% of identified proteins and more than 100 PTMs were under temporal regulation in *Populus*. This study provides insight into how plants respond to environmental cues at the molecular level.

While the biotic and abiotic environment can influence this overall goal of bioenergy research by impacting the health and fitness of the bioenergy plant, there is still a need for efficient conversion of biomass to get viable and sustainable biofuels and bioproducts. The bottleneck in this conversion has been due to the plant's lignin content which is a crucial secondary plant cell wall component, providing structural and physical integrity. Reducing the lignin through the genetic modification of enzymes involved in lignin biosynthesis can be a viable approach but often introduces several unwanted pleiotropic effects such as stunted plant growth and an associated yield penalty (Ha et al., 2021, Muro-Villanueva et al., 2019, Vanholme et al., 2019). In chapter 07, bottom-up proteomics was utilized to gain a deeper understanding of how knockdown of C3H/APX, a lignin biosynthesis enzyme of *Brachypodium*, leads to a negative growth phenotype. The study revealed that C3H/APX-knockdown leads to substantial alterations in the *Brachypodium* stem proteome, and major changes can be observed in the phenylpropanoid biosynthesis pathway, transcriptional regulators, phytohormone signaling, senescence process, photosynthesis, DNA repair, and plant defense. C3H/APX-KD might have regulated these changes via the disruption of redox homeostasis potentially by elevating cellular H₂O₂ levels. To link the overall focus of the dissertation, future experiments could be performed to see if the C3H/APX-KD impacts the plant's ability to interact with pathogenic microbes.

Altogether, the advancements and findings described in Chapters 3-7 will greatly improve our understanding of bioenergy plants and their dynamic environment. Additionally, the work provides a positive step towards understanding the current limitations of bioenergy plants to produce sustainable biofuels and bioproducts. These findings will provide a springboard for numerous future studies aiming to gain a deeper understanding of bioenergy plants. In addition, this final

chapter outlines the remaining challenges with mass spectrometry-based proteomics and PTM analysis as well as a discussion on what has been accomplished in this dissertation and how that impacts bioenergy plant research.

8.2 Remaining challenges and future considerations

Although mass spectrometry-based proteins and PTMs characterization have emerged as a promising technique for systems biology, there is still room for improvement in order to utilize MS to the fullest. As mentioned earlier in chapters 1 and 2, the initial advancements in proteomics research was due to advancements in a few key areas i.e. i) robust separation of proteins using two-dimensional gel electrophoresis technique (Sperling, 2001), ii) advancement of MS methods for the analysis of accurate mass and chemical structure (Loo et al., 1999, Aicher et al., 1998), iii) substantial effort on large-scale genome research that helped to generate high-quality databases (Zhang et al., 2002) and iv) progress in developing bioinformatics tools that helped to analyze large volumes of MS data (Nefedov et al., 2011). These advancements span around the three main steps of proteomics research i.e., (i) proteomics sample preparation, (ii) peptide separation and tandem mass spectrometry data acquisition, and (iii) data analysis and biological interpretation. The bottleneck in any of these steps can significantly drag the efficiency of other steps so, the next set of advancements should aim to improve all of them.

Global bottom-up proteomics are relatively sensitive and provide in-depth coverage for a single organism using the standard protein extraction method. In standard LC-MS/MS measurement, around 40-60% of a total proteome can be measured which can be considered reasonably deep under the assumption that the organism will not express 100% of its total proteome at a given condition. In plant holobiont research, the ability to holistically study inter-microbial interactions, plant-microbial interactions and identify molecular mechanisms used by plants and microbes to sense and respond in the complex environment requires expansive proteome coverage. While we get reasonably deep proteome coverage for a single organism, this is not entirely true for metaproteomic experiments. Usually, in metaproteomics, we get deep enough coverage for a few topmost abundant organisms because in a complex community of plant and microbes the measured proteome depth can be affected by the growth of microbes, metabolic rate, or experimental design

that favors only some abundant organisms. This limits our ability to provide molecular and functional level information only to those few abundantly identified organisms. Similarly in the presence of plants, more abundant plant proteins can mask low-abundance microbial proteins. An effort on increasing the LC separation can be helpful to minimize the co-elution of peptides and ion suppression events. Moreover, more sensitive instruments can be beneficial. In the case of one or two proteins dominating the chromatography and MS run, selective depletion can be performed so that low abundant proteins can be measured. For example, while studying the leaf proteome, often the RuBisCO protein is depleted so that enough coverage can be obtained for other proteins (Sehrawat et al., 2013).

In the context of proteome coverage, the overall coverage for membrane proteins or other hydrophobic proteins in the global proteomics run is relatively poor. Even though the hydrophilic C and N terminus are often detected in MS run, a significant portion of membrane or hydrophobic proteins might not be detected because the sample preparation methods are not optimal for these sets of proteins. There are diverse proteomics sample preparation methods available, however finding a robust method that doesn't compromise in any aspect can be difficult. Careful consideration is needed to understand the limitation of the sample preparation method being utilized and decide if the method utilized can answer the specific biological questions being asked.

The throughput of the mass spectrometry instrument has improved drastically over the last few years. A typical label free bottom-up proteomics run that would take ~22 hours with 11 step salt pulses in MudPIT a few years ago, can be performed in 4-6 hours nowadays with similar or better coverage. These sorts of improvements are expected to continuously improve with the development of newer instruments with better resolution, accuracy, higher dynamic range, scan speed, detection limit, and duty cycle. However, for the high throughput experiments, even faster measurements or alternative methods that can sample multiple samples simultaneously might be desired. Multiplexing has been an available option for a while; however, it comes with a hefty price and added challenges such as measurement artifacts (Pappireddi et al., 2019). Similarly, high throughput experiments require an improvement and optimization in the proteomic sample preparation step. An improvement in cell lysis and tissue homogenization steps can help to increase reproducibility, reliability and high-throughput that minimize cross-contamination and improve

the number of proteins obtained. BeatBox sample homogenizer by PreOmics is a rapid and scalable instrument for tissues and cells disruption for up to 96 samples (Raynie, 2022). It has a small footprint, low noise and can homogenize samples in around 10 minutes with minimal cross-contamination (Raynie, 2022). In that context, PIXUL by Active Motif is a 96 samples sonicator that can be utilized for consistent shearing of the samples for high-throughput proteomics work.

While access to efficient cell preparation techniques, extraction methods and state of the art instrumentation is key to quality proteomics research, recognition of the limitations within data analysis and informatics is equally important. Data analysis and biological interpretation of the data represent the crucial aspect of bottom-up proteomics. Failure in proper data analysis can render the well-executed sample preparation and MS run meaningless and can mask or even worse provide the false biological information. To improve the data analysis efficiency, while planning the experiment it might always be a good idea to perform the power analysis to identify the optimal number of replicates needed for a study to obtain the reproducibility. At the lowest, three biological replicates are needed for the statistical analysis; however, this can be limiting if any of the sample run fails for some unexpected reason or acts as an outlier. Therefore, careful planning before the sample preparation can be beneficial during the data analysis. Another challenge in data analysis and informatics is a proper utilization of entire data obtained in MS run for researchers wishing to understand or optimize biological systems. Researchers often tend to make a story based on the biggest changing proteins in terms of the abundance obtained from differentially abundant analysis. While there is nothing wrong with this practice, not all biologically relevant changes have to be biggest. Therefore, it is important to carefully look at the proteins with small fold changes as well.

A key challenge in terms of biological interpretation has been with the annotations. The description and the functional annotations of the genes can be ambiguous. While working with the bioenergy plants like poplar, switchgrass etc. most of the annotations are derived from the model plant *Arabidopsis* based on the sequence homolog. They can be ambiguous, incorrect, and missing. Most widely used functional annotation based in Gene Ontology (GO) can provide the information on cellular localization, biological processes, and molecular function. However, this annotation can be vague and misleading, especially when only generic functional terms from an irrelevant part of

the gene ontology hierarchy, such as ‘response to abiotic stress’, are considered. KEGG can be a useful alternative to GO however, KEGG also suffers from the limitation of multiple terms mapping to the same protein although not as extreme as GO. For the PTMs analysis, a large number of PTMs identified are still lacking any functional information in the public database. This makes it challenging to gain biological meaning from any given PTMs. The proteomics community should work together to build PTM repositories or tools that can help to link the PTM identified to meaningful information. Moreover, focus should also be given on studying PTMs other than the selective few commonly studied ones. One path towards this goal would be to document detailed information about the PTMs across multiple experimental conditions and identify the PTMs that are more reliably and consistently detected than others. Once the detected PTMs are experimentally validated, make the results easily findable so that others can utilize these resources for further experiments.

After the experiments are done and research is published, it is important to share the data so that other researchers can utilize these as the resources for future studies. While proper data sharing for -omics technology is becoming more common, the ability to FAIRify (Findable, Accessible, Interoperable and Re-usable) mass spectrometry data from proteomics still represents a significant challenge for the research community. It is the opinion of the author that data sharing using the FAIR principle will see tremendous growth in the near future not only for proteomics but also for PTM specific data which will improve the reproducibility and sustainability of scientific research.

8.3 Final thoughts

It is an exciting time to be utilizing mass-spectrometry for the systems biology studies. In this dissertation, mass-spectrometry was implemented in various biological samples (plants, fungi, bacteria) and to acquire different levels of information (global proteomics, single PTMs analysis, global PTM profiling). The biological samples and different types of information obtained herein are just a small representative of what could be done using mass spectrometry. Mass spectrometry-based proteomics is truly a versatile and robust technique that is shaping a new generation of scientists who can utilize the analytical technique to obtain the large-scale data that has to be

interpreted using the advance bioinformatic skills to uncover the underlying fundamental biological questions.

While there is and will always be a room for improvement in the mass spectrometry-proteomics and PTM techniques, we are currently experiencing a high pace advancement in cutting-edge technology where the limit of mass spectrometry and tools associated with it are pushed to the new limit every year. Unlike a couple of decades ago, scientists are now more receptive to the idea of omics. As a result, scientists are open towards the comprehensive and reproducible protein datasets that are generated via MS-based proteomics. There has been slowly but surely a paradigm shift from myopic investigations focusing on a small set of proteins to the system wide understanding of the intricacy of the biological system. To that end, even if the goal is to focus on the small set of proteins of interest, mass spectrometry based specialized techniques like targeted proteomics via multiple reaction monitoring can easily be adapted towards that goal. In addition, targeted proteomics can not only be a replacement but be advantageous over traditional techniques like Western Blot and ELISA assays due to its high specificity, sensitivity, robustness, and ability to identify proteins of interest without the need of antibodies. Over the last few years, we have observed a tremendous pace of advancements and range of questions answered using omic technologies. As a next step towards the holistic understanding of systems biology, proteomics results should be integrated with the other omics like genomics, transcriptomics, metabolomics and phenomics to better connect the genotype to a phenotype. The development of open community resource platforms such as Department of Energy Systems Biology Knowledgebase (KBase, <http://kbase.us>) (Arkin et al., 2018) is expected to be tremendously useful for integrating and analyzing these multi-omics data. By studying the combined behavior of the different biological molecules that create networks of interactions, systems biology is expected to precisely compute the system's response to perturbations. This can be revolutionary for the field of biological research including biomedical research focusing on understanding diseases or disorders; cell biology aiming to understand cellular growth, development, or differentiation; targeted therapeutics intending for precision medicines; single cell research trying to understand the molecular and cellular heterogeneity.

LIST OF REFERENCES

- ABRAHAM, P. E., YIN, H., BORLAND, A. M., WEIGHILL, D., LIM, S. D., DE PAOLI, H. C., ENGLE, N., JONES, P. C., AGH, R., WESTON, D. J., WULLSCHLEGER, S. D., TSCHAPLINSKI, T., JACOBSON, D., CUSHMAN, J. C., HETTICH, R. L., TUSKAN, G. A. & YANG, X. 2016. Transcript, protein and metabolite temporal dynamics in the CAM plant Agave. *Nat Plants*, 2, 16178.
- AEBERSOLD, R., AGAR, J. N., AMSTER, I. J., BAKER, M. S., BERTOZZI, C. R., BOJA, E. S., COSTELLO, C. E., CRAVATT, B. F., FENSELAU, C., GARCIA, B. A., GE, Y., GUNAWARDENA, J., HENDRICKSON, R. C., HERGENROTHER, P. J., HUBER, C. G., IVANOV, A. R., JENSEN, O. N., JEWETT, M. C., KELLEHER, N. L., KIESSLING, L. L., KROGAN, N. J., LARSEN, M. R., LOO, J. A., OGORZALEK LOO, R. R., LUNDBERG, E., MACCOSS, M. J., MALLICK, P., MOOTHA, V. K., MRKSICH, M., MUIR, T. W., PATRIE, S. M., PESAVENTO, J. J., PITTERI, S. J., RODRIGUEZ, H., SAGHATELIAN, A., SANDOVAL, W., SCHLUTER, H., SECHI, S., SLAVOFF, S. A., SMITH, L. M., SNYDER, M. P., THOMAS, P. M., UHLEN, M., VAN EYK, J. E., VIDAL, M., WALT, D. R., WHITE, F. M., WILLIAMS, E. R., WOHLSCHLAGER, T., WYSOCKI, V. H., YATES, N. A., YOUNG, N. L. & ZHANG, B. 2018. How many human proteoforms are there? *Nat Chem Biol*, 14, 206-214.
- AEBERSOLD, R. & MANN, M. 2016. Mass-spectrometric exploration of proteome structure and function. *Nature*, 537, 347-55.
- AGATI, G., AZZARELLO, E., POLLASTRI, S. & TATTINI, M. 2012. Flavonoids as antioxidants in plants: location and functional significance. *Plant Sci*, 196, 67-76.
- AGRAWAL, G. K., JWA, N. S., IWAHASHI, H. & RAKWAL, R. 2003. Importance of ascorbate peroxidases OsAPX1 and OsAPX2 in the rice pathogen response pathways and growth and reproduction revealed by their transcriptional profiling. *Gene*, 322, 93-103.
- AGUILAR, M. I. 2004. Reversed-phase high-performance liquid chromatography. *Methods Mol Biol*, 251, 9-22.
- AHRNE, E., MULLER, M. & LISACEK, F. 2010. Unrestricted identification of modified proteins using MS/MS. *Proteomics*, 10, 671-86.
- AICHER, L., WAHL, D., ARCE, A., GRENET, O. & STEINER, S. 1998. New insights into cyclosporine A nephrotoxicity by proteome analysis. *Electrophoresis*, 19, 1998-2003.
- AKIYAMA, K., MATSUZAKI, K. & HAYASHI, H. 2005. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature*, 435, 824-7.
- ALCOFORADO DINIZ, J., LIU, Y. C. & COULTHURST, S. J. 2015. Molecular weaponry: diverse effectors delivered by the Type VI secretion system. *Cell Microbiol*, 17, 1742-51.

- ALLEN, J. F. 1992. Protein phosphorylation in regulation of photosynthesis. *Biochim Biophys Acta*, 1098, 275-335.
- ANDERSON, N. G., MATHESON, A. & ANDERSON, N. L. 2001. Back to the future: the human protein index (HPI) and the agenda for post-proteomic biology. *Proteomics*, 1, 3-12.
- ANDERSON, N. L. & ANDERSON, N. G. 2002. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics*, 1, 845-67.
- ANDERSSON, L. & PORATH, J. 1986. Isolation of phosphoproteins by immobilized metal (Fe³⁺) affinity chromatography. *Anal Biochem*, 154, 250-4.
- ANILKUMAR, R. R., EDISON, L. K. & PRADEEP, N. 2017. Exploitation of fungi and actinobacteria for sustainable agriculture. *Microbial Biotechnology*. Springer.
- ARDITO, F., GIULIANI, M., PERRONE, D., TROIANO, G. & LO MUZIO, L. 2017. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int J Mol Med*, 40, 271-280.
- ARKIN, A. P., COTTINGHAM, R. W., HENRY, C. S., HARRIS, N. L., STEVENS, R. L., MASLOV, S., DEHAL, P., WARE, D., PEREZ, F., CANON, S., SNEDDON, M. W., HENDERSON, M. L., RIEHL, W. J., MURPHY-OLSON, D., CHAN, S. Y., KAMIMURA, R. T., KUMARI, S., DRAKE, M. M., BRETTIN, T. S., GLASS, E. M., CHIVIAN, D., GUNTER, D., WESTON, D. J., ALLEN, B. H., BAUMOHL, J., BEST, A. A., BOWEN, B., BRENNER, S. E., BUN, C. C., CHANDONIA, J. M., CHIA, J. M., COLASANTI, R., CONRAD, N., DAVIS, J. J., DAVISON, B. H., DEJONGH, M., DEVOID, S., DIETRICH, E., DUBCHAK, I., EDIRISINGHE, J. N., FANG, G., FARIA, J. P., FRYBARGER, P. M., GERLACH, W., GERSTEIN, M., GREINER, A., GURTOWSKI, J., HAUN, H. L., HE, F., JAIN, R., JOACHIMIAK, M. P., KEEGAN, K. P., KONDO, S., KUMAR, V., LAND, M. L., MEYER, F., MILLS, M., NOVICHKOV, P. S., OH, T., OLSEN, G. J., OLSON, R., PARRELLO, B., PASTERNAK, S., PEARSON, E., POON, S. S., PRICE, G. A., RAMAKRISHNAN, S., RANJAN, P., RONALD, P. C., SCHATZ, M. C., SEAVER, S. M. D., SHUKLA, M., SUTORMIN, R. A., SYED, M. H., THOMASON, J., TINTLE, N. L., WANG, D., XIA, F., YOO, H., YOO, S. & YU, D. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol*, 36, 566-569.
- ARO, E. M. & OHAD, I. 2003. Redox regulation of thylakoid protein phosphorylation. *Antioxid Redox Signal*, 5, 55-67.
- BACKER, R., ROKEM, J. S., ILANGUMARAN, G., LAMONT, J., PRASLICKOVA, D., RICCI, E., SUBRAMANIAN, S. & SMITH, D. L. 2018. Plant Growth-Promoting Rhizobacteria: Context, Mechanisms of Action, and Roadmap to Commercialization of Biostimulants for Sustainable Agriculture. *Front Plant Sci*, 9, 1473.

- BAIS, H. P., WEIR, T. L., PERRY, L. G., GILROY, S. & VIVANCO, J. M. 2006. The role of root exudates in rhizosphere interactions with plants and other organisms. *Annu Rev Plant Biol*, 57, 233-66.
- BAKKENIST, C. J. & KASTAN, M. B. 2003. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature*, 421, 499-506.
- BAKKER, P. A., BERENDSEN, R. L., DOORNBOS, R. F., WINTERMANS, P. C. & PIETERSE, C. M. 2013. The rhizosphere revisited: root microbiomics. *Front Plant Sci*, 4, 165.
- BALLIF, B. A., CAREY, G. R., SUNYAEV, S. R. & GYGI, S. P. 2008. Large-scale identification and evolution indexing of tyrosine phosphorylation sites from murine brain. *J Proteome Res*, 7, 311-8.
- BANERJEE, S. & MAZUMDAR, S. 2012. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem*, 2012, 282574.
- BARROS, J., ESCAMILLA-TREVINO, L., SONG, L., RAO, X., SERRANI-YARCE, J. C., PALACIOS, M. D., ENGLE, N., CHOUDHURY, F. K., TSCHAPLINSKI, T. J., VENABLES, B. J., MITTLER, R. & DIXON, R. A. 2019. 4-Coumarate 3-hydroxylase in the lignin biosynthesis pathway is a cytosolic ascorbate peroxidase. *Nature Communication*, 10, 1994.
- BARROS, J., SERK, H., GRANLUND, I. & PESQUET, E. 2015. The cell biology of lignification in higher plants. *Ann Bot*, 115, 1053-74.
- BARROS, J., SERRANI-YARCE, J. C., CHEN, F., BAXTER, D., VENABLES, B. J. & DIXON, R. A. 2016. Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. *Nat Plants*, 2, 16050.
- BARROS, J., SHRESTHA, H. K., SERRANI-YARCE, J. C., ENGLE, N., ABRAHAM, P. E., TSCHAPLINSKI, T. J., HETTICH, R. L. & DIXON, R. A. 2022. Proteomic and metabolic disturbances in lignin-modified *Brachypodium distachyon*. *Plant Cell*.
- BATEMAN, N. W., GOULDING, S. P., SHULMAN, N. J., GADOK, A. K., SZUMLINSKI, K. K., MACCOSS, M. J. & WU, C. C. 2014. Maximizing peptide identification events in proteomic workflows using data-dependent acquisition (DDA). *Mol Cell Proteomics*, 13, 329-38.
- BATTH, T. S., TOLLENAERE, M. X., RUTHER, P., GONZALEZ-FRANQUESA, A., PRABHAKAR, B. S., BEKKER-JENSEN, S., DESHMUKH, A. S. & OLSEN, J. V. 2019. Protein Aggregation Capture on Microparticles Enables Multipurpose Proteomics Sample Preparation. *Mol Cell Proteomics*, 18, 1027-1035.

- BAUER, W. D. & MATHESIUS, U. 2004. Plant responses to bacterial quorum sensing signals. *Curr Opin Plant Biol*, 7, 429-33.
- BECKER, J., EISENHAUER, N., SCHEU, S. & JOUSSET, A. 2012. Increasing antagonistic interactions cause bacterial communities to collapse at high diversity. *Ecol Lett*, 15, 468-74.
- BECKHAM, G. T., JOHNSON, C. W., KARP, E. M., SALVACHUA, D. & VARDON, D. R. 2016. Opportunities and challenges in biological lignin valorization. *Curr Opin Biotechnol*, 42, 40-53.
- BELLANDE, K., BONO, J. J., SAVELLI, B., JAMET, E. & CANUT, H. 2017. Plant Lectins and Lectin Receptor-Like Kinases: How Do They Sense the Outside? *Int J Mol Sci*, 18.
- BERENDSEN, R. L., PIETERSE, C. M. & BAKKER, P. A. 2012. The rhizosphere microbiome and plant health. *Trends Plant Sci*, 17, 478-86.
- BERG, D., YODIM, M. B. & RIEDERER, P. 2004. Redox imbalance. *Cell Tissue Res*, 318, 201-13.
- BERLEMAN, J. E. & BAUER, C. E. 2005. Involvement of a Che-like signal transduction cascade in regulating cyst cell development in *Rhodospirillum rubrum*. *Mol Microbiol*, 56, 1457-66.
- BESSEAU, S., HOFFMANN, L., GEOFFROY, P., LAPIERRE, C., POLLET, B. & LEGRAND, M. 2007. Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. *Plant Cell*, 19, 148-62.
- BEYNON, R. J. & PRATT, J. M. 2005. Metabolic labeling of proteins for proteomics. *Mol Cell Proteomics*, 4, 857-72.
- BI, S. & LAI, L. 2015. Bacterial chemoreceptors and chemoeffectors. *Cell Mol Life Sci*, 72, 691-708.
- BIBLE, A. N., FLETCHER, S. J., PELLETIER, D. A., SCHADT, C. W., JAWDY, S. S., WESTON, D. J., ENGLE, N. L., TSCHAPLINSKI, T., MASYUKO, R., POLISETTI, S., BOHN, P. W., COUTINHO, T. A., DOKTYCZ, M. J. & MORRELL-FALVEY, J. L. 2016. A Carotenoid-Deficient Mutant in *Pantoea* sp. YR343, a Bacteria Isolated from the Rhizosphere of *Populus deltoides*, Is Defective in Root Colonization. *Front Microbiol*, 7, 491.
- BIEMANN, K. 1988. Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom*, 16, 99-111.
- BINDEA, G., MLECNIK, B., HACKL, H., CHAROENTONG, P., TOSOLINI, M., KIRILOVSKY, A., FRIDMAN, W. H., PAGES, F., TRAJANOSKI, Z. & GALON, J.

2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, 25, 1091-3.
- BINNS, A. N., CHEN, R. H., WOOD, H. N. & LYNN, D. G. 1987. Cell division promoting activity of naturally occurring dehydrodiconiferyl glucosides: do cell wall components control cell division? *Proceedings of the National Academy of Sciences*, 84, 980-4.
- BLAIR, P. M., LAND, M. L., PIATEK, M. J., JACOBSON, D. A., LU, T. S., DOKTYCZ, M. J. & PELLETIER, D. A. 2018. Exploration of the Biosynthetic Potential of the Populus Microbiome. *mSystems*, 3.
- BLASIUS, D., FEIL, W., KOTTKE, I. & OBERWINKLER, F. 1986. Hartig net structure and formation in fully ensheathed ectomycorrhizas. *Nordic Journal of Botany*, 6, 837-842.
- BLIN, K., SHAW, S., STEINKE, K., VILLEBRO, R., ZIEMERT, N., LEE, S. Y., MEDEMA, M. H. & WEBER, T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res*, 47, W81-W87.
- BOERJAN, W., RALPH, J. & BAUCHER, M. 2003. Lignin biosynthesis. *Annu Rev Plant Biol*, 54, 519-46.
- BONAWITZ, N. D. & CHAPPLE, C. 2010. The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu Rev Genet*, 44, 337-63.
- BONAWITZ, N. D. & CHAPPLE, C. 2013. Can genetic engineering of lignin deposition be accomplished without an unacceptable yield penalty? *Curr Opin Biotechnol*, 24, 336-43.
- BONAWITZ, N. D., KIM, J. I., TOBIMATSU, Y., CIESIELSKI, P. N., ANDERSON, N. A., XIMENES, E., MAEDA, J., RALPH, J., DONOHOE, B. S., LADISCH, M. & CHAPPLE, C. 2014. Disruption of Mediator rescues the stunted growth of a lignin-deficient Arabidopsis mutant. *Nature*, 509, 376-80.
- BONAWITZ, N. D., SOLTAU, W. L., BLATCHLEY, M. R., POWERS, B. L., HURLOCK, A. K., SEALS, L. A., WENG, J. K., STOUT, J. & CHAPPLE, C. 2012. REF4 and RFR1, subunits of the transcriptional coregulatory complex mediator, are required for phenylpropanoid homeostasis in Arabidopsis. *J Biol Chem*, 287, 5434-45.
- BONIFACIO, A., MARTINS, M. O., RIBEIRO, C. W., FONTENELE, A. V., CARVALHO, F. E., MARGIS-PINHEIRO, M. & SILVEIRA, J. A. 2011. Role of peroxidases in the compensation of cytosolic ascorbate peroxidase knockdown in rice plants under abiotic stress. *Plant Cell Environ*, 34, 1705-22.
- BOREHAM, D. R. & MARTIN, B. K. 1969. The kinetics of elimination of salicylic acid and the formation of gentisic acid. *Br J Pharmacol*, 37, 294-300.

- BORTOLO, T., MARCHIOSI, R., VIGANO, J., SIQUEIRA-SOARES, R. C., FERRO, A. P., BARRETO, G. E., BIDO, G. S., ABRAHAO, J., DOS SANTOS, W. D. & FERRARESE-FILHO, O. 2018. Trans-aconitic acid inhibits the growth and photosynthesis of *Glycine max*. *Plant Physiol Biochem*, 132, 490-496.
- BOUWMEESTER, K. & GOVERS, F. 2009. Arabidopsis L-type lectin receptor kinases: phylogeny, classification, and expression profiles. *J Exp Bot*, 60, 4383-96.
- BOWSHER, A. W., KEARNS, P. J. & SHADE, A. 2019. 16S rRNA/rRNA Gene Ratios and Cell Activity Staining Reveal Consistent Patterns of Microbial Activity in Plant-Associated Soil. *mSystems*, 4.
- BRAZELTON, J. N., PFEUFER, E. E., SWEAT, T. A., GARDENER, B. B. & COENEN, C. 2008. 2,4-diacetylphloroglucinol alters plant root development. *Mol Plant Microbe Interact*, 21, 1349-58.
- BROWN, C. W., SRIDHARA, V., BOUTZ, D. R., PERSON, M. D., MARCOTTE, E. M., BARRICK, J. E. & WILKE, C. O. 2017. Large-scale analysis of post-translational modifications in *E. coli* under glucose-limiting conditions. *Bmc Genomics*, 18.
- BROWN, L., WOLF, J. M., PRADOS-ROSALES, R. & CASADEVALL, A. 2015. Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. *Nat Rev Microbiol*, 13, 620-30.
- BROWN, S. D., UTTURKAR, S. M., KLINGEMAN, D. M., JOHNSON, C. M., MARTIN, S. L., LAND, M. L., LU, T. Y., SCHADT, C. W., DOKTYCZ, M. J. & PELLETIER, D. A. 2012. Twenty-one genome sequences from *Pseudomonas* species and 19 genome sequences from diverse bacteria isolated from the rhizosphere and endosphere of *Populus deltoides*. *J Bacteriol*, 194, 5991-3.
- BRUNETTI, C., DI FERDINANDO, M., FINI, A., POLLASTRI, S. & TATTINI, M. 2013. Flavonoids as antioxidants and developmental regulators: relative significance in plants and humans. *Int J Mol Sci*, 14, 3540-55.
- BRYANT, N. D., PU, Y., TSCHAPLINSKI, T. J., TUSKAN, G. A., MUCHERO, W., KALLURI, U. C., YOO, C. G. & RAGAUSKAS, A. J. 2020. Transgenic Poplar Designed for Biofuels. *Trends Plant Sci*, 25, 881-896.
- BULGARELLI, D., ROTT, M., SCHLAEPPI, K., VER LOREN VAN THEMAAT, E., AHMADINEJAD, N., ASSENZA, F., RAUF, P., HUETTEL, B., REINHARDT, R., SCHMELZER, E., PEPLIES, J., GLOECKNER, F. O., AMANN, R., EICKHORST, T. & SCHULZE-LEFERT, P. 2012. Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature*, 488, 91-5.
- BUNNIK, E. M. & LE ROCH, K. G. 2013. An Introduction to Functional Genomics and Systems Biology. *Adv Wound Care (New Rochelle)*, 2, 490-498.

- BUSCOT, F., MUNCH, J. C., CHARCOSSET, J. Y., GARDES, M., NEHLS, U. & HAMPP, R. 2000. Recent advances in exploring physiology and biodiversity of ectomycorrhizas highlight the functioning of these symbioses in ecosystems. *FEMS Microbiol Rev*, 24, 601-14.
- CAHILL, M. A., WOZNY, W., SCHWALL, G., SCHROER, K., HOLZER, K., POZNANOVIC, S., HUNZINGER, C., VOGT, J. A., STEGMANN, W., MATTHIES, H. & SCHRATTENHOLZ, A. 2003. Analysis of relative isotopologue abundances for quantitative profiling of complex protein mixtures labelled with the acrylamide/D3-acrylamide alkylation tag system. *Rapid Commun Mass Spectrom*, 17, 1283-1290.
- CALDERON, C. E., TIENDA, S., HEREDIA-PONCE, Z., ARREBOLA, E., CARCAMO-OYARCE, G., EBERL, L. & CAZORLA, F. M. 2019. The Compound 2-Hexyl, 5-Propyl Resorcinol Has a Key Role in Biofilm Formation by the Biocontrol Rhizobacterium *Pseudomonas chlororaphis* PCL1606. *Front Microbiol*, 10, 396.
- CAMPBELL, W., HAWS, B., ASAY, K. & HANSEN, J. 1984. A review of black grass bug resistance in forage grasses [*Labops hesperius*, *irbisia brachycera*]. *Rangeland Ecology & Management/Journal of Range Management Archives*, 37, 365-369.
- CANARINI, A., KAISER, C., MERCHANT, A., RICHTER, A. & WANEK, W. 2019. Root Exudation of Primary Metabolites: Mechanisms and Their Roles in Plant Responses to Environmental Stimuli. *Front Plant Sci*, 10, 157.
- CARMONA, C., LANGAN, P., SMITH, J. C. & PETRIDIS, L. 2015. Why genetic modification of lignin leads to low-recalcitrance biomass. *Physical Chemistry Chemical Physic*, 17, 358-64.
- CARPER, D. L., WESTON, D. J., BARDE, A., TIMM, C. M., LU, T. Y., BURDICK, L. H., JAWDY, S. S., KLINGEMAN, D. M., ROBESON, M. S., 2ND, VEACH, A. M., CREGGER, M. A., KALLURI, U. C., SCHADT, C. W., PODAR, M., DOKTYCZ, M. J. & PELLETIER, D. A. 2021. Cultivating the Bacterial Microbiota of Populus Roots. *mSystems*, e0130620.
- CARR, J. F., GREGORY, S. T. & DAHLBERG, A. E. 2005. Severity of the streptomycin resistance and streptomycin dependence phenotypes of ribosomal protein S12 of *Thermus thermophilus* depends on the identity of highly conserved amino acid residues. *J Bacteriol*, 187, 3548-50.
- CASCALES, E. 2008. The type VI secretion toolkit. *EMBO Rep*, 9, 735-41.
- CERDAN, P. D. & CHORY, J. 2003. Regulation of flowering time by light quality. *Nature*, 423, 881-5.
- CHAIT, B. T. 2006. Chemistry. Mass spectrometry: bottom-up or top-down? *Science*, 314, 65-6.

- CHAIWONGSAR, S., STROHM, A. K., SU, S. H. & KRYSAN, P. J. 2012. Genetic analysis of the Arabidopsis protein kinases MAP3Kepsilon1 and MAP3Kepsilon2 indicates roles in cell expansion and embryo development. *Front Plant Sci*, 3, 228.
- CHAKRABORTY, J., GHOSH, P., SEN, S., NANDI, A. K. & DAS, S. 2019. CaMPK9 increases the stability of CaWRKY40 transcription factor which triggers defense response in chickpea upon *Fusarium oxysporum* f. sp. *ciceri* Race1 infection. *Plant Mol Biol*, 100, 411-431.
- CHASTAIN, C. J., XU, W., PARSLEY, K., SARATH, G., HIBBERD, J. M. & CHOLLET, R. 2008. The pyruvate, orthophosphate dikinase regulatory proteins of Arabidopsis possess a novel, unprecedented Ser/Thr protein kinase primary structure. *Plant J*, 53, 854-63.
- CHATNAPARAT, T., PRATHUANGWONG, S., IONESCU, M. & LINDOW, S. E. 2012. XagR, a LuxR homolog, contributes to the virulence of *Xanthomonas axonopodis* pv. *glycines* to soybean. *Mol Plant Microbe Interact*, 25, 1104-17.
- CHEN, F. & DIXON, R. A. 2007. Lignin modification improves fermentable sugar yields for biofuel production. *Nat Biotechnol*, 25, 759-61.
- CHEN, H., LAI, Z., SHI, J., XIAO, Y., CHEN, Z. & XU, X. 2010. Roles of arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. *BMC Plant Biol*, 10, 281.
- CHEN, I. A., CHU, K., PALANIAPPAN, K., RATNER, A., HUANG, J., HUNTEMANN, M., HAJEK, P., RITTER, S., VARGHESE, N., SESHADRI, R., ROUX, S., WOYKE, T., ELOE-FADROSH, E. A., IVANOVA, N. N. & KYRPIDES, N. C. 2021. The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res*, 49, D751-D763.
- CHEN, X., SHANG, J., CHEN, D., LEI, C., ZOU, Y., ZHAI, W., LIU, G., XU, J., LING, Z., CAO, G., MA, B., WANG, Y., ZHAO, X., LI, S. & ZHU, L. 2006. A B-lectin receptor kinase gene conferring rice blast resistance. *Plant J*, 46, 794-804.
- CHENG, X., WU, Y., GUO, J., DU, B., CHEN, R., ZHU, L. & HE, G. 2013. A rice lectin receptor-like kinase that is involved in innate immune responses also contributes to seed germination. *Plant J*, 76, 687-98.
- CHENG, Y., YAM, J. K. H., CAI, Z., DING, Y., ZHANG, L. H., DENG, Y. & YANG, L. 2019. Population dynamics and transcriptomic responses of *Pseudomonas aeruginosa* in a complex laboratory microbial community. *NPJ Biofilms Microbiomes*, 5, 1.
- COHEN, P. 2000. The regulation of protein function by multisite phosphorylation--a 25 year update. *Trends Biochem Sci*, 25, 596-601.

- COLE, S. J. & DIENER, A. C. 2013. Diversity in receptor-like kinase genes is a major determinant of quantitative resistance to *Fusarium oxysporum* f.sp. *matthioli*. *New Phytol*, 200, 172-184.
- CONSIDINE, M. J. & FOYER, C. H. 2014. Redox regulation of plant development. *Antioxid Redox Signal*, 21, 1305-26.
- COOK, R. J., THOMASHOW, L. S., WELLER, D. M., FUJIMOTO, D., MAZZOLA, M., BANGERA, G. & KIM, D. S. 1995. Molecular mechanisms of defense by rhizobacteria against root disease. *Proc Natl Acad Sci U S A*, 92, 4197-201.
- CORNELIS, P. 2010. Iron uptake and metabolism in pseudomonads. *Appl Microbiol Biotechnol*, 86, 1637-45.
- CORNELIS, P. & MATTHIJS, S. 2002. Diversity of siderophore-mediated iron uptake systems in fluorescent pseudomonads: not only pyoverdines. *Environ Microbiol*, 4, 787-98.
- COSIO, C., RANOCHA, P., FRANCOZ, E., BURLAT, V., ZHENG, Y., PERRY, S. E., RIPOLL, J. J., YANOFSKY, M. & DUNAND, C. 2017. The class III peroxidase PRX17 is a direct target of the MADS-box transcription factor AGAMOUS-LIKE15 (AGL15) and participates in lignified tissue formation. *New Phytol*, 213, 250-263.
- COUTINHO, B. G., MEVERS, E., SCHAEFER, A. L., PELLETIER, D. A., HARWOOD, C. S., CLARDY, J. & GREENBERG, E. P. 2018. A plant-responsive bacterial-signaling system senses an ethanolamine derivative. *Proc Natl Acad Sci U S A*, 115, 9785-9790.
- CREASY, D. M. & COTTRELL, J. S. 2004. Unimod: Protein modifications for mass spectrometry. *Proteomics*, 4, 1534-6.
- CREGGER, M. A., VEACH, A. M., YANG, Z. K., CROUCH, M. J., VILGALYS, R., TUSKAN, G. A. & SCHADT, C. W. 2018. The *Populus* holobiont: dissecting the effects of plant niches and genotype on the microbiome. *Microbiome*, 6, 31.
- DAKHIYA, Y., HUSSIN, D., FRIDMAN, E., KIFLAWI, M. & GREEN, R. 2017. Correlations between Circadian Rhythms and Growth in Challenging Environments. *Plant Physiol*, 173, 1724-1734.
- DAVIN, L. B., JOURDES, M., PATTEN, A. M., KIM, K. W., VASSAO, D. G. & LEWIS, N. G. 2008. Dissection of lignin macromolecular configuration and assembly: comparison to related biochemical processes in allyl/propenyl phenol and lignan biosynthesis. *Nat Prod Rep*, 25, 1015-90.
- DAVLETOVA, S., RIZHSKY, L., LIANG, H., SHENGQIANG, Z., OLIVER, D. J., COUTU, J., SHULAEV, V., SCHLAUCH, K. & MITTLER, R. 2005. Cytosolic ascorbate peroxidase 1 is a central component of the reactive oxygen gene network of *Arabidopsis*. *Plant Cell*, 17, 268-81.

- DE HOFF, P. L., BRILL, L. M. & HIRSCH, A. M. 2009. Plant lectins: the ties that bind in root symbiosis and plant defense. *Mol Genet Genomics*, 282, 1-15.
- DE HOFFMANN, E. & STROOBANT, V. 2007. *Mass spectrometry: principles and applications*, John Wiley & Sons.
- DE MEESTER, B., DE VRIES, L., OZPARPUCU, M., GIERLINGER, N., CORNEILLIE, S., PALLIDIS, A., GOEMINNE, G., MORREEL, K., DE BRUYNE, M., DE RYCKE, R., VANHOLME, R. & BOERJAN, W. 2018. Vessel-Specific Reintroduction of CINNAMOYL-COA REDUCTASE1 (CCR1) in Dwarfed *ccr1* Mutants Restores Vessel and Xylary Fiber Integrity and Increases Biomass. *Plant Physiol*, 176, 611-633.
- DE WEERT, S., VERMEIREN, H., MULDER, I. H., KUIPER, I., HENDRICKX, N., BLOEMBERG, G. V., VANDERLEYDEN, J., DE MOT, R. & LUGTENBERG, B. J. 2002. Flagella-driven chemotaxis towards exudate components is an important trait for tomato root colonization by *Pseudomonas fluorescens*. *Mol Plant Microbe Interact*, 15, 1173-80.
- DE ZELICOURT, A., AL-YOUSIF, M., & HIRT, H. 2013. Rhizosphere microbes as essential partners for plant stress tolerance. *Molecular plant*, 6(2), 242-245.
- DEB, S., SANKARANARAYANAN, S., WEWALA, G., WIDDUP, E. & SAMUEL, M. A. 2014. The S-Domain Receptor Kinase Arabidopsis Receptor Kinase2 and the U Box/Armadillo Repeat-Containing E3 Ubiquitin Ligase9 Module Mediates Lateral Root Development under Phosphate Starvation in Arabidopsis. *Plant Physiol*, 165, 1647-1656.
- DELAHUNTY, C. M. & YATES, J. R., 3RD 2007. MudPIT: multidimensional protein identification technology. *Biotechniques*, 43, 563, 565, 567 passim.
- DELESSERT, C., WILSON, I. W., VAN DER STRAETEN, D., DENNIS, E. S. & DOLFERUS, R. 2004. Spatial and temporal analysis of the local response to wounding in Arabidopsis leaves. *Plant Mol Biol*, 55, 165-81.
- DELOM, F. & CHEVET, E. 2006. Phosphoprotein analysis: from proteins to proteomes. *Proteome Sci*, 4, 15.
- DEN RIDDER, M., DARAN-LAPUJADE, P. & PABST, M. 2020. Shot-gun proteomics: why thousands of unidentified signals matter. *FEMS Yeast Res*, 20.
- DES MARAIS, D. L., HERNANDEZ, K. M. & JUENGER, T. E. 2013. Genotype-by-environment interaction and plasticity: exploring genomic responses of plants to the abiotic environment. *Annu Rev Ecol Evol Syst*, 44, 5-29.
- DEVABHAKTUNI, A., LIN, S., ZHANG, L., SWAMINATHAN, K., GONZALEZ, C. G., OLSSON, N., PEARLMAN, S. M., RAWSON, K. & ELIAS, J. E. 2019. TagGraph reveals

- vast protein modification landscapes from large tandem mass spectrometry datasets. *Nat Biotechnol*, 37, 469-479.
- DEZNABI, I., ARABACI, B., KOYUTURK, M. & TASTAN, O. 2020. DeepKinZero: zero-shot learning for predicting kinase-phosphosite associations involving understudied kinases. *Bioinformatics*, 36, 3652-3661.
- DIXON, R. A. & BARROS, J. 2019. Lignin biosynthesis: old roads revisited and new roads explored. *Open biology*, 9, 190215.
- DO, C. T., POLLET, B., THEVENIN, J., SIBOUT, R., DENOUE, D., BARRIERE, Y., LAPIERRE, C. & JOUANIN, L. 2007. Both caffeoyl Coenzyme A 3-O-methyltransferase 1 and caffeic acid O-methyltransferase 1 are involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in Arabidopsis. *Planta*, 226, 1117-29.
- DOERR, A. 2015. DIA mass spectrometry. *Nature Methods*, 12, 35-35.
- DOMANOVA, W., KRYCER, J., CHAUDHURI, R., YANG, P., VAFAGE, F., FAZAKERLEY, D., HUMPHREY, S., JAMES, D. & KUNCIC, Z. 2016. Unraveling Kinase Activation Dynamics Using Kinase-Substrate Relationships from Temporal Large-Scale Phosphoproteomics Studies. *PLoS One*, 11, e0157763.
- DONCHEVA, N. T., MORRIS, J. H., GORODKIN, J. & JENSEN, L. J. 2019. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J Proteome Res*, 18, 623-632.
- DORFER, V., PICHLER, P., STRANZL, T., STADLMANN, J., TAUS, T., WINKLER, S. & MECHTLER, K. 2014. MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J Proteome Res*, 13, 3679-84.
- DU, Z., ZHOU, X., LING, Y., ZHANG, Z. & SU, Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res*, 38, W64-70.
- DUAN, G. & WALTHER, D. 2015. The roles of post-translational modifications in the context of protein interaction networks. *PLoS Comput Biol*, 11, e1004049.
- DUPREE, E. J., JAYATHIRTHA, M., YORKEY, H., MIHASAN, M., PETRE, B. A. & DARIE, C. C. 2020. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of this Field. *Proteomes*, 8.
- EDELMANN, M. J. 2011. Strong cation exchange chromatography in analysis of posttranslational modifications: innovations and perspectives. *J Biomed Biotechnol*, 2011, 936508.
- EDGE, R., MCGARVEY, D. J., & TRUSCOTT, T. G. 1997. The carotenoids as anti-oxidants—a review. . *Journal of Photochemistry and Photobiology B: Biology*, 41(3), 189-200.

- EDWARDS, J., JOHNSON, C., SANTOS-MEDELLIN, C., LURIE, E., PODISHETTY, N. K., BHATNAGAR, S., EISEN, J. A. & SUNDARESAN, V. 2015. Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc Natl Acad Sci U S A*, 112, E911-20.
- EL HOUARI, I., BOERJAN, W. & VANHOLME, B. 2021. Behind the Scenes: The Impact of Bioactive Phenylpropanoids on the Growth Phenotypes of Arabidopsis Lignin Mutants. *Front Plant Sci*, 12, 734070.
- ENDRES, R. G. & WINGREEN, N. S. 2006. Precise adaptation in bacterial chemotaxis through "assistance neighborhoods". *Proc Natl Acad Sci U S A*, 103, 13040-4.
- ENG, J. K., MCCORMACK, A. L. & YATES, J. R. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom*, 5, 976-89.
- ENGSTROM, P. & HAZELBAUER, G. L. 1980. Multiple methylation of methyl-accepting chemotaxis proteins during adaptation of *E. coli* to chemical stimuli. *Cell*, 20, 165-71.
- ERCE, M. A., PANG, C. N., HART-SMITH, G. & WILKINS, M. R. 2012. The methylproteome and the intracellular methylation network. *Proteomics*, 12, 564-86.
- ESCAMILLA-TREVINO, L. L., SHEN, H., UPPALAPATI, S. R., RAY, T., TANG, Y., HERNANDEZ, T., YIN, Y., XU, Y. & DIXON, R. A. 2010. Switchgrass (*Panicum virgatum*) possesses a divergent family of cinnamoyl CoA reductases with distinct biochemical properties. *New Phytol*, 185, 143-55.
- FAHAD, S., HUSSAIN, S., BANO, A., SAUD, S., HASSAN, S., SHAN, D., KHAN, F. A., KHAN, F., CHEN, Y., WU, C., TABASSUM, M. A., CHUN, M. X., AFZAL, M., JAN, A., JAN, M. T. & HUANG, J. 2015. Potential role of phytohormones and plant growth-promoting rhizobacteria in abiotic stresses: consequences for changing environment. *Environ Sci Pollut Res Int*, 22, 4907-21.
- FAN, L., LINKER, R., GEPSTEIN, S., TANIMOTO, E., YAMAMOTO, R. & NEUMANN, P. M. 2006. Progressive inhibition by water deficit of cell wall extensibility and growth along the elongation zone of maize roots is related to increased lignin metabolism and progressive stelar accumulation of wall phenolics. *Plant Physiol*, 140, 603-12.
- FEIST, P. & HUMMON, A. B. 2015. Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int J Mol Sci*, 16, 3537-63.
- FENN, J. B., MANN, M., MENG, C. K., WONG, S. F. & WHITEHOUSE, C. M. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246, 64-71.
- FENYO, D. & BEAVIS, R. C. 2015. The GPMDB REST interface. *Bioinformatics*, 31, 2056-8.

- FERNANDEZ-PEREZ, F., POMAR, F., PEDRENO, M. A. & NOVO-UZAL, E. 2015. The suppression of AtPrx52 affects fibers but not xylem lignification in Arabidopsis by altering the proportion of syringyl units. *Physiol Plant*, 154, 395-406.
- FICARRO, S. B., MCCLELAND, M. L., STUKENBERG, P. T., BURKE, D. J., ROSS, M. M., SHABANOWITZ, J., HUNT, D. F. & WHITE, F. M. 2002. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol*, 20, 301-5.
- FICHMAN, Y., MILLER, G. & MITTLER, R. 2019. Whole-Plant Live Imaging of Reactive Oxygen Species. *Mol Plant*, 12, 1203-1210.
- FICHMAN, Y., ZANDALINAS, S. I., PECK, S., LUAN, S. & MITTLER, R. 2022. HPCA1 is required for systemic reactive oxygen species and calcium cell-to-cell signaling and plant acclimation to stress. *Plant Cell*.
- FILLOUX, A., HACHANI, A. & BLEVES, S. 2008. The bacterial type VI secretion machine: yet another player for protein transport across membranes. *Microbiology (Reading)*, 154, 1570-1583.
- FORNELLI, L., TOBY, T. K., SCHACHNER, L. F., DOUBLEDAY, P. F., SRZENTIC, K., DEHART, C. J. & KELLEHER, N. L. 2018. Top-down proteomics: Where we are, where we are going? *J Proteomics*, 175, 3-4.
- FOYER, C. H. 2018. Reactive oxygen species, oxidative signaling and the regulation of photosynthesis. *Environ Exp Bot*, 154, 134-142.
- FRANKE, R., HEMM, M. R., DENAULT, J. W., RUEGGER, M. O., HUMPHREYS, J. M. & CHAPPLE, C. 2002. Changes in secondary metabolism and deposition of an unusual lignin in the ref8 mutant of Arabidopsis. *Plant J*, 30, 47-59.
- FRS, L. 1882. On the equilibrium of liquid conducting masses charged with electricity', Lond. *Edinb. Dublin Philos. Mag. J. Sci*, 14, 87.
- FUJIOKA, S. & YOKOTA, T. 2003. Biosynthesis and metabolism of brassinosteroids. *Annu Rev Plant Biol*, 54, 137-64.
- FUQUA, C., PARSEK, M. R. & GREENBERG, E. P. 2001. Regulation of gene expression by cell-to-cell communication: acyl-homoserine lactone quorum sensing. *Annu Rev Genet*, 35, 439-68.
- FURANO, A. V. 1975. Content of Elongation-Factor Tu in Escherichia-Coli. *Proceedings of the National Academy of Sciences of the United States of America*, 72, 4780-4784.

- GAIERO, J. R., MCCALL, C. A., THOMPSON, K. A., DAY, N. J., BEST, A. S. & DUNFIELD, K. E. 2013. Inside the root microbiome: bacterial root endophytes and plant growth promotion. *Am J Bot*, 100, 1738-50.
- GALLEGO-GIRALDO, L., JIKUMARU, Y., KAMIYA, Y., TANG, Y. & DIXON, R. A. 2011. Selective lignin downregulation leads to constitutive defense response expression in alfalfa (*Medicago sativa* L.). *New Phytol*, 190, 627-39.
- GALLEGO-GIRALDO, L., LIU, C., POSE-ALBACETE, S., PATTATHIL, S., PERALTA, A. G., YOUNG, J., WESTPHELING, J., HAHN, M. G., RAO, X., KNOX, J. P., DE MEESTER, B., BOERJAN, W. & DIXON, R. A. 2020. ARABIDOPSIS DEHISCENCE ZONE POLYGALACTURONASE 1 (ADPG1) releases latent defense signals in stems with reduced lignin content. *Proc Natl Acad Sci U S A*, 117, 3281-3290.
- GALLIE, D. R. 2013. The role of L-ascorbic acid recycling in responding to environmental stress and in promoting plant growth. *J Exp Bot*, 64, 433-43.
- GAN, S. & AMASINO, R. M. 1995. Inhibition of leaf senescence by autoregulated production of cytokinin. *Science*, 270, 1986-8.
- GAN, S. & AMASINO, R. M. 1997. Making Sense of Senescence (Molecular Genetic Regulation and Manipulation of Leaf Senescence). *Plant Physiol*, 113, 313-319.
- GARCIA-MEDEL, P. L., BARUCH-TORRES, N., PERALTA-CASTRO, A., TRASVINA-ARENAS, C. H., TORRES-LARIOS, A. & BRIEBA, L. G. 2019. Plant organellar DNA polymerases repair double-stranded breaks by microhomology-mediated end-joining. *Nucleic Acids Res*, 47, 3028-3044.
- GE, W., WOLF, A., FENG, T., HO, C. H., SEKIRNIK, R., ZAYER, A., GRANATINO, N., COCKMAN, M. E., LOENARZ, C., LOIK, N. D., HARDY, A. P., CLARIDGE, T. D. W., HAMED, R. B., CHOWDHURY, R., GONG, L., ROBINSON, C. V., TRUDGIAN, D. C., JIANG, M., MACKEEN, M. M., MCCULLAGH, J. S., GORDIYENKO, Y., THALHAMMER, A., YAMAMOTO, A., YANG, M., LIU-YI, P., ZHANG, Z., SCHMIDT-ZACHMANN, M., KESSLER, B. M., RATCLIFFE, P. J., PRESTON, G. M., COLEMAN, M. L. & SCHOFIELD, C. J. 2012. Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat Chem Biol*, 8, 960-962.
- GEHLENBORG, N., O'DONOGHUE, S. I., BALIGA, N. S., GOESMANN, A., HIBBS, M. A., KITANO, H., KOHLBACHER, O., NEUWEGER, H., SCHNEIDER, R. & TENENBAUM, D. 2010. Visualization of omics data for systems biology. *Nature methods*, 7, S56-S68.
- GILARDONI, P. A., HETTENHAUSEN, C., BALDWIN, I. T. & BONAVENTURE, G. 2011. *Nicotiana attenuata* LECTIN RECEPTOR KINASE1 suppresses the insect-mediated inhibition of induced defense responses during *Manduca sexta* herbivory. *Plant Cell*, 23, 3512-32.

- GOBBATO, E., MARSH, J. F., VERNIE, T., WANG, E., MAILLET, F., KIM, J., MILLER, J. B., SUN, J., BANO, S. A., RATET, P., MYSORE, K. S., DENARIE, J., SCHULTZE, M. & OLDROYD, G. E. 2012. A GRAS-type transcription factor with a specific function in mycorrhizal signaling. *Curr Biol*, 22, 2236-41.
- GOHLKE, R. S. 1959. Time-of-flight mass spectrometry and gas-liquid partition chromatography. *Analytical Chemistry*, 31, 535-541.
- GONZALEZ, E., PITRE, F. E., PAGE, A. P., MARLEAU, J., GUIDI NISSIM, W., ST-ARNAUD, M., LABRECQUE, M., JOLY, S., YERGEAU, E. & BRERETON, N. J. B. 2018. Trees, fungi and bacteria: tripartite metatranscriptomics of a root microbiome responding to soil contamination. *Microbiome*, 6, 53.
- GONZALEZ, J. F. & VENTURI, V. 2013. A novel widespread interkingdom signaling circuit. *Trends Plant Sci*, 18, 167-74.
- GOODSTEIN, D. M., SHU, S., HOWSON, R., NEUPANE, R., HAYES, R. D., FAZO, J., MITROS, T., DIRKS, W., HELLSTEN, U., PUTNAM, N. & ROKHSAR, D. S. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*, 40, D1178-86.
- GOUDA, S., KERRY, R. G., DAS, G., PARAMITHIOTIS, S., SHIN, H. S. & PATRA, J. K. 2018. Revitalization of plant growth promoting rhizobacteria for sustainable development in agriculture. *Microbiol Res*, 206, 131-140.
- GOYAL, A., WANG, T., DUBINKINA, V. & MASLOV, S. 2021. Ecology-guided prediction of cross-feeding interactions in the human gut microbiome. *Nat Commun*, 12, 1335.
- GRANDCHAMP, G. M., CARO, L. & SHANK, E. A. 2017. Pirated Siderophores Promote Sporulation in *Bacillus subtilis*. *Appl Environ Microbiol*, 83.
- GRAVES, P. R. & HAYSTEAD, T. A. 2002. Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev*, 66, 39-63; table of contents.
- GRIFFITHS, J. 2008. A brief history of mass spectrometry. *Anal Chem*, 80, 5678-83.
- GUIDARELLI, M., ZOLI, L., ORLANDINI, A., BERTOLINI, P. & BARALDI, E. 2014. The mannose-binding lectin gene FaMBL1 is involved in the resistance of unripe strawberry fruits to *Colletotrichum acutatum*. *Mol Plant Pathol*, 15, 832-40.
- GUO, Y. R., G.; ZHANG, K.; LI, Z., MIAO, Y.; & GUO, H. 2021. Leaf senescence: Progression, regulation, and application. *Molecular Horticulture*, 1(1), 1-25.
- GUPTA, N., TANNER, S., JAITLY, N., ADKINS, J. N., LIPTON, M., EDWARDS, R., ROMINE, M., OSTERMAN, A., BAFNA, V., SMITH, R. D. & PEVZNER, P. A. 2007.

- Whole proteome analysis of post-translational modifications: applications of mass-spectrometry for proteogenomic annotation. *Genome Res*, 17, 1362-77.
- HA, C. M., ESCAMILLA-TREVINO, L., YARCE, J. C., KIM, H., RALPH, J., CHEN, F. & DIXON, R. A. 2016. An essential role of caffeoyl shikimate esterase in monolignol biosynthesis in *Medicago truncatula*. *Plant J*, 86, 363-75.
- HA, C. M., RAO, X., SAXENA, G. & DIXON, R. A. 2021. Growth-defense trade-offs and yield loss in plants with engineered cell walls. *New Phytol*, 231, 60-74.
- HAMANN, T. 2015. The plant cell wall integrity maintenance mechanism--a case study of a cell wall plasma membrane signaling network. *Phytochemistry*, 112, 100-9.
- HAMEL, L. P., NICOLE, M. C., DUPLESSIS, S. & ELLIS, B. E. 2012. Mitogen-activated protein kinase signaling in plant-interacting fungi: distinct messages from conserved messengers. *Plant Cell*, 24, 1327-51.
- HAN, G., YE, M. & ZOU, H. 2008. Development of phosphopeptide enrichment techniques for phosphoproteome analysis. *Analyst*, 133, 1128-38.
- HAN, X., HE, L., XIN, L., SHAN, B. & MA, B. 2011. PeaksPTM: Mass spectrometry-based identification of peptides with unspecified modifications. *J Proteome Res*, 10, 2930-6.
- HARDMAN, M. & MAKAROV, A. A. 2003. Interfacing the orbitrap mass analyzer to an electrospray ion source. *Anal Chem*, 75, 1699-705.
- HART-SMITH, G. & BLANKSBY, S. J. 2012. Mass analysis. *Mass spectrometry in polymer chemistry*, 5-32.
- HARVEY, K. L., JAROCKI, V. M., CHARLES, I. G. & DJORDJEVIC, S. P. 2019. The Diverse Functional Roles of Elongation Factor Tu (EF-Tu) in Microbial Pathogenesis. *Front Microbiol*, 10, 2351.
- HASIM, S., ALLISON, D. P., MENDEZ, B., FARMER, A. T., PELLETIER, D. A., RETTERER, S. T., CAMPAGNA, S. R., REYNOLDS, T. B. & DOKTYCZ, M. J. 2018. Elucidating Duramycin's Bacterial Selectivity and Mode of Action on the Bacterial Cell Envelope. *Front Microbiol*, 9, 219.
- HEIN, K. R. 2005. Future energy supply in Europe—challenge and chances. *Fuel*, 84, 1189-1194.
- HIGGINS, D. & DWORKIN, J. 2012. Recent progress in *Bacillus subtilis* sporulation. *FEMS Microbiol Rev*, 36, 131-48.
- HILAL, M., PARRADO, M. F., ROSA, M., GALLARDO, M., ORCE, L., MASSA, E. M., GONZALEZ, J. A. & PRADO, F. E. 2004. Epidermal lignin deposition in quinoa cotyledons in response to UV-B radiations. *Photochem Photobiol*, 79, 205-10.

- HINSINGER, P. & MARSCHNER, P. 2006. Rhizosphere--perspectives and challenges--a tribute to Lorenz Hiltner 12-17 September 2004--Munich, Germany. *Plant and Soil*, 283.
- HIRUMA, K., NISHIUCHI, T., KATO, T., BEDNAREK, P., OKUNO, T., SCHULZE-LEFERT, P. & TAKANO, Y. 2011. Arabidopsis ENHANCED DISEASE RESISTANCE 1 is required for pathogen-induced expression of plant defensins in nonhost resistance, and acts through interference of MYC2-mediated repressor function. *Plant J*, 67, 980-92.
- HOFFMANN, L., BESSEAU, S., GEOFFROY, P., RITZENTHALER, C., MEYER, D., LAPIERRE, C., POLLET, B. & LEGRAND, M. 2004. Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinic acid hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell*, 16, 1446-65.
- HOFFMANN, N., BENSKE, A., BETZ, H., SCHUETZ, M. & SAMUELS, A. L. 2020. Laccases and Peroxidases Co-Localize in Lignified Secondary Cell Walls throughout Stem Development. *Plant Physiol*, 184, 806-822.
- HOUTZ, R. L., MAGNANI, R., NAYAK, N. R. & DIRK, L. M. 2008. Co- and post-translational modifications in Rubisco: unanswered questions. *J Exp Bot*, 59, 1635-45.
- HSIA, M. M., O'MALLEY, R., CARTWRIGHT, A., NIEU, R., GORDON, S. P., KELLY, S., WILLIAMS, T. G., WOOD, D. F., ZHAO, Y., BRAGG, J., JORDAN, M., PAULY, M., ECKER, J. R., GU, Y. & VOGEL, J. P. 2017. Sequencing and functional validation of the JGI Brachypodium distachyon T-DNA collection. *Plant J*, 91, 361-370.
- HU, A., NOBLE, W. S. & WOLF-YADLIN, A. 2016. Technical advances in proteomics: new developments in data-independent acquisition. *F1000Res*, 5.
- HU, Q., NOLL, R. J., LI, H., MAKAROV, A., HARDMAN, M. & GRAHAM COOKS, R. 2005. The Orbitrap: a new mass spectrometer. *J Mass Spectrom*, 40, 430-43.
- HUOT, B., YAO, J., MONTGOMERY, B. L. & HE, S. Y. 2014. Growth-defense tradeoffs in plants: a balancing act to optimize fitness. *Mol Plant*, 7, 1267-1287.
- IAVICOLI, A., BOUTET, E., BUCHALA, A. & METRAUX, J. P. 2003. Induced systemic resistance in Arabidopsis thaliana in response to root inoculation with Pseudomonas fluorescens CHA0. *Mol Plant Microbe Interact*, 16, 851-8.
- IMURA, Y., KOBAYASHI, Y., YAMAMOTO, S., FURUTANI, M., TASAKA, M., ABE, M. & ARAKI, T. 2012. CRYPTIC PRECOCIOUS/MED12 is a novel flowering regulator with multiple target steps in Arabidopsis. *Plant Cell Physiol*, 53, 287-303.
- ISAACSON, T., DAMASCENO, C. M., SARAVANAN, R. S., HE, Y., CATALA, C., SALADIE, M. & ROSE, J. K. 2006. Sample extraction techniques for enhanced proteomic analysis of plant tissues. *Nat Protoc*, 1, 769-74.

- ISSAQ, H. J., CHAN, K. C., JANINI, G. M., CONRAD, T. P. & VEENSTRA, T. D. 2005. Multidimensional separation of peptides for effective proteomic analysis. *J Chromatogr B Analyt Technol Biomed Life Sci*, 817, 35-47.
- JACOBS, M. & RUBERY, P. H. 1988. Naturally occurring auxin transport regulators. *Science*, 241, 346-9.
- JALMI, S. K. & SINHA, A. K. 2015. ROS mediated MAPK signaling in abiotic and biotic stress-striking similarities and differences. *Front Plant Sci*, 6, 769.
- JANI, A. J. & COTTER, P. A. 2010. Type VI secretion: not just for pathogenesis anymore. *Cell Host Microbe*, 8, 2-6.
- JIANG, L., HE, L. & FOUNTOULAKIS, M. 2004. Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. *J Chromatogr A*, 1023, 317-20.
- JING, Y. & LIN, R. 2015. The VQ Motif-Containing Protein Family of Plant-Specific Transcriptional Regulators. *Plant Physiol*, 169, 371-8.
- JONES, D. & HINSINGER, P. 2008. The rhizosphere: complex by design. Springer.
- JONES, L., ENNOS, A. R. & TURNER, S. R. 2001. Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of Arabidopsis. *Plant J*, 26, 205-16.
- JOSEPH, T. A., SHENHAV, L., XAVIER, J. B., HALPERIN, E. & PE'ER, I. 2020. Compositional Lotka-Volterra describes microbial dynamics in the simplex. *PLoS Comput Biol*, 16, e1007917.
- JOSHI, R., GANGABHAGIRATHI, R., VENU, S., ADHIKARI, S. & MUKHERJEE, T. 2012. Antioxidant activity and free radical scavenging reactions of gentisic acid: in-vitro and pulse radiolysis studies. *Free Radic Res*, 46, 11-20.
- JOSHI, S., CHOUKIMATH, A., ISENEGGER, D., PANOZZO, J., SPANGENBERG, G. & KANT, S. 2019. Improved Wheat Growth and Yield by Delayed Leaf Senescence Using Developmentally Regulated Expression of a Cytokinin Biosynthesis Gene. *Front Plant Sci*, 10, 1285.
- KALL, L., CANTERBURY, J. D., WESTON, J., NOBLE, W. S. & MACCOSS, M. J. 2007. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods*, 4, 923-5.
- KAMATH, K. S., VASAVADA, M. S. & SRIVASTAVA, S. 2011. Proteomic databases and tools to decipher post-translational modifications. *J Proteomics*, 75, 127-44.
- KANDEL, S. L., JOUBERT, P. M. & DOTY, S. L. 2017. Bacterial endophyte colonization and distribution within plants. *Microorganisms*, 5, 77.

- KANSIN, E., GIGUERE, S., JING, C., TYERS, M. & THIBAUT, P. 2017. Machine Learning of Global Phosphoproteomic Profiles Enables Discrimination of Direct versus Indirect Kinase Substrates. *Mol Cell Proteomics*, 16, 786-798.
- KARAS, M., BACHMANN, D., BAHR, U. & HILLENKAMP, F. 1987. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *International journal of mass spectrometry and ion processes*, 78, 53-68.
- KARAS, M. & HILLENKAMP, F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical chemistry*, 60, 2299-2301.
- KARASOV, T. L., CHAE, E., HERMAN, J. J. & BERGELSON, J. 2017. Mechanisms to Mitigate the Trade-Off between Growth and Defense. *Plant Cell*, 29, 666-680.
- KARP, A. & SHIELD, I. 2008. Bioenergy from plants and the sustainable yield challenge. *New Phytol*, 179, 15-32.
- KARPIEVITCH, Y. V., POLPITIYA, A. D., ANDERSON, G. A., SMITH, R. D. & DABNEY, A. R. 2010. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. *Ann Appl Stat*, 4, 1797-1823.
- KEBARLE, P. & VERKERK, U. H. 2009. Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom Rev*, 28, 898-917.
- KESHAVAN, N. D., CHOWDHARY, P. K., HAINES, D. C. & GONZALEZ, J. E. 2005. L-Canavanine made by *Medicago sativa* interferes with quorum sensing in *Sinorhizobium meliloti*. *J Bacteriol*, 187, 8427-36.
- KIDD, B. N., EDGAR, C. I., KUMAR, K. K., AITKEN, E. A., SCHENK, P. M., MANNERS, J. M. & KAZAN, K. 2009. The mediator complex subunit PFT1 is a key regulator of jasmonate-dependent defense in Arabidopsis. *Plant Cell*, 21, 2237-52.
- KIM, H., HWANG, H., HONG, J. W., LEE, Y. N., AHN, I. P., YOON, I. S., YOO, S. D., LEE, S., LEE, S. C. & KIM, B. G. 2012. A rice orthologue of the ABA receptor, OsPYL/RCAR5, is a positive regulator of the ABA signal transduction pathway in seed germination and early seedling growth. *J Exp Bot*, 63, 1013-24.
- KIM, H. S., JUNG, M. S., LEE, S. M., KIM, K. E., BYUN, H., CHOI, M. S., PARK, H. C., CHO, M. J. & CHUNG, W. S. 2009. An S-locus receptor-like kinase plays a role as a negative regulator in plant defense responses. *Biochem Biophys Res Commun*, 381, 424-8.
- KIM, J. I., SHEN, Y., HAN, Y. J., PARK, J. E., KIRCHENBAUER, D., SOH, M. S., NAGY, F., SCHAFER, E. & SONG, P. S. 2004. Phytochrome phosphorylation modulates light signaling by influencing the protein-protein interaction. *Plant Cell*, 16, 2629-40.

- KIM, Y. J., BJORKLUND, S., LI, Y., SAYRE, M. H. & KORNBERG, R. D. 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, 77, 599-608.
- KIMURA, S. & SAKAGUCHI, K. 2006. DNA repair in plants. *Chem Rev*, 106, 753-66.
- KINTER, M. & SHERMAN, N. E. 2005. *Protein sequencing and identification using tandem mass spectrometry*, John Wiley & Sons.
- KIRSCHBAUM, M. U. 2011. Does enhanced photosynthesis enhance growth? Lessons learned from CO₂ enrichment studies. *Plant Physiol*, 155, 117-24.
- KISER, J. Z., POST, M., WANG, B. & MIYAGI, M. 2009. *Streptomyces erythraeus* trypsin for proteomics applications. *J Proteome Res*, 8, 1810-7.
- KLEINER, M. 2019. Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities. *mSystems*, 4.
- KLEINER, M., THORSON, E., SHARP, C. E., DONG, X., LIU, D., LI, C. & STROUS, M. 2017. Assessing species biomass contributions in microbial communities via metaproteomics. *Nat Commun*, 8, 1558.
- KLOPPHOLZ, S., KUHN, H. & REQUENA, N. 2011. A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Curr Biol*, 21, 1204-9.
- KONG, K. F., SCHNEPER, L. & MATHEE, K. 2010. Beta-lactam antibiotics: from antibiosis to resistance and bacteriology. *APMIS*, 118, 1-36.
- KORNBERG, R. D. 2007. The molecular basis of eukaryotic transcription. *Proc Natl Acad Sci U S A*, 104, 12955-61.
- KOWALAK, J. A. & WALSH, K. A. 1996. Beta-methylthio-aspartic acid: identification of a novel posttranslational modification in ribosomal protein S12 from *Escherichia coli*. *Protein Sci*, 5, 1625-32.
- KRAAL, B., LIPPMANN, C. & KLEANTHOUS, C. 1999. Translational regulation by modifications of the elongation factor Tu. *Folia Microbiol (Praha)*, 44, 131-41.
- KRAHMER, J., HINDLE, M., PERBY, L.K., NIELSEN, T.H., HALLIDAY, K.J., VAN OOIJEN, G., LE BIHAN, T. AND MILLAR, A.J. 2021. The circadian clock gene circuit controls protein and phosphoprotein rhythms in *Arabidopsis thaliana*. *bioRxiv*, p.760892.
- KRAMER, J., OZKAYA, O. & KUMMERLI, R. 2020. Bacterial siderophores in community and host interactions. *Nat Rev Microbiol*, 18, 152-163.
- KRINSKY, N. I. 1989. Antioxidant functions of carotenoids. *Free Radic Biol Med*, 7, 617-35.

- KUGAN, H. M., REJAB, N. A., SAHRUZAINI, N. A., HARIKRISHNA, J. A., BAISAKH, N. & CHENG, A. 2021. Circadian Rhythms in Legumes: What Do We Know and What Else Should We Explore? *Int J Mol Sci*, 22.
- KUMAR, M., JI, B., ZENGLER, K. & NIELSEN, J. 2019a. Modelling approaches for studying the microbiome. *Nat Microbiol*, 4, 1253-1267.
- KUMAR, S. V., TAYLOR, G., HASIM, S., COLLIER, C. P., FARMER, A. T., CAMPAGNA, S. R., BIBLE, A. N., DOKTYCZ, M. J. & MORRELL-FALVEY, J. 2019b. Loss of carotenoids from membranes of *Pantoea* sp. YR343 results in altered lipid composition and changes in membrane biophysical properties. *Biochim Biophys Acta Biomembr*, 1861, 1338-1345.
- LABBE, J., MUCHERO, W., CZARNECKI, O., WANG, J., WANG, X., BRYAN, A. C., ZHENG, K., YANG, Y., XIE, M., ZHANG, J., WANG, D., MEIDL, P., WANG, H., MORRELL-FALVEY, J. L., COPE, K. R., MAIA, L. G. S., ANE, J. M., MEWALAL, R., JAWDY, S. S., GUNTER, L. E., SCHACKWITZ, W., MARTIN, J., LE TACON, F., LI, T., ZHANG, Z., RANJAN, P., LINDQUIST, E., YANG, X., JACOBSON, D. A., TSCHAPLINSKI, T. J., BARRY, K., SCHMUTZ, J., CHEN, J. G. & TUSKAN, G. A. 2019. Mediation of plant-mycorrhizal interaction by a lectin receptor-like kinase. *Nat Plants*, 5, 676-680.
- LANQUAR, V. & FROMMER, W. B. 2010. Adjusting ammonium uptake via phosphorylation. *Plant Signal Behav*, 5, 736-8.
- LAUVERGEAT, V., LACOMME, C., LACOMBE, E., LASSERRE, E., ROBY, D. & GRIMA-PETTENATI, J. 2001. Two cinnamoyl-CoA reductase (CCR) genes from *Arabidopsis thaliana* are differentially expressed during development and in response to infection with pathogenic bacteria. *Phytochemistry*, 57, 1187-95.
- LAZAR, C., GATTO, L., FERRO, M., BRULEY, C. & BURGER, T. 2016. Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *J Proteome Res*, 15, 1116-25.
- LEE, M. H., JEON, H. S., KIM, S. H., CHUNG, J. H., ROPPOLO, D., LEE, H. J., CHO, H. J., TOBIMATSU, Y., RALPH, J. & PARK, O. K. 2019. Lignin-based barrier restricts pathogens to the infection site and confers resistance in plants. *EMBO J*, 38, e101948.
- LEE, Y., MUNNIK, T. & LEE, Y. 2010. Plant phosphatidylinositol 3-kinase. *Lipid signaling in plants*. Springer.
- LEON, J., GAYUBAS, B. & CASTILLO, M. C. 2020. Valine-Glutamine Proteins in Plant Responses to Oxygen and Nitric Oxide. *Front Plant Sci*, 11, 632678.
- LEVINE, M. & DAVIDSON, E. H. 2005. Gene regulatory networks for development. *Proceedings of the National Academy of Sciences*, 102, 4936-4942.

- LEVY, A., SALAS GONZALEZ, I., MITTELVIEFHAUS, M., CLINGENPEEL, S., HERRERA PAREDES, S., MIAO, J., WANG, K., DEVESCOVI, G., STILLMAN, K., MONTEIRO, F., RANGEL ALVAREZ, B., LUNDBERG, D. S., LU, T. Y., LEBEIS, S., JIN, Z., MCDONALD, M., KLEIN, A. P., FELTCHER, M. E., RIO, T. G., GRANT, S. R., DOTY, S. L., LEY, R. E., ZHAO, B., VENTURI, V., PELLETIER, D. A., VORHOLT, J. A., TRINGE, S. G., WOYKE, T. & DANGL, J. L. 2017. Genomic features of bacterial adaptation to plants. *Nat Genet*, 50, 138-150.
- LI, C., PAN, D., LI, M., WANG, Y., SONG, L., YU, D., ZUO, Y., WANG, K., LIU, Y., WEI, Z., LU, Z., ZHU, L. & SHEN, X. 2021. Aerobactin-Mediated Iron Acquisition Enhances Biofilm Formation, Oxidative Stress Resistance, and Virulence of *Yersinia pseudotuberculosis*. *Front Microbiol*, 12, 699913.
- LI, K. W., GONZALEZ-LOZANO, M. A., KOOPMANS, F. & SMIT, A. B. 2020. Recent Developments in Data Independent Acquisition (DIA) Mass Spectrometry: Application of Quantitative Analysis of the Brain Proteome. *Front Mol Neurosci*, 13, 564446.
- LI, M., PU, Y. & RAGAUSKAS, A. J. 2016. Current Understanding of the Correlation of Lignin Structure with Biomass Recalcitrance. *Frontiers in chemistry*, 4, 45.
- LI, S. & TANG, H. 2016. Computational Methods in Mass Spectrometry-Based Proteomics. *Adv Exp Med Biol*, 939, 63-89.
- LI, X., BONAWITZ, N. D., WENG, J. K. & CHAPPLE, C. 2010. The growth reduction associated with repressed lignin biosynthesis in *Arabidopsis thaliana* is independent of flavonoids. *Plant Cell*, 22, 1620-32.
- LI, Z., PENG, J., WEN, X. & GUO, H. 2013. Ethylene-insensitive3 is a senescence-associated gene that accelerates age-dependent leaf senescence by directly repressing miR164 transcription in *Arabidopsis*. *Plant Cell*, 25, 3311-28.
- LI, Z., WANG, Y., YAO, Q., JUSTICE, N. B., AHN, T. H., XU, D., HETTICH, R. L., BANFIELD, J. F. & PAN, C. 2014. Diverse and divergent protein post-translational modifications in two growth stages of a natural microbial community. *Nat Commun*, 5, 4405.
- LIAAEN-JENSEN, S. & ANDREWES, A. G. 1972. Microbial carotenoids. *Annu Rev Microbiol*, 26, 225-48.
- LIANG, M., DAVIS, E., GARDNER, D., CAI, X. & WU, Y. 2006. Involvement of AtLAC15 in lignin synthesis in seeds and in root elongation of *Arabidopsis*. *Planta*, 224, 1185-96.
- LIEBLER, D. C. 2001. *Introduction to proteomics: tools for the new biology*, Springer Science & Business Media.
- LIM, P. O., KIM, H. J. & NAM, H. G. 2007. Leaf senescence. *Annu Rev Plant Biol*, 58, 115-36.

- LIN, W. H., YE, R., MA, H., XU, Z. H. & XUE, H. W. 2004. DNA chip-based expression profile analysis indicates involvement of the phosphatidylinositol signaling pathway in multiple plant responses to hormone and abiotic treatments. *Cell Res*, 14, 34-45.
- LIU, C., XU, Z. & CHUA, N. H. 1993. Auxin Polar Transport Is Essential for the Establishment of Bilateral Symmetry during Early Plant Embryogenesis. *Plant Cell*, 5, 621-630.
- LIU, L., LIU, X. J., SUN, W., LI, M. X. & GAO, Y. H. 2013. Unrestrictive identification of post-translational modifications in the urine proteome without enrichment. *Proteome Science*, 11.
- LIU, W., ZHOU, X., LI, G., LI, L., KONG, L., WANG, C., ZHANG, H. & XU, J. R. 2011. Multiple plant surface signals are sensed by different mechanisms in the rice blast fungus for appressorium formation. *PLoS Pathog*, 7, e1001261.
- LIU, Y., WU, H., CHEN, H., LIU, Y., HE, J., KANG, H., SUN, Z., PAN, G., WANG, Q., HU, J., ZHOU, F., ZHOU, K., ZHENG, X., REN, Y., CHEN, L., WANG, Y., ZHAO, Z., LIN, Q., WU, F., ZHANG, X., GUO, X., CHENG, X., JIANG, L., WU, C., WANG, H. & WAN, J. 2015. A gene cluster encoding lectin receptor kinases confers broad-spectrum and durable insect resistance in rice. *Nat Biotechnol*, 33, 301-5.
- LONG, S. R. 1989. Rhizobium-legume nodulation: life together in the underground. *Cell*, 56, 203-14.
- LOO, J. A., DEJOHN, D. E., DU, P., STEVENSON, T. I. & OGORZALEK LOO, R. R. 1999. Application of mass spectrometry for target identification and characterization. *Med Res Rev*, 19, 307-19.
- LOPER, J. E., AND BUYER, J.S. 1991. Siderophores in microbial interactions on plant surfaces. *Mol Plant-Microbe Interact*, 4: 5 – 13.
- LOPEZ, D. & KOLTER, R. 2010. Functional microdomains in bacterial membranes. *Genes Dev*, 24, 1893-902.
- LOW, T. Y., MOHTAR, M. A., LEE, P. Y., OMAR, N., ZHOU, H. & YE, M. 2021. Widening the Bottleneck of Phosphoproteomics: Evolving Strategies for Phosphopeptide Enrichment. *Mass Spectrom Rev*, 40, 309-333.
- LU, S., FAN, S. B., YANG, B., LI, Y. X., MENG, J. M., WU, L., LI, P., ZHANG, K., ZHANG, M. J., FU, Y., LUO, J., SUN, R. X., HE, S. M. & DONG, M. Q. 2015. Mapping native disulfide bonds at a proteome scale. *Nat Methods*, 12, 329-31.
- LU, T., KE, M., LAVOIE, M., JIN, Y., FAN, X., ZHANG, Z., FU, Z., SUN, L., GILLINGS, M., PENUELAS, J., QIAN, H. & ZHU, Y. G. 2018. Rhizosphere microorganisms can influence the timing of plant flowering. *Microbiome*, 6, 231.

- LUGTENBERG, B. & KAMILOVA, F. 2009. Plant-growth-promoting rhizobacteria. *Annu Rev Microbiol*, 63, 541-56.
- LUNDBERG, D. S., LEBEIS, S. L., PAREDES, S. H., YOURSTONE, S., GEHRING, J., Malfatti, S., TREMBLAY, J., ENGELBREKTSON, A., KUNIN, V., DEL RIO, T. G., EDGAR, R. C., EICKHORST, T., LEY, R. E., HUGENHOLTZ, P., TRINGE, S. G. & DANGL, J. L. 2012. Defining the core Arabidopsis thaliana root microbiome. *Nature*, 488, 86-90.
- LUO, S., COUTINHO, B. G., DADHWAL, P., ODA, Y., REN, J., SCHAEFER, A. L., GREENBERG, E. P., HARWOOD, C. S. & TONG, L. 2021. Structural basis for a bacterial Pip system plant effector recognition protein. *Proc Natl Acad Sci U S A*, 118.
- MA, B., ZHANG, K., HENDRIE, C., LIANG, C., LI, M., DOHERTY-KIRBY, A. & LAJOIE, G. 2003. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom*, 17, 2337-42.
- MA, C. W. & LAM, H. 2014. Hunting for unexpected post-translational modifications by spectral library searching with tier-wise scoring. *J Proteome Res*, 13, 2262-71.
- MACEK, B., FORCHHAMMER, K., HARDOUIN, J., WEBER-BAN, E., GRANGEASSE, C. & MIJAKOVIC, I. 2019. Protein post-translational modifications in bacteria. *Nat Rev Microbiol*, 17, 651-664.
- MAKAROV, A. 2000. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem*, 72, 1156-62.
- MALINOVSKY, F. G., FANGEL, J. U. & WILLATS, W. G. 2014. The role of the cell wall in plant immunity. *Frontiers in plant science*, 5, 178.
- MAN HA, C., FINE, D., BHATIA, A., RAO, X., MARTIN, M. Z., ENGLE, N. L., WHERRITT, D. J., TSCHAPLINSKI, T. J., SUMNER, L. W. & DIXON, R. A. 2019. Ectopic Defense Gene Expression Is Associated with Growth Defects in Medicago truncatula Lignin Pathway Mutants. *Plant Physiol*, 181, 63-84.
- MANDYAM, K. G. & JUMPPONEN, A. 2014. Mutualism-parasitism paradigm synthesized from results of root-endophyte models. *Front Microbiol*, 5, 776.
- MANT, C. T., CHEN, Y., YAN, Z., POPA, T. V., KOVACS, J. M., MILLS, J. B., TRIPET, B. P. & HODGES, R. S. 2007. HPLC analysis and purification of peptides. *Methods Mol Biol*, 386, 3-55.
- MARK, G. L., DOW, J. M., KIELY, P. D., HIGGINS, H., HAYNES, J., BAYSSE, C., ABBAS, A., FOLEY, T., FRANKS, A., MORRISSEY, J. & O'GARA, F. 2005. Transcriptome profiling of bacterial responses to root exudates identifies genes involved in microbe-plant interactions. *Proc Natl Acad Sci U S A*, 102, 17454-9.

- MARLOW, J. J., SKENNERTON, C. T., LI, Z., CHOUREY, K., HETTICH, R. L., PAN, C. & ORPHAN, V. J. 2016. Proteomic Stable Isotope Probing Reveals Biosynthesis Dynamics of Slow Growing Methane Based Microbial Communities. *Front Microbiol*, 7, 563.
- MARNETT, L. J. 2000. Oxyradicals and DNA damage. *Carcinogenesis*, 21, 361-70.
- MARTINEZ-KLIMOVA, E., RODRÍGUEZ-PEÑA, K. & SÁNCHEZ, S. 2017. Endophytes as sources of antibiotics. *Biochemical pharmacology*, 134, 1-17.
- MARTÍNEZ-NÚÑEZ, M. A., & Y LÓPEZ, V. E. L. 2016. Nonribosomal peptides synthetases and their applications in industry. *Sustainable Chemical Processes*, 4(1), 1-8.
- MATA-PEREZ, C., SANCHEZ-CALVO, B., BEGARA-MORALES, J. C., LUQUE, F., JIMENEZ-RUIZ, J., PADILLA, M. N., FIERRO-RISCO, J., VALDERRAMA, R., FERNANDEZ-OCANA, A., CORPAS, F. J. & BARROSO, J. B. 2015. Transcriptomic profiling of linolenic acid-responsive genes in ROS signaling from RNA-seq data in Arabidopsis. *Front Plant Sci*, 6, 122.
- MAVRODI, D. V., PAREJKO, J. A., MAVRODI, O. V., KWAK, Y. S., WELLER, D. M., BLANKENFELDT, W. & THOMASHOW, L. S. 2013. Recent insights into the diversity, frequency and ecological roles of phenazines in fluorescent *Pseudomonas* spp. *Environmental Microbiology*, 15, 675-686.
- MAYRHOFFER, S., TEUBER, M., ZIMMER, I., LOUIS, S., FISCHBACH, R. J. & SCHNITZLER, J. P. 2005. Diurnal and seasonal variation of isoprene biosynthesis-related genes in grey poplar leaves. *Plant Physiol*, 139, 474-84.
- MCCLUNG, C. R. 2006. Plant circadian rhythms. *Plant Cell*, 18, 792-803.
- MECHIN, V., DAMERVAL, C. & ZIVY, M. 2007. Total protein extraction with TCA-acetone. *Methods Mol Biol*, 355, 1-8.
- MEIJER, H. J. & MUNNIK, T. 2003. Phospholipid-based signaling in plants. *Annu Rev Plant Biol*, 54, 265-306.
- MELO, A. M., BANDEIRAS, T. M. & TEIXEIRA, M. 2004. New insights into type II NAD(P)H:quinone oxidoreductases. *Microbiol Mol Biol Rev*, 68, 603-16.
- MENDES, R., KRUIJT, M., DE BRUIJN, I., DEKKERS, E., VAN DER VOORT, M., SCHNEIDER, J. H., PICENO, Y. M., DESANTIS, T. Z., ANDERSEN, G. L., BAKKER, P. A. & RAAIJMAKERS, J. M. 2011. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science*, 332, 1097-100.
- MENG, X. & ZHANG, S. 2013. MAPK cascades in plant disease resistance signaling. *Annu Rev Phytopathol*, 51, 245-66.

- MI, H., EBERT, D., MURUGANUJAN, A., MILLS, C., ALBOU, L. P., MUSHAYAMAHA, T. & THOMAS, P. D. 2021. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res*, 49, D394-D403.
- MICHAL, G. & SCHOMBURG, D. 2013. *Biochemical pathways: an atlas of biochemistry and molecular biology*, Wiley.
- MIJAKOVIC, I., GRANGEASSE, C. & TURGAY, K. 2016. Exploring the diversity of protein modifications: special bacterial phosphorylation systems. *FEMS Microbiol Rev*, 40, 398-417.
- MILLAR, A. H., HEAZLEWOOD, J. L., GIGLIONE, C., HOLDSWORTH, M. J., BACHMAIR, A. & SCHULZE, W. X. 2019. The Scope, Functions, and Dynamics of Posttranslational Protein Modifications. *Annu Rev Plant Biol*, 70, 119-151.
- MILLER, G., SUZUKI, N., RIZHSKY, L., HEGIE, A., KOUSSEVITZKY, S. & MITTLER, R. 2007. Double mutants deficient in cytosolic and thylakoid ascorbate peroxidase reveal a complex mode of interaction between reactive oxygen species, plant development, and response to abiotic stresses. *Plant Physiol*, 144, 1777-85.
- MILO, R. 2013. What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays*, 35, 1050-5.
- MISCHNIK, M., SACCO, F., COX, J., SCHNEIDER, H. C., SCHAFER, M., HENDLICH, M., CROWTHER, D., MANN, M. & KLABUNDE, T. 2016. IKAP: A heuristic framework for inference of kinase activities from Phosphoproteomics data. *Bioinformatics*, 32, 424-31.
- MOLLE, V., FUJITA, M., JENSEN, S. T., EICHENBERGER, P., GONZALEZ-PASTOR, J. E., LIU, J. S. & LOSICK, R. 2003. The Spo0A regulon of *Bacillus subtilis*. *Mol Microbiol*, 50, 1683-701.
- MORI, I. C., MURATA, Y., YANG, Y., MUNEMASA, S., WANG, Y. F., ANDREOLI, S., TIRIAC, H., ALONSO, J. M., HARPER, J. F., ECKER, J. R., KWAK, J. M. & SCHROEDER, J. I. 2006. CDPKs CPK6 and CPK3 function in ABA regulation of guard cell S-type anion- and Ca(2+)-permeable channels and stomatal closure. *PLoS Biol*, 4, e327.
- MOTOYAMA, A., VENABLE, J. D., RUSE, C. I. & YATES, J. R., 3RD 2006. Automated ultra-high-pressure multidimensional protein identification technology (UHP-MudPIT) for improved peptide identification of proteomic samples. *Anal Chem*, 78, 5109-18.
- MURO-VILLANUEVA, F., MAO, X. & CHAPPLE, C. 2019. Linking phenylpropanoid metabolism, lignin deposition, and plant growth inhibition. *Curr Opin Biotechnol*, 56, 202-208.

- MURRAY, K. K., BOYD, R. K., EBERLIN, M. N., LANGLEY, G. J., LI, L. & NAITO, Y. 2013. Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). *Pure and Applied Chemistry*, 85, 1515-1609.
- NA, S., BANDEIRA, N. & PAEK, E. 2012. Fast multi-blind modification search through tandem mass spectrometry. *Mol Cell Proteomics*, 11, M111 010199.
- NAITHANI, S., CHOOKAJORN, T., RIPOLL, D. R. & NASRALLAH, J. B. 2007. Structural modules for receptor dimerization in the S-locus receptor kinase extracellular domain. *Proc Natl Acad Sci U S A*, 104, 12211-6.
- NAKAMICHI, N., TAKAO, S., KUDO, T., KIBA, T., WANG, Y., KINOSHITA, T. & SAKAKIBARA, H. 2016. Improvement of Arabidopsis Biomass and Cold, Drought and Salinity Stress Tolerance by Modified Circadian Clock-Associated PSEUDO-RESPONSE REGULATORS. *Plant Cell Physiol*, 57, 1085-97.
- NASSER, W. & REVERCHON, S. 2007. New insights into the regulatory mechanisms of the LuxR family of quorum sensing regulators. *Anal Bioanal Chem*, 387, 381-90.
- NAVARRO-GOCHICOA, M. T., CAMUT, S., TIMMERS, A. C., NIEBEL, A., HERVE, C., BOUTET, E., BONO, J. J., IMBERTY, A. & CULLIMORE, J. V. 2003. Characterization of four lectin-like receptor kinases expressed in roots of *Medicago truncatula*. Structure, location, regulation of expression, and potential role in the symbiosis with *Sinorhizobium meliloti*. *Plant Physiol*, 133, 1893-910.
- NEFEDOV, A. V., GILSKI, M. J. & SADYGOV, R. G. 2011. Bioinformatics Tools for Mass Spectrometry-Based High-Throughput Quantitative Proteomics Platforms. *Curr Proteomics*, 8, 125-137.
- NEIDHARDT, F. C., BLOCH, P. L. & SMITH, D. F. 1974. Culture medium for enterobacteria. *J Bacteriol*, 119, 736-47.
- NERVA, L., GARCIA, J. F., FAVARETTO, F., GIUDICE, G., MOFFA, L., SANDRINI, M., CANTU, D., ZANZOTTO, A., GARDIMAN, M., VELASCO, R., GAMBINO, G. & CHITARRA, W. 2022. The hidden world within plants: metatranscriptomics unveils the complexity of wood microbiomes. *J Exp Bot*, 73, 2682-2697.
- NEUROHR, G. E. & AMON, A. 2020. Relevance and Regulation of Cell Density. *Trends Cell Biol*, 30, 213-225.
- NEVEROVA, I. & VAN EYK, J. E. 2002. Application of reversed phase high performance liquid chromatography for subproteomic analysis of cardiac muscle. *Proteomics*, 2, 22-31.
- NIESSEN, W. M. & FALCK, D. 2015. Introduction to mass spectrometry, a tutorial. *Analyzing biomolecular interactions by mass spectrometry*, 1-54.

- NISA, M. U., HUANG, Y., BENHAMED, M. & RAYNAUD, C. 2019. The Plant DNA Damage Response: Signaling Pathways Leading to Growth Inhibition and Putative Role in Response to Stress Conditions. *Front Plant Sci*, 10, 653.
- NISHIZUKA, Y. 1984. The role of protein kinase C in cell surface signal transduction and tumour promotion. *Nature*, 308, 693-8.
- NOBLE, W. S. 2009. How does multiple testing correction work? *Nature biotechnology*, 27, 1135-1137.
- NOESKE, J., WASSERMAN, M. R., TERRY, D. S., ALTMAN, R. B., BLANCHARD, S. C. & CATE, J. H. 2015. High-resolution structure of the Escherichia coli ribosome. *Nat Struct Mol Biol*, 22, 336-41.
- OLDROYD, G. E. 2013. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat Rev Microbiol*, 11, 252-63.
- OLSEN, J. V. & MANN, M. 2013. Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol Cell Proteomics*, 12, 3444-52.
- ORTIZ-CASTRO, R., PELAGIO-FLORES, R., MENDEZ-BRAVO, A., RUIZ-HERRERA, L. F., CAMPOS-GARCIA, J. & LOPEZ-BUCIO, J. 2014. Pyocyanin, a virulence factor produced by Pseudomonas aeruginosa, alters root development through reactive oxygen species and ethylene signaling in Arabidopsis. *Mol Plant Microbe Interact*, 27, 364-78.
- PAIZS, B. & SUHAI, S. 2005. Fragmentation pathways of protonated peptides. *Mass Spectrom Rev*, 24, 508-48.
- PANDA, C., LI, X., WAGER, A., CHEN, H. Y. & LI, X. 2020. An importin-beta-like protein mediates lignin-modification-induced dwarfism in Arabidopsis. *Plant J*, 102, 1281-1293.
- PANDESWARI, P. B. & SABAREESH, V. 2018. Middle-down approach: a choice to sequence and characterize proteins/proteomes by mass spectrometry. *RSC Adv*, 9, 313-344.
- PAPPIREDDI, N., MARTIN, L. & WUHR, M. 2019. A Review on Quantitative Multiplexed Proteomics. *Chembiochem*, 20, 1210-1224.
- PARKER, K. C., PATTERSON, D., WILLIAMSON, B., MARCHESE, J., GRABER, A., HE, F., JACOBSON, A., JUHASZ, P. & MARTIN, S. 2004. Depth of proteome issues: a yeast isotope-coded affinity tag reagent study. *Mol Cell Proteomics*, 3, 625-59.
- PATANKAR, A. V. & GONZALEZ, J. E. 2009. Orphan LuxR regulators of quorum sensing. *FEMS Microbiol Rev*, 33, 739-56.
- PATTERSON, S. D. & AEBERSOLD, R. 1995. Mass spectrometric approaches for the identification of gel-separated proteins. *Electrophoresis*, 16, 1791-814.

- PEREIRA, A. 2016. Plant Abiotic Stress Challenges from the Changing Environment. *Front Plant Sci*, 7, 1123.
- PETRICKA, J. J. & NELSON, T. M. 2007. Arabidopsis nucleolin affects plant development and patterning. *Plant Physiol*, 144, 173-86.
- PETRIK, D. L., KARLEN, S. D., CASS, C. L., PADMAKSHAN, D., LU, F., LIU, S., LE BRIS, P., ANTELME, S., SANTORO, N., WILKERSON, C. G., SIBOUT, R., LAPIERRE, C., RALPH, J. & SEDBROOK, J. C. 2014. p-Coumaroyl-CoA:monolignol transferase (PMT) acts specifically in the lignin biosynthetic pathway in *Brachypodium distachyon*. *Plant J*, 77, 713-26.
- PIERRET, A., DOUSSAN, C., CAPOWIEZ, Y., BASTARDIE, F. & PAGÈS, L. 2007. Root functional architecture: a framework for modeling the interplay between roots and soil. *Vadose Zone Journal*, 6, 269-281.
- PIETERSE, C. M., ZAMIOUDIS, C., BERENDSEN, R. L., WELLER, D. M., VAN WEES, S. C. & BAKKER, P. A. 2014. Induced systemic resistance by beneficial microbes. *Annu Rev Phytopathol*, 52, 347-75.
- PII, Y., MIMMO, T., TOMASI, N., TERZANO, R., CESCO, S., & CRECCHIO, C. 2015. Microbial interactions in the rhizosphere: beneficial influences of plant growth-promoting rhizobacteria on nutrient acquisition process. *Biology and Fertility of Soils*, 51(4), 403-415.
- PINEDA, A., SOLER, R., POZO, M. J., RASMANN, S. & TURLINGS, T. C. 2015. Editorial: Above-belowground interactions involving plants, microbes and insects. *Front Plant Sci*, 6, 318.
- PIQUEMAL, J., LAPIERRE, C., MYTON, K., O'CONNELL, A., SCHUCH, W., GRIMA-PETTENATI, J., & BOUDET, A. M. 1998. Down-regulation of cinnamoyl-CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants. *The Plant Journal*, 13(1), 71-83.
- PLETT, J. M., KEMPPAINEN, M., KALE, S. D., KOHLER, A., LEGUE, V., BRUN, A., TYLER, B. M., PARDO, A. G. & MARTIN, F. 2011. A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Curr Biol*, 21, 1197-203.
- PNUELI, L., LIANG, H., ROZENBERG, M. & MITTLER, R. 2003. Growth suppression, altered stomatal responses, and augmented induction of heat shock proteins in cytosolic ascorbate peroxidase (Apx1)-deficient Arabidopsis plants. *Plant J*, 34, 187-203.
- PODHORECKA, M., SKLADANOWSKI, A. & BOZKO, P. 2010. H2AX Phosphorylation: Its Role in DNA Damage Response and Cancer Therapy. *J Nucleic Acids*, 2010.
- POLEVODA, B. & SHERMAN, F. 2007. Methylation of proteins involved in translation. *Mol Microbiol*, 65, 590-606.

- POLPITIYA, A. D., QIAN, W. J., JAITLY, N., PETYUK, V. A., ADKINS, J. N., CAMP, D. G., 2ND, ANDERSON, G. A. & SMITH, R. D. 2008. DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics*, 24, 1556-8.
- PORETSKY, R., RODRIGUEZ, R. L., LUO, C., TSEMENTZI, D. & KONSTANTINIDIS, K. T. 2014. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One*, 9, e93827.
- POSEWITZ, M. C. & TEMPST, P. 1999. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem*, 71, 2883-92.
- PRABAKARAN, S., LIPPENS, G., STEEN, H., GUNAWARDENA, J. J. W. I. R. S. B. & MEDICINE 2012. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. 4, 565-583.
- QIAN, Y., LYNCH, J. H., GUO, L., RHODES, D., MORGAN, J. A. & DUDAREVA, N. 2019. Completion of the cytosolic post-chorismate phenylalanine biosynthetic pathway in plants. *Nat Commun*, 10, 15.
- QIAO, Z., YATES, T. B., SHRESTHA, H. K., ENGLE, N. L., FLANAGAN, A., MORRELL-FALVEY, J. L., SUN, Y., TSCHAPLINSKI, T. J., ABRAHAM, P. E., LABBE, J., WANG, Z. Y., HETTICH, R. L., TUSKAN, G. A., MUCHERO, W. & CHEN, J. G. 2021. Towards engineering ectomycorrhization into switchgrass bioenergy crops via a lectin receptor-like kinase. *Plant Biotechnol J*, 19, 2454-2468.
- RAAIJMAKERS, J. M., PAULITZ, T. C., STEINBERG, C., ALABOUVETTE, C. & MOËNNE-LOCCOZ, Y. 2009. The rhizosphere: a playground and battlefield for soilborne pathogens and beneficial microorganisms. *Plant and soil*, 321, 341-361.
- RAE, G. M., DAVID, K., & WOOD, M. 2013. The dormancy marker DRM1/ARP associated with dormancy but a broader role in planta. *Developmental Biology Journal*, 632524, 12.
- RAGAUSKAS, A. J., BECKHAM, G. T., BIDDY, M. J., CHANDRA, R., CHEN, F., DAVIS, M. F., DAVISON, B. H., DIXON, R. A., GILNA, P., KELLER, M., LANGAN, P., NASKAR, A. K., SADDLER, J. N., TSCHAPLINSKI, T. J., TUSKAN, G. A. & WYMAN, C. E. 2014. Lignin valorization: improving lignin processing in the biorefinery. *Science*, 344, 1246843.
- RALPH, J., LUNDQUIST, K., BRUNOW, G., LU, F., KIM, H., SCHATZ, P. F., MARITA, J. M., HATFIELD, R. D., RALPH, S. A. & CHRISTENSEN, J. H. 2004. Lignins: natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids. *Phytochemistry reviews*, 3(1), 29-60.
- RANF, S., GISCH, N., SCHAFFER, M., ILLIG, T., WESTPHAL, L., KNIREL, Y. A., SANCHEZ-CARBALLO, P. M., ZHRINGER, U., HUCKELHOVEN, R., LEE, J. &

- SCHEEL, D. 2015. A lectin S-domain receptor kinase mediates lipopolysaccharide sensing in *Arabidopsis thaliana*. *Nat Immunol*, 16, 426-33.
- RANOCHA, P., DENANCE, N., VANHOLME, R., FREYDIER, A., MARTINEZ, Y., HOFFMANN, L., KOHLER, L., POUZET, C., RENOU, J. P., SUNDBERG, B., BOERJAN, W. & GOFFNER, D. 2010. Walls are thin 1 (WAT1), an *Arabidopsis* homolog of *Medicago truncatula* NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. *Plant J*, 63, 469-83.
- RANOCHA, P., DIMA, O., NAGY, R., FELTEN, J., CORRATGE-FAILLIE, C., NOVAK, O., MORREEL, K., LACOMBE, B., MARTINEZ, Y., PFRUNDER, S., JIN, X., RENOU, J. P., THIBAUD, J. B., LJUNG, K., FISCHER, U., MARTINOIA, E., BOERJAN, W. & GOFFNER, D. 2013. *Arabidopsis* WAT1 is a vacuolar auxin transport facilitator required for auxin homeostasis. *Nat Commun*, 4, 2625.
- RAUDVERE, U., KOLBERG, L., KUZMIN, I., ARAK, T., ADLER, P., PETERSON, H. & VILO, J. 2019. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*, 47, W191-W198.
- RAVIKUMAR, V., MACEK, B. & MIJAKOVIC, I. 2016. Resources for Assignment of Phosphorylation Sites on Peptides and Proteins. *Methods Mol Biol*, 1355, 293-306.
- RAYNIE, D. 2022. New Sample Preparation Products and Accessories For 2021–2022. *LCGC North America*, 40, 218–222-218–222.
- REASONER, D. J. & GELDREICH, E. E. 1985. A new medium for the enumeration and subculture of bacteria from potable water. *Appl Environ Microbiol*, 49, 1-7.
- REDDY, M. S., NAITHANI, S., TULI, R. & SANE, P. V. 2000. Diurnal regulation of plastid genes in *Populus deltoides*. *Indian J Biochem Biophys*, 37, 453-8.
- REDMOND, J. W., BATLEY, M., DJORDJEVIC, M. A., INNES, R. W., KUEMPEL, P. L. & ROLFE, B. G. 1986. Flavones induce expression of nodulation genes in *Rhizobium*. *Nature*, 323, 632-635.
- REGALADO, J., LUNDBERG, D. S., DEUSCH, O., KERSTEN, S., KARASOV, T., POERSCH, K., SHIRSEKAR, G. & WEIGEL, D. 2020. Combining whole-genome shotgun sequencing and rRNA gene amplicon analyses to improve detection of microbe-microbe interaction networks in plant leaves. *ISME J*, 14, 2116-2130.
- REID, W. V., ALI, M. K. & FIELD, C. B. 2020. The future of bioenergy. *Glob Chang Biol*, 26, 274-286.
- RIBEIRO, C. W., CARVALHO, F. E., ROSA, S. B., ALVES-FERREIRA, M., ANDRADE, C. M., RIBEIRO-ALVES, M., SILVEIRA, J. A., MARGIS, R. & MARGIS-PINHEIRO, M.

2012. Modulation of genes related to specific metabolic pathways in response to cytosolic ascorbate peroxidase knockdown in rice plants. *Plant Biol (Stuttg)*, 14, 944-55.
- RILEY, N. M. & COON, J. J. 2016. Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Anal Chem*, 88, 74-94.
- ROEPSTORFF, P. & FOHLMAN, J. 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom*, 11, 601.
- ROGERS, J. C. & BOMGARDEN, R. D. 2016. Sample Preparation for Mass Spectrometry-Based Proteomics; from Proteomes to Peptides. *Adv Exp Med Biol*, 919, 43-62.
- ROSENBLUETH, M. & MARTÍNEZ-ROMERO, E. 2006. Bacterial endophytes and their interactions with hosts. *Molecular plant-microbe interactions*, 19, 827-837.
- ROSSEN, L., SHEARMAN, C. A., JOHNSTON, A. W. & DOWNIE, J. A. 1985. The nodD gene of *Rhizobium leguminosarum* is autoregulatory and in the presence of plant exudate induces the nodA,B,C genes. *EMBO J*, 4, 3369-73.
- ROTH-BEJERANO, N., LIVNE, D. & KAGAN-ZUR, V. 1990. *Helianthemum*—*Terfezia* relations in different growth media. *New phytologist*, 114, 235-238.
- RUNNING, W. E. & REILLY, J. P. 2009. Ribosomal Proteins of *Deinococcus radiodurans*: Their Solvent Accessibility and Reactivity. *Journal of Proteome Research*, 8, 1228-1246.
- RUSH, J., MORITZ, A., LEE, K. A., GUO, A., GOSS, V. L., SPEK, E. J., ZHANG, H., ZHA, X. M., POLAKIEWICZ, R. D. & COMB, M. J. 2005. Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol*, 23, 94-101.
- RUYTER-SPIRA, C., AL-BABILI, S., VAN DER KROL, S. & BOUWMEESTER, H. 2013. The biology of strigolactones. *Trends Plant Sci*, 18, 72-83.
- SACCO, F., PERFETTO, L. & CESARENI, G. 2018. Combining Phosphoproteomics Datasets and Literature Information to Reveal the Functional Connections in a Cell Phosphorylation Network. *Proteomics*, 18, e1700311.
- SAKAI, T., HONING, H., NISHIOKA, M., UEHARA, Y., TAKAHASHI, M., FUJISAWA, N., SAJI, K., SEKI, M., SHINOZAKI, K., JONES, M. A., SMIRNOFF, N., OKADA, K. & WASTENEYS, G. O. 2008. Armadillo repeat-containing kinesins and a NIMA-related kinase are required for epidermal-cell morphogenesis in *Arabidopsis*. *Plant J*, 53, 157-71.
- SALAH UD-DIN, A. I. M. & ROUJEINIKOVA, A. 2017. Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archaea. *Cell Mol Life Sci*, 74, 3293-3303.

- SANTOS-MEDELLIN, C., EDWARDS, J., NGUYEN, B. & SUNDARESAN, V. 2022. Acquisition of a complex root microbiome reshapes the transcriptomes of rice plants. *New Phytol.*
- SARATH, G., MITCHELL, R. B., SATTTLER, S. E., FUNNELL, D., PEDERSEN, J. F., GRAYBOSCH, R. A. & VOGEL, K. P. 2008. Opportunities and roadblocks in utilizing forages and small grains for liquid fuels. *J Ind Microbiol Biotechnol*, 35, 343-354.
- SCHAEFER, A. L., LAPPALA, C. R., MORLEN, R. P., PELLETIER, D. A., LU, T. Y., LANKFORD, P. K., HARWOOD, C. S. & GREENBERG, E. P. 2013. LuxR- and luxI-type quorum-sensing circuits are prevalent in members of the *Populus deltoides* microbiome. *Appl Environ Microbiol*, 79, 5745-52.
- SCHAEFER, A. L., ODA, Y., COUTINHO, B. G., PELLETIER, D. A., WEIBURG, J., VENTURI, V., GREENBERG, E. P. & HARWOOD, C. S. 2016. A LuxR Homolog in a Cottonwood Tree Endophyte That Activates Gene Expression in Response to a Plant Signal or Specific Peptides. *mBio*, 7.
- SCHILMILLER, A. L., STOUT, J., WENG, J. K., HUMPHREYS, J., RUEGGER, M. O. & CHAPPLE, C. 2009. Mutations in the cinnamate 4-hydroxylase gene impact metabolism, growth and development in *Arabidopsis*. *Plant J*, 60, 771-82.
- SCHMIDT, R., CORDOVEZ, V., DE BOER, W., RAAIJMAKERS, J. & GARBEVA, P. 2015. Volatile affairs in microbial interactions. *ISME J*, 9, 2329-35.
- SCHMITZ, A. M. & HARRISON, M. J. 2014. Signaling events during initiation of arbuscular mycorrhizal symbiosis. *J Integr Plant Biol*, 56, 250-61.
- SCHUSTER, M., SEXTON, D. J., DIGGLE, S. P. & GREENBERG, E. P. 2013. Acyl-homoserine lactone quorum sensing: from evolution to application. *Annu Rev Microbiol*, 67, 43-63.
- SCIGELOVA, M. & MAKAROV, A. 2006. Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics*, 6 Suppl 2, 16-21.
- SEHRAWAT, A., ABAT, J. K. & DESWAL, R. 2013. RuBisCO depletion improved proteome coverage of cold responsive S-nitrosylated targets in *Brassica juncea*. *Front Plant Sci*, 4, 342.
- SELIN, C., HABIBIAN, R., PORITSANOS, N., ATHUKORALA, S. N., FERNANDO, D. & DE KIEVIT, T. R. 2010. Phenazines are not essential for *Pseudomonas chlororaphis* PA23 biocontrol of *Sclerotinia sclerotiorum*, but do play a role in biofilm formation. *FEMS Microbiol Ecol*, 71, 73-83.
- SHALOM, G., SHAW, J. G. & THOMAS, M. S. 2007. In vivo expression technology identifies a type VI secretion system locus in *Burkholderia pseudomallei* that is induced upon invasion of macrophages. *Microbiology (Reading)*, 153, 2689-2699.

- SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. & IDEKER, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13, 2498-504.
- SHEVCHENKO, A., WILM, M., VORM, O. & MANN, M. 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem*, 68, 850-8.
- SHORESH, M., HARMAN, G. E. & MASTOURI, F. 2010. Induced systemic resistance and plant responses to fungal biocontrol agents. *Annu Rev Phytopathol*, 48, 21-43.
- SHYU, A. B., WILKINSON, M. F. & VAN HOOF, A. 2008. Messenger RNA regulation: to translate or to degrade. *EMBO J*, 27, 471-81.
- SITARAMAN, R. 2015. Pseudomonas spp. as models for plant-microbe interactions. *Front Plant Sci*, 6, 787.
- SIVEEN, K. S., PRABHU, K. S., ACHKAR, I. W., KUTTIKRISHNAN, S., SHYAM, S., KHAN, A. Q., MERHI, M., DERMIME, S. & UDDIN, S. 2018. Role of Non Receptor Tyrosine Kinases in Hematological Malignances and its Targeting by Natural Products. *Mol Cancer*, 17, 31.
- SMITH, L. M., KELLEHER, N. L. & CONSORTIUM FOR TOP DOWN, P. 2013. Proteoform: a single term describing protein complexity. *Nat Methods*, 10, 186-7.
- SMOLKA, M. B., ZHOU, H., PURKAYASTHA, S. & AEBERSOLD, R. 2001. Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. *Anal Biochem*, 297, 25-31.
- SOLOVYEVA, E. M., LOBAS, A. A., KOPYLOV, A. T., ILINA, I. Y., LEVITSKY, L. I., MOSHKOVSKII, S. A. & GORSHKOV, M. V. 2018. FractionOptimizer: a method for optimal peptide fractionation in bottom-up proteomics. *Anal Bioanal Chem*, 410, 3827-3833.
- SPAEPEN, S., BOSSUYT, S., ENGELEN, K., MARCHAL, K. & VANDERLEYDEN, J. 2014. Phenotypical and molecular responses of Arabidopsis thaliana roots as a result of inoculation with the auxin-producing bacterium Azospirillum brasilense. *New Phytol*, 201, 850-861.
- SPERLING, K. 2001. From proteomics to genomics. *Electrophoresis*, 22, 2835-2837.
- SRIVASTAVA, D., SHAMIM, M., KUMAR, M., MISHRA, A., MAURYA, R., SHARMA, D., ... & SINGH, K. N. 2019. Role of circadian rhythm in plant system: An update from development to stress response. *Environmental and Experimental Botany*, 162, 256-271.
- STAMP, N. 2003. Out of the quagmire of plant defense hypotheses. *Q Rev Biol*, 78, 23-55.

- STANTON, B. J., BOURQUE, A., COLEMAN, M., EISENBIES, M., EMERSON, R. M., ESPINOZA, J., ... & ZERPA, J. 2020. The practice and economics of hybrid poplar biomass production for biofuels and bioproducts in the Pacific Northwest. *BioEnergy Research*, 1-18.
- STOLPOVSKY, K., MARTINEZ-LAVANCHY, P., HEIPIEPER, H. J., VAN CAPPELLEN, P., & THULLNER, M. 2011. Incorporating dormancy in dynamic microbial community models. *Ecological Modelling*, 222(17), 3092-3102.
- STRADER, M. B., VERBERKMOES, N. C., TABB, D. L., CONNELLY, H. M., BARTON, J. W., BRUCE, B. D., PELLETIER, D. A., DAVISON, B. H., HETTICH, R. L., LARIMER, F. W. & HURST, G. B. 2004. Characterization of the 70S Ribosome from *Rhodospseudomonas palustris* using an integrated "top-down" and "bottom-up" mass spectrometric approach. *J Proteome Res*, 3, 965-78.
- SU, W., SUN, J., SHIMIZU, K. & KADOTA, K. 2019. TCC-GUI: a Shiny-based application for differential expression analysis of RNA-Seq count data. *BMC Res Notes*, 12, 133.
- SUBRAMONI, S., GONZALEZ, J. F., JOHNSON, A., PECHY-TARR, M., ROCHAT, L., PAULSEN, I., LOPER, J. E., KEEL, C. & VENTURI, V. 2011. Bacterial subfamily of LuxR regulators that respond to plant compounds. *Appl Environ Microbiol*, 77, 4579-88.
- SUBRAMONI, S. & VENTURI, V. 2009. LuxR-family 'solos': bachelor sensors/regulators of signalling molecules. *Microbiology (Reading)*, 155, 1377-1385.
- SUGIYAMA, Y., KATAYAMA, S., KAMESHITA, I., MORISAWA, K., HIGUCHI, T., TODAKA, H., KINOSHITA, E., KINOSHITA-KIKUTA, E., KOIKE, T., TANIGUCHI, T. & SAKAMOTO, S. 2015. Expression and phosphorylation state analysis of intracellular protein kinases using Multi-PK antibody and Phos-tag SDS-PAGE. *MethodsX*, 2, 469-74.
- SUN, Y., QIAO, Z., MUCHERO, W. & CHEN, J. G. 2020. Lectin Receptor-Like Kinases: The Sensor and Mediator at the Plant Cell Surface. *Front Plant Sci*, 11, 596301.
- SUPEK, F., BOSNJAK, M., SKUNCA, N. & SMUC, T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, 6, e21800.
- SZKLARCZYK, D., GABLE, A. L., NASTOU, K. C., LYON, D., KIRSCH, R., PYYSALO, S., DONCHEVA, N. T., LEGEAY, M., FANG, T., BORK, P., JENSEN, L. J. & VON MERING, C. 2021. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*, 49, D605-D612.
- TAN, I. S. & RAMAMURTHI, K. S. 2014. Spore formation in *Bacillus subtilis*. *Environ Microbiol Rep*, 6, 212-25.

- TAO, W. A., ZHANG, Y. & WILEY, J. 2019. *Mass Spectrometry-Based Chemical Proteomics*, Wiley Online Library.
- TARTAGLIA, M., BASTIDA, F., SCIARRILLO, R. & GUARINO, C. 2020. Soil Metaproteomics for the Study of the Relationships Between Microorganisms and Plants: A Review of Extraction Protocols and Ecological Insights. *Int J Mol Sci*, 21.
- TAYLOR, L., NUNES-NESE, A., PARSLEY, K., LEISS, A., LEACH, G., COATES, S., WINGLER, A., FERNIE, A. R. & HIBBERD, J. M. 2010. Cytosolic pyruvate,orthophosphate dikinase functions in nitrogen remobilization during leaf senescence and limits individual seed growth and nitrogen content. *Plant J*, 62, 641-52.
- TAYLOR, L. P. & GROTEWOLD, E. 2005. Flavonoids as developmental regulators. *Curr Opin Plant Biol*, 8, 317-23.
- TECON, R., MITRI, S., CICCARESE, D., OR, D., VAN DER MEER, J. R. & JOHNSON, D. R. 2019. Bridging the Holistic-Reductionist Divide in Microbial Ecology. *mSystems*, 4.
- THEVIS, M., OGORZALEK LOO, R. R. & LOO, J. A. 2003. In-gel derivatization of proteins for cysteine-specific cleavages and their analysis by mass spectrometry. *J Proteome Res*, 2, 163-72.
- THOMAS, S. G. & SUN, T. P. 2004. Update on gibberellin signaling. A tale of the tall and the short. *Plant Physiol*, 135, 668-76.
- TIAN, T., LIU, Y., YAN, H., YOU, Q., YI, X., DU, Z., XU, W. & SU, Z. 2017. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res*, 45, W122-W129.
- TIMMS-WILSON, T. M., ELLIS, R. J., RENWICK, A., RHODES, D. J., MAVRODI, D. V., WELLER, D. M., THOMASHOW, L. S. & BAILEY, M. J. 2000. Chromosomal insertion of phenazine-1-carboxylic acid biosynthetic pathway enhances efficacy of damping-off disease control by *Pseudomonas fluorescens*. *Mol Plant Microbe Interact*, 13, 1293-300.
- TOBIMATSU, Y. & SCHUETZ, M. 2019. Lignin polymerization: how do plants manage the chemistry so well? *Curr Opin Biotechnol*, 56, 75-81.
- TODD, G. W., GETAHUN, A. & CRESS, D. C. 1971. Resistance in barley to the greenbug, *Schizaphis graminum*. 1. Toxicity of phenolic and flavonoid compounds and related substances. *Annals of the Entomological Society of America*, 64, 718-722.
- TOKUNAGA, N., KANETA, T., SATO, S. & SATO, Y. 2009. Analysis of expression profiles of three peroxidase genes associated with lignification in *Arabidopsis thaliana*. *Physiol Plant*, 136, 237-49.

- TRINGE, S. G., VON MERING, C., KOBAYASHI, A., SALAMOV, A. A., CHEN, K., CHANG, H. W., PODAR, M., SHORT, J. M., MATHUR, E. J., DETTER, J. C., BORK, P., HUGENHOLTZ, P. & RUBIN, E. M. 2005. Comparative metagenomics of microbial communities. *Science*, 308, 554-7.
- TRONTIN, C., KIANI, S., CORWIN, J. A., HEMATY, K., YANSOUNI, J., KLIEBENSTEIN, D. J. & LOUDET, O. 2014. A pair of receptor-like kinases is responsible for natural variation in shoot growth response to mannitol treatment in *Arabidopsis thaliana*. *Plant J*, 78, 121-33.
- TSAI, A. Y., OOTA, M. & SAWA, S. 2020. Chemotactic Host-Finding Strategies of Plant Endoparasites and Endophytes. *Front Plant Sci*, 11, 1167.
- TSCHAPLINSKI, T. J., STANDAERT, R. F., ENGLE, N. L., MARTIN, M. Z., SANGHA, A. K., PARKS, J. M., SMITH, J. C., SAMUEL, R., JIANG, N., PU, Y., RAGAUSKAS, A. J., HAMILTON, C. Y., FU, C., WANG, Z. Y., DAVISON, B. H., DIXON, R. A. & MIELENZ, J. R. 2012. Down-regulation of the caffeic acid O-methyltransferase gene in switchgrass reveals a novel monolignol analog. *Biotechnol Biofuels*, 5, 71.
- TSVETKOV, D., KOLPAKOV, E., KASSMANN, M., SCHUBERT, R. & GOLLASCH, M. 2019. Distinguishing Between Biological and Technical Replicates in Hypertension Research on Isolated Arteries. *Front Med (Lausanne)*, 6, 126.
- TURNER, S. R. & SOMERVILLE, C. R. 1997. Collapsed xylem phenotype of *Arabidopsis* identifies mutants deficient in cellulose deposition in the secondary cell wall. *Plant Cell*, 9, 689-701.
- TYANOVA, S., TEMU, T., SINITYN, P., CARLSON, A., HEIN, M. Y., GEIGER, T., MANN, M. & COX, J. 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*, 13, 731-40.
- UHRIG, R. G., SCHLAPFER, P., ROSCHITZKI, B., HIRSCH-HOFFMANN, M. & GRUISSEM, W. 2019. Diurnal changes in concerted plant protein phosphorylation and acetylation in *Arabidopsis* organs and seedlings. *Plant J*, 99, 176-194.
- UMEZAWA, T., YOSHIDA, R., MARUYAMA, K., YAMAGUCHI-SHINOZAKI, K. & SHINOZAKI, K. 2004. SRK2C, a SNF1-related protein kinase 2, improves drought tolerance by controlling stress-responsive gene expression in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*, 101, 17306-11.
- URBAN, J. 2022. A review on recent trends in the phosphoproteomics workflow. From sample preparation to data analysis. *Anal Chim Acta*, 1199, 338857.
- VAAHTERA, L., SCHULZ, J. & HAMANN, T. 2019. Cell wall integrity maintenance during plant development and interaction with the environment. *Nature plants*, 5, 924-932.

- VAINONEN, J. P., HANSSON, M. & VENER, A. V. 2005. STN8 protein kinase in *Arabidopsis thaliana* is specific in phosphorylation of photosystem II core proteins. *J Biol Chem*, 280, 33679-86.
- VAN DAM, N. M. & BOUWMEESTER, H. J. 2016. Metabolomics in the Rhizosphere: Tapping into Belowground Chemical Communication. *Trends Plant Sci*, 21, 256-265.
- VAN DAMME, E. J., LANNOO, N. & PEUMANS, W. J. 2008. Plant lectins. *Advances in botanical research*. Elsevier.
- VAN DER ARK, K. C. H., VAN HECK, R. G. A., MARTINS DOS SANTOS, V. A. P., BELZER, C. & DE VOS, W. M. 2017. More than just a gut feeling: constraint-based genome-scale metabolic models for predicting functions of human intestinal microbes. *Microbiome*, 5, 78.
- VAN NOORT, J. M., KRAAL, B., SINJORGIO, K. M., PERSON, N. L., JOHANNIS, E. S. & BOSCH, L. J. E. J. O. B. 1986. Methylation in vivo of elongation factor EF-Tu at lysine-56 decreases the rate of tRNA-dependent GTP hydrolysis. 160, 557-561.
- VAN STAALDUINEN, L. M. & JIA, Z. 2015. Post-translational hydroxylation by 2OG/Fe (II)-dependent oxygenases as a novel regulatory mechanism in bacteria. *J Frontiers in Microbiology*, 5, 798.
- VANHOLME, B., EL HOUARI, I. & BOERJAN, W. 2019. Bioactivity: phenylpropanoids' best kept secret. *Curr Opin Biotechnol*, 56, 156-162.
- VANHOLME, R., CESARINO, I., RATAJ, K., XIAO, Y., SUNDIN, L., GOEMINNE, G., KIM, H., CROSS, J., MORREEL, K., ARAUJO, P., WELSH, L., HAUSTRAETE, J., MCCLELLAN, C., VANHOLME, B., RALPH, J., SIMPSON, G. G., HALPIN, C. & BOERJAN, W. 2013. Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. *Science*, 341, 1103-6.
- VANHOLME, R., DEMEDTS, B., MORREEL, K., RALPH, J. & BOERJAN, W. 2010. Lignin biosynthesis and structure. *Plant Physiol*, 153, 895-905.
- VELVIZHI, G., BALAKUMAR, K., SHETTI, N. P., AHMAD, E., KISHORE PANT, K. & AMINABHAVI, T. M. 2022. Integrated biorefinery processes for conversion of lignocellulosic biomass to value added materials: Paving a path towards circular economy. *Bioresour Technol*, 343, 126151.
- VENABLE, J. D., DONG, M. Q., WOHLSCHEGEL, J., DILLIN, A. & YATES, J. R. 2004. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat Methods*, 1, 39-45.
- VENNE, A. S., KOLLIPARA, L. & ZAHEDI, R. P. 2014. The next level of complexity: crosstalk of posttranslational modifications. *Proteomics*, 14, 513-24.

- VENTURI, V. & KEEL, C. 2016. Signaling in the Rhizosphere. *Trends Plant Sci*, 21, 187-198.
- VILLALOBOS SOLIS, M. I., GIANNONE, R. J., HETTICH, R. L. & ABRAHAM, P. E. 2019. Exploiting the Dynamic Relationship between Peptide Separation Quality and Peptide Coisolation in a Multiple-Peptide Matches-per-Spectrum Approach Offers a Strategy To Optimize Bottom-Up Proteomics Throughput and Depth. *Anal Chem*, 91, 7273-7279.
- VILLALOBOS SOLIS, M. I., POUDEL, S., BONNOT, C., SHRESTHA, H. K., HETTICH, R. L., VENEULT-FOURREY, C., MARTIN, F. & ABRAHAM, P. E. 2020. A Viable New Strategy for the Discovery of Peptide Proteolytic Cleavage Products in Plant-Microbe Interactions. *Mol Plant Microbe Interact*, 33, 1177-1188.
- VON MERING, C., HUYNEN, M., JAEGGI, D., SCHMIDT, S., BORK, P. & SNEL, B. 2003. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res*, 31, 258-61.
- VYSE, S., DESMOND, H. & HUANG, P. H. 2017. Advances in mass spectrometry based strategies to study receptor tyrosine kinases. *IUCrJ*, 4, 119-130.
- WAAS, M., PERECKAS, M., JONES LIPINSKI, R. A., ASHWOOD, C. & GUNDRY, R. L. 2019. SP2: Rapid and Automatable Contaminant Removal from Peptide Samples for Proteomic Analyses. *J Proteome Res*, 18, 1644-1656.
- WADHAMS, G. H. & ARMITAGE, J. P. 2004. Making sense of it all: bacterial chemotaxis. *Nature reviews Molecular cell biology*, 5, 1024-1037.
- WALDIE, T., MCCULLOCH, H. & LEYSER, O. 2014. Strigolactones and the control of plant development: lessons from shoot branching. *Plant J*, 79, 607-22.
- WALSH, C. T., GARNEAU-TSODIKOVA, S. & GATTO, G. J., JR. 2005. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl*, 44, 7342-72.
- WANG, E., SCHORNACK, S., MARSH, J. F., GOBBATO, E., SCHWESSINGER, B., EASTMOND, P., SCHULTZE, M., KAMOUN, S. & OLDROYD, G. E. 2012. A common signaling process that promotes mycorrhizal and oomycete colonization of plants. *Current Biology*, 22, 2242-2246.
- WANG, J., CARPER, D. L., BURDICK, L. H., SHRESTHA, H. K., APPIDI, M. R., ABRAHAM, P. E., ... & DOKTYCZ, M. J. 2021. Formation, characterization and modeling of emergent synthetic microbial communities. . *Computational and Structural Biotechnology Journal*.
- WANG, N., TANG, C., FAN, X., HE, M., GAN, P., ZHANG, S., HU, Z., WANG, X., YAN, T., SHU, W., YU, L., ZHAO, J., HE, J., LI, L., WANG, J., HUANG, X., HUANG, L., ZHOU, J. M., KANG, Z. & WANG, X. 2022. Inactivation of a wheat protein kinase gene confers broad-spectrum resistance to rust fungi. *Cell*, 185, 2961-2974 e19.

- WASHBURN, M. P., WOLTERS, D. & YATES, J. R., 3RD 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*, 19, 242-7.
- WATHUGALA, D. L., HEMSLEY, P. A., MOFFAT, C. S., CREMELIE, P., KNIGHT, M. R. & KNIGHT, H. 2012. The Mediator subunit SFR6/MED16 controls defence gene expression mediated by salicylic acid and jasmonate responsive pathways. *New Phytol*, 195, 217-30.
- WEBB-ROBERTSON, B. J., WIBERG, H. K., MATZKE, M. M., BROWN, J. N., WANG, J., MCDERMOTT, J. E., SMITH, R. D., RODLAND, K. D., METZ, T. O., POUNDS, J. G. & WATERS, K. M. 2015. Review, evaluation, and discussion of the challenges of missing value imputation for mass spectrometry-based label-free global proteomics. *J Proteome Res*, 14, 1993-2001.
- WEILAND-BRAUER, N. 2021. Friends or Foes-Microbial Interactions in Nature. *Biology (Basel)*, 10.
- WEINER, H. L. & ZAGZAG, D. 2000. Growth factor receptor tyrosine kinases: cell adhesion kinase family suggests a novel signaling mechanism in cancer. *Cancer Invest*, 18, 544-54.
- WELLER, D. M., MAVRODI, D. V., VAN PELT, J. A., PIETERSE, C. M., VAN LOON, L. C. & BAKKER, P. A. 2012. Induced systemic resistance in *Arabidopsis thaliana* against *Pseudomonas syringae* pv. tomato by 2,4-diacetylphloroglucinol-producing *Pseudomonas fluorescens*. *Phytopathology*, 102, 403-12.
- WHETTEN, R. & SEDEROFF, R. 1995. Lignin Biosynthesis. *Plant Cell*, 7, 1001-1013.
- WILKINS, M. R., SANCHEZ, J. C., GOOLEY, A. A., APPEL, R. D., HUMPHERY-SMITH, I., HOCHSTRASSER, D. F. & WILLIAMS, K. L. 1996. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev*, 13, 19-50.
- WIRBEL, J., CUTILLAS, P. & SAEZ-RODRIGUEZ, J. 2018. Phosphoproteomics-Based Profiling of Kinase Activities in Cancer Cells. *Methods Mol Biol*, 1711, 103-132.
- WIREDJA, D. D., KOYUTURK, M. & CHANCE, M. R. 2017. The KSEA App: a web-based tool for kinase activity inference from quantitative phosphoproteomics. *Bioinformatics*, 33, 3489-3491.
- WITHERS, S., LU, F., KIM, H., ZHU, Y., RALPH, J. & WILKERSON, C. G. 2012. Identification of grass-specific enzyme that acylates monolignols with p-coumarate. *J Biol Chem*, 287, 8347-55.
- WOHLBRAND, L., TRAUTWEIN, K. & RABUS, R. 2013. Proteomic tools for environmental microbiology--a roadmap from sample preparation to protein identification and quantification. *Proteomics*, 13, 2700-30.

- WONG, W. S., GUO, D., WANG, X. L., YIN, Z. Q., XIA, B. & LI, N. 2005. Study of cis-cinnamic acid in *Arabidopsis thaliana*. *Plant Physiol Biochem*, 43, 929-37.
- XIE, M., ZHANG, J., TSCHAPLINSKI, T. J., TUSKAN, G. A., CHEN, J. G. & MUCHERO, W. 2018. Regulation of Lignin Biosynthesis and Its Role in Growth-Defense Tradeoffs. *Front Plant Sci*, 9, 1427.
- XUE, J., LUO, D., XU, D., ZENG, M., CUI, X., LI, L. & HUANG, H. 2015. CCR1, an enzyme required for lignin biosynthesis in *Arabidopsis*, mediates cell proliferation exit for leaf development. *Plant J*, 83, 375-87.
- YAMAZAKI, S., MARDANI-KORRANI, H., KAIDA, R., OCHIAI, K., KOBAYASHI, M., NAGANO, A. J., FUJII, Y., SUGIYAMA, A. & AOKI, Y. 2021. Field multi-omics analysis reveals a close association between bacterial communities and mineral properties in the soybean rhizosphere. *Sci Rep*, 11, 8878.
- YANG, L., ZHAO, X., RAN, L., LI, C., FAN, D. & LUO, K. 2017. PtoMYB156 is involved in negative regulation of phenylpropanoid metabolism and secondary cell wall biosynthesis during wood formation in poplar. *Sci Rep*, 7, 41209.
- YANG, X., MEDFORD, J. I., MARKEL, K., SHIH, P. M., DE PAOLI, H. C., TRINH, C. T., ... & TUSKAN, G. A. 2020. Plant Biosystems Design Research Roadmap 1.0. *BioDesign Research*.
- YAO, X., FREAS, A., RAMIREZ, J., DEMIREV, P. A. & FENSELAU, C. 2001. Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem*, 73, 2836-42.
- YOUNG, C. C. & BERNLOHR, R. W. 1991. Elongation Factor-Tu Is Methylated in Response to Nutrient Deprivation in *Escherichia-Coli*. *Journal of Bacteriology*, 173, 3096-3100.
- YU, D., CHEN, C. & CHEN, Z. 2001. Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression. *Plant Cell*, 13, 1527-40.
- ZAMIOUDIS, C. & PIETERSE, C. M. 2012. Modulation of host immunity by beneficial microbes. *Mol Plant Microbe Interact*, 25, 139-50.
- ZEILINGER, S., GUPTA, V. K., DAHMS, T. E., SILVA, R. N., SINGH, H. B., UPADHYAY, R. S., GOMES, E. V., TSUI, C. K. & NAYAK, S. C. 2016. Friends or foes? Emerging insights from fungal interactions with plants. *FEMS Microbiol Rev*, 40, 182-207.
- ZHANG, H., SUN, Y., XIE, X., KIM, M. S., DOWD, S. E. & PARE, P. W. 2009. A soil bacterium regulates plant acquisition of iron via deficiency-inducible mechanisms. *Plant J*, 58, 568-77.

- ZHANG, J., TUSKAN, G. A., TSCHAPLINSKI, T. J., MUCHERO, W. & CHEN, J. G. 2020a. Transcriptional and Post-transcriptional Regulation of Lignin Biosynthesis Pathway Genes in Populus. *Front Plant Sci*, 11, 652.
- ZHANG, J., XIE, M., LI, M., DING, J., PU, Y., BRYAN, A. C., ROTTMANN, W., WINKELER, K. A., COLLINS, C. M., SINGAN, V., LINDQUIST, E. A., JAWDY, S. S., GUNTER, L. E., ENGLE, N. L., YANG, X., BARRY, K., TSCHAPLINSKI, T. J., SCHMUTZ, J., TUSKAN, G. A., MUCHERO, W. & CHEN, J. G. 2020b. Overexpression of a Prefoldin beta subunit gene reduces biomass recalcitrance in the bioenergy crop Populus. *Plant Biotechnol J*, 18, 859-871.
- ZHANG, J., XIN, L., SHAN, B., CHEN, W., XIE, M., YUEN, D., ZHANG, W., ZHANG, Z., LAJOIE, G. A. & MA, B. 2012. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics*, 11, M111 010587.
- ZHANG, K., HALITSCHKE, R., YIN, C., LIU, C. J. & GAN, S. S. 2013. Salicylic acid 3-hydroxylase regulates Arabidopsis leaf longevity by mediating salicylic acid catabolism. *Proc Natl Acad Sci U S A*, 110, 14807-12.
- ZHANG, L., JIA, Y., WANG, L. & FANG, R. 2007. A proline iminopeptidase gene upregulated in planta by a LuxR homologue is essential for pathogenicity of Xanthomonas campestris pv. campestris. *Mol Microbiol*, 65, 121-36.
- ZHANG, N., AEBERSOLD, R. & SCHWIKOWSKI, B. 2002. ProbID: a probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data. *Proteomics*, 2, 1406-12.
- ZHANG, R., VIVANCO, J. M. & SHEN, Q. 2017. The unseen rhizosphere root-soil-microbe interactions for crop production. *Curr Opin Microbiol*, 37, 8-14.
- ZHANG, W. P., SUN, J., CAO, H. L., TIAN, R. M., CAI, L., DING, W. & QIAN, P. Y. 2016. Post-translational modifications are enriched within protein functional groups important to bacterial adaptation within a deep-sea hydrothermal vent environment. *Microbiome*, 4.
- ZHANG, X., XIAO, S., RAMEAU, R. D., DEVANY, E., NADEEM, Z., CAGLAR, E., NG, K., KLEIMAN, F. E. & SAXENA, A. 2018. Nucleolin phosphorylation regulates PARN deadenylase activity during cellular stress response. *RNA Biol*, 15, 251-260.
- ZHANG, Y., RUYTER-SPIRA, C. & BOUWMEESTER, H. J. 2015. Engineering the plant rhizosphere. *Curr Opin Biotechnol*, 32, 136-142.
- ZHANG, Y., XU, J., WANG, E. & WANG, N. 2020c. Mechanisms Underlying the Rhizosphere-To-Rhizoplane Enrichment of Cellvibrio Unveiled by Genome-Centric Metagenomics and Metatranscriptomics. *Microorganisms*, 8.

- ZHAO, J., ZHANG, W., ZHAO, Y., GONG, X., GUO, L., ZHU, G., WANG, X., GONG, Z., SCHUMAKER, K. S. & GUO, Y. 2007. SAD2, an importin -like protein, is required for UV-B response in Arabidopsis by mediating MYB4 nuclear trafficking. *Plant Cell*, 19, 3805-18.
- ZHAO, Q. & DIXON, R. A. 2011. Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends Plant Sci*, 16, 227-33.
- ZHAO, Q. & DIXON, R. A. 2014. Altering the cell wall and its impact on plant disease: from forage to bioenergy. *Annu Rev Phytopathol*, 52, 69-91.
- ZHAO, Q., NAKASHIMA, J., CHEN, F., YIN, Y., FU, C., YUN, J., SHAO, H., WANG, X., WANG, Z. Y. & DIXON, R. A. 2013. Laccase is necessary and nonredundant with peroxidase for lignin polymerization during vascular development in Arabidopsis. *Plant Cell*, 25, 3976-87.
- ZHOU, J., LIU, W., DENG, Y., JIANG, Y. H., XUE, K., HE, Z., VAN NOSTRAND, J. D., WU, L., YANG, Y. & WANG, A. 2013. Stochastic assembly leads to alternative communities with distinct functions in a bioreactor microbial community. *mBio*, 4.
- ZHOU, J., THOMPSON, D. K., XU, Y. & TIEDJE, J. M. 2004. *Microbial functional genomics*, Wiley Online Library.
- ZHOU, R., JACKSON, L., SHADLE, G., NAKASHIMA, J., TEMPLE, S., CHEN, F. & DIXON, R. A. 2010. Distinct cinnamoyl CoA reductases involved in parallel routes to lignin in *Medicago truncatula*. *Proc Natl Acad Sci U S A*, 107, 17803-8.
- ZHOU, Y. & SHAO, H. B. 2008. The responding relationship between plants and environment is the essential principle for agricultural sustainable development on the globe. *C R Biol*, 331, 321-8.
- ZHUO, C., WANG, X., DOCAMPO-PALACIOS, M., SANDERS, B. C., ENGLE, N. L., TSCHAPLINSKI, T. J., HENDRY, J. I., MARANAS, C. D., CHEN, F. & DIXON, R. A. 2022. Developmental changes in lignin composition are driven by both monolignol supply and laccase specificity. *Sci Adv*, 8, eabm8145.
- ZOGHLAMI, A. & PAES, G. 2019. Lignocellulosic Biomass: Understanding Recalcitrance and Predicting Hydrolysis. *Frontiers in chemistry*, 7, 874.

VITA

Him K. Shrestha was born in Pokhara, Nepal. He graduated from Novel Academy with an Advance level degree in 2013. He moved to the United State for his college education. He earned a Bachelor of Science in Nanoscience at Northwest Missouri State University. During his undergraduate, he was introduced to mass spectrometry and its biological application. Following the completion of his B.S. in 2018, he enrolled in the graduate school of Genome Science and Technology (GST) at the University of Tennessee, Knoxville (UTK). While in graduate school, he worked in biological mass spectrometry under Dr. Robert Hettich at Oak Ridge National Laboratory. He expects to receive his Ph.D. in December 2022.