

**DEVELOPING AND APPLYING MASS SPECTROMETRY-BASED OMICS TO STUDY
PLANT-MICROBIOME INTERACTIONS IN *POPULUS* SPP. USING A
REDUCTIONIST APPROACH**

A DISSERTATION PRESENTED FOR THE

DOCTOR OF PHILOSOPHY

DEGREE

THE UNIVERSITY OF TENNESSEE, KNOXVILLE

MANASA REDDY APPIDI

DECEMBER 2024

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Dr. Robert Hettich, for his unwavering guidance and support throughout my PhD journey. His mentorship, encouragement and understanding during challenging times have been invaluable, and I am deeply appreciative of the opportunities he has provided me over the past five years.

I am also profoundly grateful to Dr. Paul Abraham for his mentorship and insightful guidance, which were instrumental in the completion of my dissertation. My sincere thanks go to the remaining members of my PhD committee- Dr. Jennifer Morrell-Falvey, Dr. Gladys Alexandre, Dr. Margaret Staton for their valuable feedback and support throughout this process. I would also like to thank the Genome science and technology (GST) program and all my project collaborators.

I extend my thanks to the other staff members and students of the Hettich lab group for their invaluable contributions to my PhD journey. Their constructive discussions, thoughtful suggestions and willingness to share their expertise created a collaborative environment that greatly enriched my PhD experience.

I am also deeply appreciative of the support and friendship of my GST cohort. The shared experiences, insightful conversations and constant encouragement fostered a sense of community that I will always cherish.

Finally, I am immensely thankful for the unwavering love and support of my family, who have been my constant source of strength throughout my education and career. A special thanks to my husband, Sandeep Kannuri, for his patience, encouragement, and steadfast belief in me during the challenges of my PhD journey. His support has been the foundation of my success, and I am forever grateful. To my son, Aadvik, who fills my life with immense joy and inspires me to strive for excellence every day- thank you for being my light and my motivation to achieve my best.

ABSTRACT

Microbes are ubiquitously associated with plants and play a crucial role in plant's health and fitness. Due to their pivotal roles in plant health, it is important to understand how the microbiomes assemble and establish themselves in various plant niches. Of the various plant niches, the rhizosphere serves as a reservoir of beneficial microbes which are selectively recruited by the plant. The molecular mechanisms that facilitate these associations between the plant and microbes are not fully known. Studying plant-microbe interactions in nature is challenging due to the complexity and technical limitations. However, reductionist approaches using constructed microbial communities and laboratory-controlled experiments have proven to be valuable in advancing our knowledge of plant-microbe interactions as they simplify the complexity of natural systems and thereby allowing us to focus on specific interactions. High-throughput omics technologies provide invaluable information in studying these complex and dynamic plant microbiomes. Metaproteomics can identify and quantify thousands of proteins and provide insights into microbiome composition and functions of individual members of the microbiome. Metabolomics helps in studying the chemicals released by plants and microbes that serve as nutrients and/or chemical signals to establish interactions with each other. In this dissertation, liquid chromatography coupled with mass spectrometry-based metaproteomics and metabolomics approaches have been used to advance our understanding of interactions between the *Populus* spp. and its rhizosphere microbiome. The research presented here describes (1) application of metaproteomics in understanding microbial functions in a constructed microbial community and assembly of stable microbiomes, (2) methodological advancements to achieve spatially resolved metaproteomes along plant root axis, (3) understanding microbial functions and responses to plant root chemicals using metaproteomics, (4) application of untargeted metabolomics to characterize root and exudate metabolite profiles of *Populus trichocarpa*. In summary, this dissertation has contributed significantly to understanding the responses of rhizosphere bacterial species associated with *Populus* spp., development of novel methods with new capabilities and experimental platforms that can be easily applied to other plant-microbe studies.

TABLE OF CONTENTS

Chapter 1: Introduction to understanding plant-microbe interactions using liquid chromatography coupled mass spectrometry-based metaproteomics and metabolomics.....	1
Plant-microbe interactions	1
Importance of studying <i>Populus</i> spp. and its associated microbiome	2
Rhizosphere microbiome	2
Bridging the holistic and reductionist approaches to study plant associated microbiome	4
Omics technologies to study plant-associated microbiomes	7
Characterizing plant microbiome functions using metaproteomics.....	9
Metabolomics to study the metabolites produced by plant and microbes in plant holobiont...	10
Overview of dissertation	11
Chapter 2: Methodologies for mass spectrometry-based metaproteomics and untargeted metabolomics analysis	14
Overview of Liquid chromatography coupled mass spectrometry (LC-MS/MS)	14
Liquid chromatography.....	14
One-dimensional liquid chromatography	15
Multidimensional chromatography	16
Mass spectrometry	18
Ionization sources	18
Electrospray ionization (ESI).....	19
Mass analyzer and detector	21
Analytical figures of merit.....	22
Q Exactive plus mass spectrometer	23
Tandem mass spectrometry MS/MS and data acquisition	25
General considerations for LC-MS/MS experiments	26
LC-MS/MS based proteomics and metaproteomics	29
Bottom-up proteomics workflow	30
Sample collection and sample preparation for proteomics	30
Peptide separation using liquid chromatography followed by tandem mass spectrometry analysis.....	33

Data processing and downstream analysis of proteomics data	34
Reference database search for peptide identification and quantification.....	34
Data processing and statistical analysis	35
LC-MS/MS based metabolomics	36
Untargeted metabolomics workflow	37
Sample preparation for metabolomics	37
LC-MS/MS measurements for Untargeted metabolomics.....	39
Untargeted metabolomics data processing and analysis	40
Summary	42
Chapter 3: A Metaproteomic approach to study microbial community structure and inter- microbial interactions in constructed microbial communities	43
Author contributions	43
Introduction.....	43
Materials and methods	48
Preparation of 3-member communities with varying initial ratios	48
Cellular protein extraction for proteomics	50
Protein identification and quantification.....	50
Statistical analysis	51
Results.....	51
Differential proteomic analysis reveals enhanced metabolic activity in synthetic microbial community compared to monoculture in nutrient-rich conditions	51
The influence of various initial inoculum ratios on the resulting community structures and temporal dynamics assessed by metaproteomics	59
Increased metabolic activity of <i>Pseudomonas</i> sp. GM17 contributes to its dominance in the community	61
Conclusions.....	64
Chapter 4: Development of an experimental approach to achieve spatially resolved plant root- associated metaproteomics using an agar-plate system	66
Author contributions	66
Introduction.....	66

Materials and methods	69
Preparation of plants, bacterial cultures, and a SynCom	69
Plant-agar culture growth conditions	69
Extracting microbiomes from agar for metaproteomics	72
Protein identification and quantification.....	74
Statistical analysis	75
Results and discussion	76
A novel approach to extract and characterize root-associated microbiomes using metaproteomics	76
Spatially resolved metaproteomics along the plant root axis.....	81
Detecting functional changes in microbial activity along the plant root axis.....	85
Conclusions.....	89
Chapter 5: Unraveling how bacterial adaptation strategies in the microbiome are shaped by the chemical environment of the plant rhizosphere	93
Author contributions	93
Introduction.....	93
Materials and methods	96
Preparation of plants, bacterial cultures, and growth conditions	96
Expression of plant-responsive promoters on <i>Populus</i> roots using filter-based system.....	96
Bacterial inoculation on <i>Populus</i> roots using a filter-based system	98
Cellular protein extraction for proteomic analyses	98
Protein identification and quantification.....	99
Statistical analysis.....	100
Root colonization assays.....	101
Results.....	101
A filter-based system enables the study of bacterial responses to the plant root chemical environment without physical contact with the plant	101
Quantification of bacterial proteomes revealed varying responses to the presence of a plant by individual isolates	103

The plant-associated chemical environment impacts the microbiome population structure	106
Bacteria shift metabolic function at a taxon-specific level in response to the plant root chemical environment	109
<i>Bacillus</i> sp. BC15 becomes metabolically active in the presence of a plant	116
<i>Pantoea</i> sp. YR343 shows increased abundance of proteins related to chemotaxis, motility, nutrient transport, and metabolism in response to the plant root chemical environment	118
Discussion	120
Chapter 6: Untargeted metabolomic analysis to Characterize the dynamic root and exudate metabolomes of <i>Populus trichocarpa</i>	125
Author contributions	125
Introduction	125
Material and methods	127
<i>Populus</i> genotypes and growth conditions	127
Experimental design	127
Metabolite extraction from root tissue	128
Metabolite extraction from agar plugs for root exudates	130
Metabolite analysis using Liquid chromatography- tandem mass spectrometry (LC-MS/MS)	130
Untargeted metabolomics data processing and analysis	131
Results and discussion	132
Global results of untargeted metabolomic analysis and data processing to obtain non-redundant metabolite feature lists	132
Multivariate analysis of metabolite profiles to assess effects of various experimental factors	137
Comparison of metabolite profiles between genotypes	141
Comparison of intracellular and extracellular metabolite profiles	143
Assessing influence of nutrient condition on root and exudate profiles of BESC-24	148
Conclusions	151
Chapter 7: Conclusions and future perspectives	154

Conclusions.....	154
Remaining challenges, limitations and future perspectives	157
References.....	161
Vita.....	189

LIST OF TABLES

Table 4.1: LIMMA results using the relative abundances of the individual taxa comparing the agar sample to the root wash.....	80
Table 4.2: ANOVA results for each member in the 10-member SynCom based on the relative abundances over time.....	84
Table 4.3: PERMANOVA results of Euclidean distances of proteins.....	87
Table 4.4: ANOVA results for each member in the 10-member SynCom based on the relative abundances at the 48h time across spatial locations.	88
Table 5.1: Bacterial strains utilized in this study	97
Table 6.1: Chemical classification results for the annotated metabolite features	138
Table 6.2: PERMANOVA results of Bray Curtis distances of metabolomics data.....	139

LIST OF FIGURES

Figure 1.1: Interactions in the rhizosphere.	3
Figure 1.2: Experimental approaches to study plant-microbe interactions.....	5
Figure 1.3: Omics approaches for studying plant associated microbiomes.....	8
Figure 2.1: Multidimensional liquid chromatography column.....	17
Figure 2.2: Schematic illustration of electrospray ionization (ESI) process	20
Figure 2.3: Schematic of Q Exactive plus hybrid quadrupole-orbitrap mass spectrometer.	24
Figure 2.4: Overview of data acquisition strategies in tandem mass spectrometry (MS/MS). ...	27
Figure 2.5: Overview of a typical bottom-up proteomics workflow	31
Figure 2.6: Overview of untargeted metabolomics workflow.	38
Figure 3.1: Three-member communities with varying initial inoculum ratios.	49
Figure 3.2: Qualitative assessment of proteomes of bacterial monocultures.....	52
Figure 3.3: Qualitative assessment of proteomes of bacterial strains in community.....	52
Figure 3.4: Overlap of proteins between monocultures and 1:1:1 community at 48 hr timepoint.	54
Figure 3.5: Volcano plots showing the significant differentially abundant proteins for each bacterial species at 48 hr timepoint in the three-member community with equal initial inoculum ratios (GM17-YR343-AP49 1-1-1) compared to their monocultures at early stationary phase in R2A medium	56
Figure 3.6: Global KEGG map of metabolic pathways for <i>Pseudomonas</i> sp. GM17	57
Figure 3.7: Global KEGG map of metabolic pathways for <i>Pantoea</i> sp. YR343	58
Figure 3.8: Assessment of population sizes of bacterial strains in three-member communities..	60
Figure 3.9: Functional analysis of <i>Pseudomonas</i> sp. GM17 in community with equal ratios.	62
Figure 3.10: Gene ontology enrichment results for <i>Pseudomonas</i> sp. in YR343- GM17-AP49 1: 0.001:1 community.	63
Figure 4.1: Experimental design to achieve spatially resolved plant root-associated microbes using an agar system.	71
Figure 4.2: Proteins extracted from root-associated microbiomes using different extraction methods for plant-agar samples at 48 h.	77
Figure 4.3: Global proteomics metrics using novel agar extraction method.	78

Figure 4.4: Assessment of microbial relative population sizes based on their summed protein abundances.....	82
Figure 4.5: Assessment of microbial relative population sizes based on their summed protein abundances.....	83
Figure 4.6: Association with plant shows higher variation in proteome expression of synthetic community.....	86
Figure 4.7: Hierarchical clustering results for <i>Pantoea</i> sp. strain YR343.	90
Figure 5.1: Overview of experimental design and proof-of-concept.....	102
Figure 5.2: Average number of proteins identified (n=3) in proteomic analysis of ten bacterial strains grown individually in the absence or presence of the plant.	104
Figure 5.3: Stacked bar charts showing the number of differentially abundant proteins for ten bacterial strains grown individually.....	105
Figure 5.4: Overview of number of proteins identified in metaproteomics analysis for ten bacterial strains grown in a mixed community in the absence or presence of a plant.	107
Figure 5.5: Relative population sizes of ten bacterial strains in the mixed community	108
Figure 5.6: Stacked bar charts showing the number of differentially abundant proteins for ten bacterial strains grown in a mixed community	110
Figure 5.7: Functional analysis of bacterial strains grown individually in the absence or presence of a plant using Clusters of Orthologous Groups (COGs).....	111
Figure 5.8: Functional analysis of bacterial strains grown in a mixed community in the absence or presence of a plant using Clusters of Orthologous Groups (COGs).....	113
Figure 5.9: Stacked bar charts illustrating the distribution of identified ABC transporters in ten bacterial strains.	115
Figure 5.10: Functional changes in <i>Bacillus</i> sp. BC15 as a response to the chemical environment of the plant root.....	117
Figure 5.11: Proteomic analysis of <i>Pantoea</i> sp. YR343 in response to the chemical environment of a plant root.	119
Figure 6.1: Schematic representation of the experimental design.	129
Figure 6.2: Number of metabolite features	134

Figure 6.3: Box plots showing Log2 abundance values across experimental conditions for most abundant identified compounds in each extraction solvent	136
Figure 6.4: Nonmetric multidimensional scaling (NMDS) plots based on Bray Curtis distances of metabolomics data	140
Figure 6.5: Upset plots showing overlap of compounds between the genotypes	142
Figure 6.6: Root trait data	144
Figure 6.7: Box plots showing Log2 abundance values across experimental conditions for Loliolide, an important metabolite feature in ethyl acetate organic fraction differentiating between the genotypes based on PLS-DA analysis	145
Figure 6.8: Upset plots showing overlap of compounds between the root and exudates samples	147
Figure 6.9: Box plots showing Log2 abundance values across experimental conditions for important metabolite features differentiating between root and exudate samples based on PLS-DA analysis	149
Figure 6.10: Box plots showing Log2 abundance values for metabolites identified in BESC-24 genotype root samples and displayed changes in abundance in response to nutrient condition-timepoint interactions.....	150
Figure 6.11: Box plots showing Log2 abundance values for metabolites identified in BESC-24 genotype exudate samples and displayed changes in abundance in response to nutrient condition-timepoint interactions.....	152

CHAPTER 1: INTRODUCTION TO UNDERSTANDING PLANT-MICROBE INTERACTIONS USING LIQUID CHROMATOGRAPHY COUPLED MASS SPECTROMETRY-BASED METAPROTEOMICS AND METABOLOMICS

Plant-microbe interactions

A plethora of microbial communities are hosted by plants and interacted with continuously in nature. These interactions include symbiotic associations that are mutually beneficial to both microbe and plant host, commensal associations in which only one species benefits while the other one neither benefits nor is harmed and pathogenic associations that are harmful to plant hosts¹. The beneficial interactions between plants and microbes are known to improve plant growth and survival by conferring multiple advantages to their plant host, such as increasing nutritional acquisition², overcoming abiotic stresses³ and activating defense pathways against pathogens⁴. The plant fitness is a consequence of associations between plant and microbes that form a complex, multi-organism system, which collectively forms a plant holobiont⁵. In the recent years, there has been an increased focus on studying the plant holobiont to unravel the interdependencies between the plant and their microbiome⁶, as the knowledge gained from these studies has the potential to drive innovations in plant productivity for sustainable agriculture^{7, 8}. The growing recognition of plant holobiont concept raises several important questions to be answered such as what are the driving factors of microbiome assembly, the processes involved in plants differentiating between symbiotic and pathogenic microbes, effects of abiotic stressors on plant-microbe interactions and effects of microbe-microbe interactions on plant-microbe interactions^{5, 9} and consequently requires application of advanced omics technologies to address them. The structure and composition of the plant holobiont undergoes changes due to inherent dynamic interactions with the biotic and abiotic factors of the environment¹⁰, during which the plant and microbes work together to adapt to the changing environmental conditions. Studying the underlying genetic, molecular, and metabolic relationships between plants and associated microbes across various environmental conditions would provide better understanding of the plant holobiont. This knowledge is crucial for enhancing plant health and fitness by developing strategies for improved and sustainable agricultural practices.

Importance of studying *Populus* spp. and its associated microbiome

Populus spp. is an economically important, leading candidate plant for bioenergy production^{11, 12}. It is a woody perennial organism and has become a model organism in plant biology due to its broad distribution across North America, fast growth, vegetative propagation potential and the relative ease with which it can be manipulated for experiments¹²⁻¹⁴. *Populus trichocarpa* was the first tree species that had its genome sequenced¹⁴. *Populus* roots and its rhizosphere serves as a hub for diverse interactions between the plant host, bacteria and fungi, which makes it an excellent model system to study the plant-microbe interactions¹⁵. Additionally, a culture collection of 3,211 unique bacterial isolates is available for investigating the interactions between *Populus* and its associated bacterial strains¹⁶. The fundamental understanding of interactions between *Populus* and its associated microbiome will provide valuable information for improving the plant health and growth which will be helpful for bioenergy production, carbon cycling and sequestration.

Rhizosphere microbiome

Microbes are ubiquitously present on almost every part of the plant both belowground and aboveground. The microbes assemble into diverse communities because of dynamic and complex associations with plant and environment¹⁷. These microbial communities can be categorized into epiphytes that are found on plant surface, endophytes that are present inside plant tissues, phyllospheric microbes that grow on leaf surfaces, and rhizospheric microbes that are found in soil surrounding the roots⁷. Each of these serve as unique niches for interacting with plant host. Among these, rhizosphere microbiome is considered to significantly impact the plant health^{18, 19} as it enhances the nutrient uptake and offer protection against pathogens²⁰. Therefore, it is essential to understand the various processes and associations between plant roots and microbes in the rhizosphere (**Figure 1.1**).

Plants release 5–21% of their photosynthetically fixed carbon into surrounding soil through its roots in the form of root exudates, sloughed cells or mucilage²¹. The root exudates consist of primary and secondary metabolites, including sugars, amino acids, and phenolics²². The root

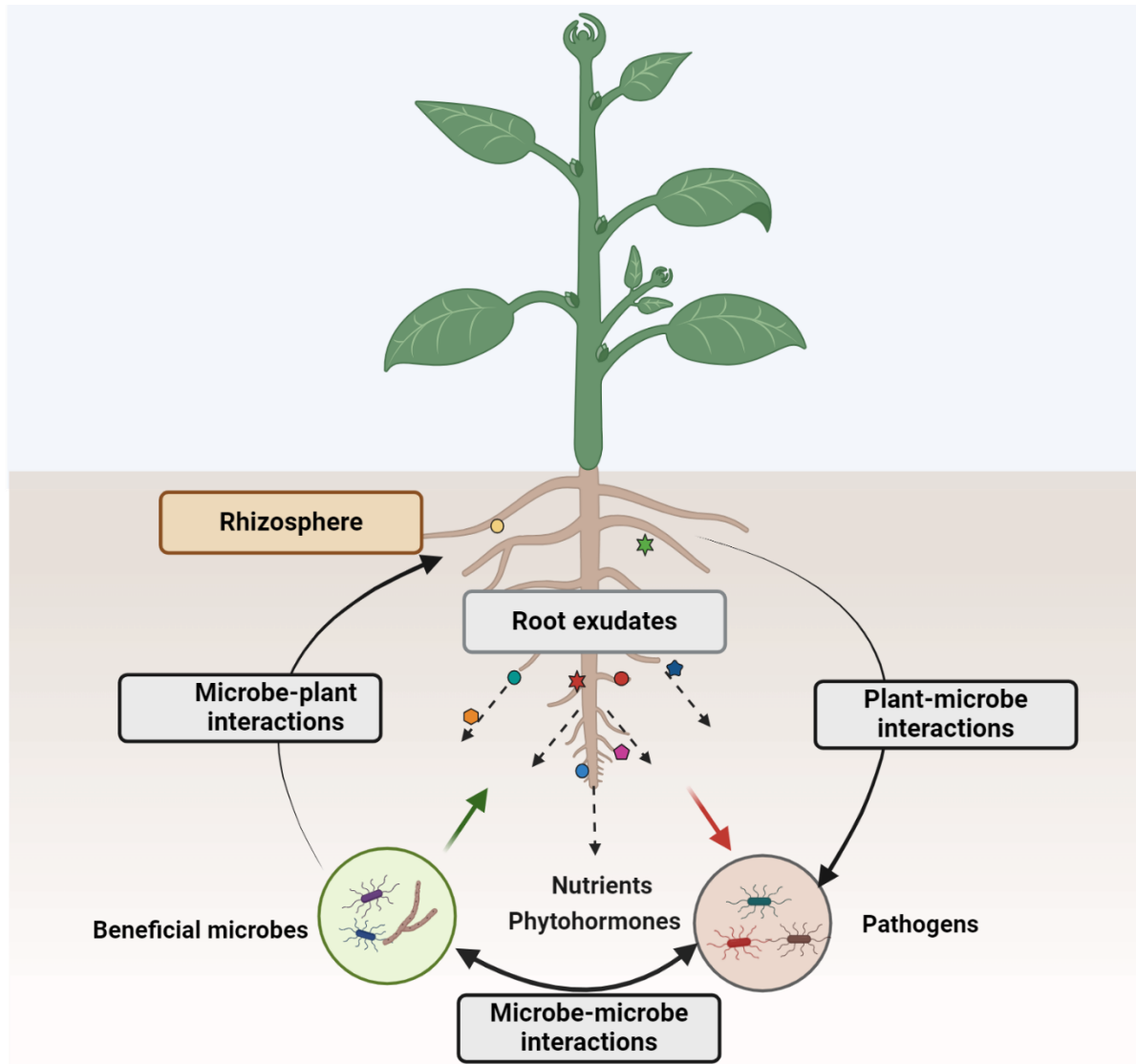


Figure 1.1: Interactions in the rhizosphere. Plant-microbe interactions: Plant roots secrete root exudates and phytochemicals that interact with microbial populations to form niches. Some metabolites released by plant roots play a role in attracting beneficial microbes (indicated by green arrow), while others repel or eliminate pathogens (indicated by red arrow). Microbe-plant interactions: Beneficial microbes promote plant growth via various mechanisms including nutrient chelation and transporting them to plants, phytohormone production, secrete antibiotics to protect plants from pathogens. Microbe-microbe interactions: Cooperative interactions by nutrient cross-feeding or competitive interactions for resources such as nutrients or space. Figure generated using BioRender.com.

exudates are known to play a crucial role in rhizosphere microbiome assembly and composition²³. Plant selectively recruits microbes from soil to their rhizosphere by releasing root exudates²⁴. The root exudates consist of metabolites that play a role in attracting beneficial microbes by providing nutrients to them and/or acting as signaling molecules for establishing interactions while others may play a role in deterring or eliminating pathogens²⁵. The rhizosphere microbes confer several advantages to plant host such as improved nutrient availability by nutrient chelation and transporting them to plants, growth promotion by producing phytohormones, secrete antibiotics to protect plants from pathogens²⁶. The rhizosphere bacteria help in mineralizing and breaking down organic forms of major nutrients such as phosphorus, nitrogen and potassium in soil and converting them into inorganic forms which the plant can uptake and utilize^{27, 28}. The bacteria in rhizosphere can stimulate plant root growth by producing indole acetic acid, a plant hormone that is involved in plant growth^{29, 30}. The beneficial microbes in the rhizosphere employ various mechanisms to protect the plant from adverse effects of pathogens by competition for space at the root surface and for essential nutrients. Some microbes produce secondary metabolites to inhibit the growth or kill pathogens and thereby protecting the plant³¹. In addition to plant-microbe interactions, complex intraspecies and interspecies microbial interactions also occur in the rhizosphere. These interactions could be cooperative by nutrient cross-feeding or competitive interactions for essential resources such as nutrients or space³². The plant associated microbial communities do not represent random assembly of microbial members but are governed by complex interactions among the microbes, the plant host and the environmental conditions^{17, 33}. However, the underlying mechanisms of rhizosphere microbiome assembly are not well understood. Studying the complex processes involved in the rhizosphere microbiome assembly are crucial as they will provide insights for improving plant health and productivity.

Bridging the holistic and reductionist approaches to study plant associated microbiome

To study plant-microbe interactions, researchers adopt either holistic approaches that embrace the natural complexity of these systems or reductionist approaches that allow more control of experimental factors in study and focus on specific interactions³⁴. Both these experimental approaches are on opposite ends of the complexity-control spectrum (**Figure 1.2**).

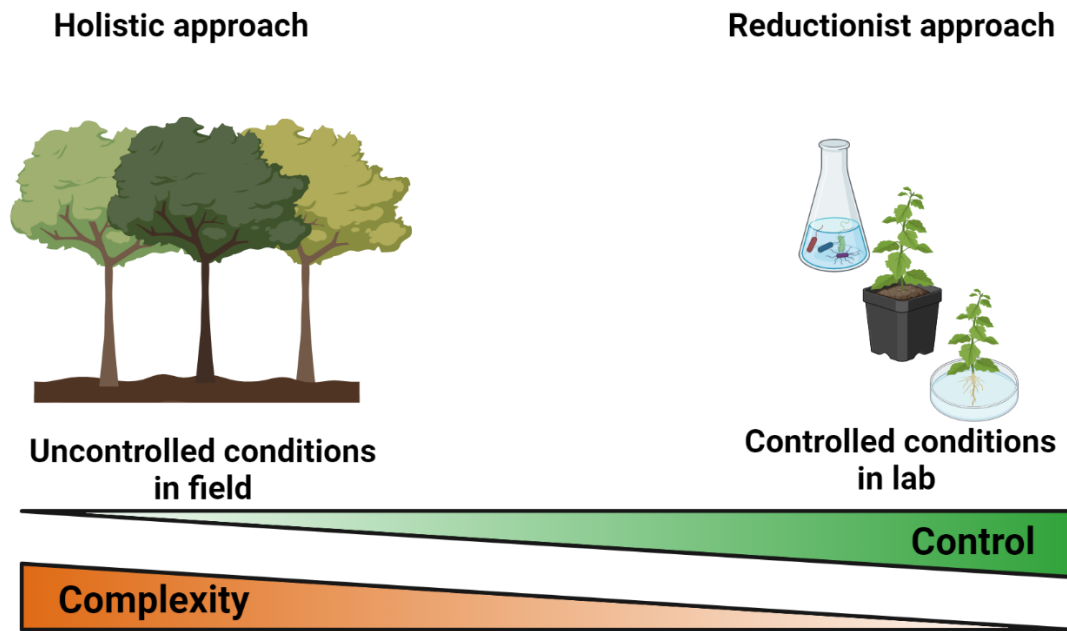


Figure 1.2: Experimental approaches to study plant-microbe interactions. Holistic approaches are inherently complex and offers less control on environmental variables. In contrast, reductionist approaches simplify the system and offer control to manipulate experimental conditions to determine the effects of specific components of plant-microbe systems. Figure generated using BioRender.com.

Use of both the reduced and high complex systems is crucial in obtaining a complete understanding of these complex plant-microbe interactions³⁵. Using the reductionist approaches, low levels of genotypic and functional diversity are studied under controlled environmental conditions, which enables reproducible investigations of plant-microbe interactions³⁶. Engineered habitats such as EcoFabs are used in reductionist approaches to control the complexity while enabling omics measurements³⁷. The simplicity and control in reductionist approaches enables interpretability and understanding of the plant-microbe interactions, however, this strategy is not suitable for studying the importance of higher-order ecological processes on plant holobiont. In contrast, the holistic approaches are used to study plant-microbe interactions across complex natural environments and capture the ecological interactions between genotype, environment and functional variables and their relative influence on the plant holobiont^{38, 39}. But the increase in complexity in holistic approaches presents a challenge for all omics techniques due to sparse data generated and methodological limitations. Additionally, it would be extremely challenging to dissect specific interactions and study them at the mechanistic molecular level. It is now very well known that plant-microbe interactions and microbial communities are highly dependent on entire ecosystem⁴⁰. Experimental designs that can act as a bridge in between the holistic and reductionist approaches are essential in addressing some of the limitations in using only either one of these approaches in advancing our understanding of processes that govern the plant microbial communities³⁴.

A potential approach to address the limitations of studies on highly complex natural communities and also those studies examining interactions between a monoculture and a plant host is to utilize constructed microbial communities. The constructed microbial communities or synthetic microbial communities (SynComs) are microbial consortia that are designed to represent the natural microbiome composition and functions at some scale, while reducing the complexity of natural microbiome and still preserving some of the important interactions between microbes and their hosts⁴¹. This approach has garnered significant attention in recent years in microbiome research³⁹. SynCom construction requires careful design, selection criteria, and available culture collections of microorganisms that can effectively represent and resemble the natural microbiome. The experimental approaches using SynComs with less complexity can be used to understand specific interactions and gradually build the complexity of the system by

adding more microbes, plant host, environmental variables etc. for a comprehensive understanding of microbial dynamics and their functionality in rhizosphere³⁹.

Omics technologies to study plant-associated microbiomes

Advancements in high throughput sequencing platforms and various omics technologies including genomics, transcriptomics, proteomics, and metabolomics have greatly improved our understanding of the complex plant-microbe systems. Each omics technology can answer different questions related to plant holobiont (**Figure 1.3**). Metagenomics helps in identifying which microbial species are present in the plant environment and reveals the functional potential of microbes by measuring the DNA from microbiomes. Metatranscriptomics measures microbial RNA transcripts and is valuable in understanding which microbial genes are actively expressed. Metaproteomics provides functional information by measuring proteins produced in the microbial community. Metabolomics further provides insights into chemical signals between plants and microbes by measuring metabolites produced by plant and microbes. Each omics technology has its own advantages and limitations to study these systems⁴². Integrated multi-omics studies can provide a more comprehensive understanding of plant holobiont but are challenging because the processes of gene expression, translation and metabolite production operate on distinct timescales from each other. Additionally, metabolism is greatly influenced by the environmental factors, which makes it difficult to isolate the effects of environment factors from genetic outcomes⁴³.

Over the past several years, our understanding of the structure and diversity of the plant endosphere, phyllosphere and rhizosphere microbiomes has expanded due to 16S ribosomal RNA (rRNA) sequencing^{44, 45}. Similarly, DNA barcoding of Internal Transcribed Spacer (ITS) region of the nuclear DNA has been instrumental for our understanding of fungal diversity⁴⁶. The core microbiome members that engage with plants consistently are identified using the high-throughput sequencing methods using marker genes (e.g., 16S rRNA, ITS)^{47, 48}, but these methods lack the functional information. Therefore, it is important to follow up or integrate the taxonomic approaches with approaches that can provide functional information such as metagenomics, metatranscriptomics, metaproteomics, and metabolomics as there is increasing

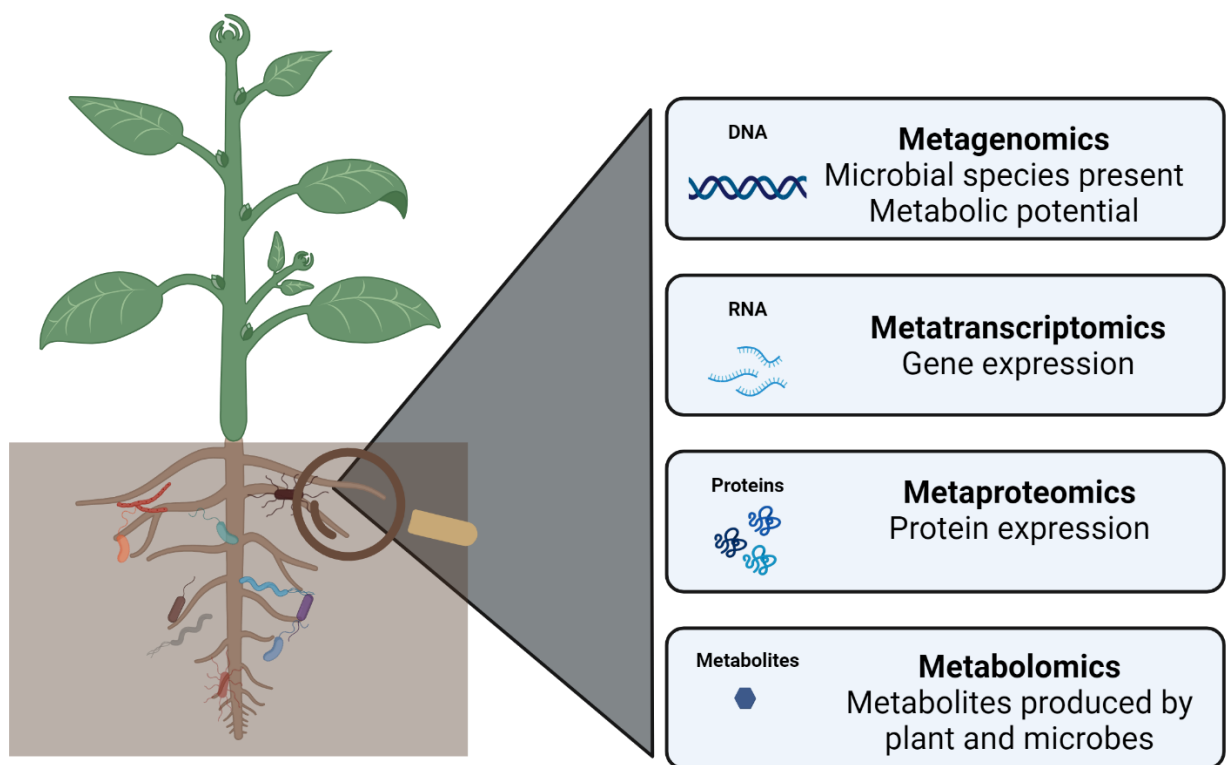


Figure 1.3: Omics approaches for studying plant associated microbiomes. Metagenomics to study which microbial species are present and metabolic potential, metatranscriptomics to determine gene expression, metaproteomics to understand function and metabolomics to study metabolites produced in the plant- microbe system. Figure generated using BioRender.com.

evidence of functional redundancy in microbiomes that have different species⁴⁹. It is predicted that integrating metagenomics with other downstream meta-omic approaches is essential for understanding the interactions between plants and microbes that shape the core microbiome taxa and their functions^{50, 51}.

Characterizing plant microbiome functions using metaproteomics

Metaproteomic analysis characterizes the entire suite of proteins expressed in a microbial community at a given timepoint, providing information on composition and functions in microbiomes⁵². The mass spectrometry-based metaproteomics is a powerful tool in characterizing microbial proteins and has become a critical component of microbiome research in various environments, as it enables the study of multiple interactions that occur simultaneously in an ecosystem⁵³. In recent years, several studies have applied metaproteomics to advance our understanding of microbial functions in plant such as maize⁵⁴, rice^{55, 56}. To perform metaproteomic experiments, several important things involved in the workflow need to be taken into consideration, including biomass availability, choice of sample preparation protocols and LC-MS/MS approaches to study each individual system⁵⁷. The biological diversity (i.e., the number of species present and their relative abundances) of the microbial communities being analyzed directly affects the proteome coverage of the population. In a typical experimental run, thousands of proteins can be identified and often the proteome coverage is biased towards proteins that are highly abundant in the dominant species of the populations. This makes measuring the low abundant proteins more challenging. The fewer low abundant proteins or proteins from less dominant species that are playing prominent functional roles in biological processes are overshadowed by highly expressed proteins. The dynamic range, which is the difference between the least abundant and most abundant signals in a given measurement and proteome coverage, number of proteins identified for large-scale metaproteomics can be increased by using multidimensional protein identification technology (MudPIT), a multidimensional chromatography technique to separate complex protein samples prior to analyzing samples on mass spectrometry⁵⁸. Another important consideration for metaproteomic experiments is the sample biomass. Typically, several micrograms of protein will be required for adequate analysis by mass spectrometry. The protein extraction from the limited biomass of

plant-associated microbiota for metaproteomics is one of the challenging steps due to the high ratio of plant to microbial proteins in the sample, which makes microbial protein detection difficult and therefore requires method optimizations⁵⁹ and new method developments. Further, availability of reference metagenomic and/or metatranscriptomic databases are required for metaproteomic analysis; these serve as reference protein databases to match peptide spectra obtained from mass spectrometry analysis⁵⁷.

Metabolomics to study the metabolites produced by plant and microbes in plant holobiont

Metabolites are the direct effectors driving the fundamental processes of recognition and communication between the plant and microbes. They are exchanged across symbiotic, commensal, parasitic and pathogenic interactions between the organisms. From the genome predictions, it is known that the plants, and their associated microbes have the ability to produce thousands of molecules that interact with and influence their ecosystems⁶⁰. Mass spectrometry-based metabolomics is one of the key technologies that is used currently to characterize these diverse chemical inventories produced by plant and microbes⁶¹. During the past few years, the insights gained from the metabolomics studies have been crucial in understanding the influence of plant metabolites in shaping their microbial communities and the effects of microbially derived metabolites on plant hosts and their ecosystems^{25, 61, 62}. In plant-microbe studies, metabolomics is often applied to characterize the plant root exudate profiles and their impact on the microbes present in the surrounding environment^{63, 64}. Additionally, the effects of associated microbes on host metabolism are also studied using metabolomic approach. While in situ analysis of metabolites provides the most relevant data for understanding the plant-microbe interactions, this can be very challenging due to several factors. Environmental factors such as soil properties are highly variable, and it would be difficult to replicate the experiments. Additionally, the metabolomic results from in situ analysis can be confounded by consumption and/or breakdown of metabolites by other organisms in soil and presence of other biomaterials in the surrounding environment⁶⁵. Consequently, most metabolomic studies employ a reductionistic approach. Studies focused on understanding the effect of rhizosphere on microbiome composition and function were mostly conducted in sterile or semi-sterile environments including soil or artificial systems such as hydroponic systems⁶⁶⁻⁶⁸. Our understanding of root

exudate compositions of *Arabidopsis*⁶⁹⁻⁷¹, rice^{72, 73} and maize⁷⁴ and their impact associated microbiomes^{75, 76} has increased due to the significant technological developments in mass spectrometry. There are only a few number of metabolomics studies conducted in field^{77, 78} or greenhouse soil environments^{79, 80}. Recent experimental advances are promising for addressing challenges related to metabolomic analysis in non-sterile soil matrices⁸¹.

Overview of dissertation

It is well-established that microbiomes have a significant influence on plant health. However, the limited understanding of specific mechanisms driving the dynamics of plant-associated microbial communities represents a critical knowledge gap. This dissertation seeks to advance our understanding of plant-microbe interactions of *Populus* spp. and its rhizosphere microbiome by developing and applying liquid chromatography coupled mass spectrometry-based metaproteomics and metabolomics methods to gain insights into factors driving microbiome assembly and the functions of microbiome that were previously unknown. Synthetic microbial communities are invaluable in plant-microbiome research as they bridge the gap between the single microbe-plant and whole community-plant experimental approaches. This dissertation employed a constructed microbial community that taxonomically and functionally resembles a natural *Populus* bacterial community for fundamental understanding of microbiome assembly under various experimental conditions and additional layers of complexity such as plant host were introduced, to address more intricate questions regarding the spatial and temporal dynamics in microbiome composition along the plant root axis, their functional responses to plant root chemical environment. Root exudates, a key player influencing microbiome assembly were also studied in this dissertation.

This dissertation is divided into seven chapters including this introduction chapter. Chapter 2 describes the experimental methods used in chapters 3-6 including proteomic and metabolomic sample preparations, chromatography, mass spectrometry acquisition and downstream processing methods.

Chapter 3 focuses on gaining fundamental understanding of microbiome assembly using metaproteomic measurements of constructed microbial communities grown under different

conditions. First, we explored the inter-microbial interactions and adaptation strategies employed by bacterial microbes in a ten-member constructed community grown in complex and minimal nutrient media conditions. The results from this study provided insights into the functions and metabolic activities of microbes that enabled them to persist in the stable microbial community observed. To further investigate whether the initial inoculum ratios of the bacterial strains have influence on which strain becomes abundant in the community, the three bacterial strains that persisted in the stable community were selected and a simplified three-member community was constructed using different initial inoculum ratios. By performing proteomic analysis on community samples collected at various timepoints, we uncovered the temporal functional dynamics of the individual bacterial strains in these simplified communities.

While the first research chapter focuses only on the microbiome composition due to microbe-microbe interactions, Chapter 4 focus on microbiome assembly in presence of plant host. In this chapter, a novel peptide extraction method for metaproteomic characterization using plant-agar systems is described. Additionally, spatial sampling of root-associated microbiomes along the plant root axis to study microbial composition and function is demonstrated. Distinct microbiome compositions were observed surrounding distinct root regions, these observations led us to further investigate the spatial dynamics in root exudates, that are known to shape microbial community composition.

Chapter 5 focused on understanding the microbial responses and community composition in response to plant root chemical environment. In this chapter, the application of a novel filter-based experimental set up on plant roots was demonstrated to specifically isolate and study the microbial responses to chemicals released by the plant roots. In this study, proteomic analysis of both monocultures and mixed community grown in presence of plant was conducted to investigate plant-specific responses. The study provided insights into functional behaviors of bacterial strains in response to plant root chemical environment that facilitated their growth and survival in rhizosphere.

Chapter 6 describes the comprehensive spatial-temporal analysis of *Populus* sp. root and exudate metabolite profiles by applying LC-MS/MS based untargeted metabolomics approach. In this study, a plant-agar system that facilitated spatial sampling of root exudates surrounding

distinct root regions was used and a biphasic extraction method was employed to capture the diversity of root exudates. This chapter provides an in-depth analysis of differences in root and exudate metabolite profiles due to the influence of genotype and nutrient conditions.

Chapter 7 summarizes the contributions of research described in this dissertation in advancing our understanding of plant-microbe interfaces of *Populus* sp. and its rhizosphere microbes and future perspectives.

CHAPTER 2: METHODOLOGIES FOR MASS SPECTROMETRY-BASED METAPROTEOMICS AND UNTARGETED METABOLOMICS ANALYSIS

Overview of Liquid chromatography coupled mass spectrometry (LC-MS/MS)

Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) is a powerful analytical technology to analyze complex biological mixtures including proteins and metabolites^{82, 83}. It effectively separates, identifies and quantifies the vast number (thousands) of analytes present in biological samples without extensive purification. The wide dynamic range i.e., the ability to detect high and low abundant compounds in a single measurement using LC-MS/MS allows for a comprehensive profiling of molecules present across a broad abundance range in a complex sample mixture. A major advantage of LC-MS/MS based analysis of biological samples is its agnostic untargeted approach of not requiring prior knowledge of composition of sample mixture and can identify the compounds by matching the fragment ions to spectral libraries. It offers both qualitative and quantitative results in a single measurement⁸⁴. In this dissertation, LC-MS/MS was applied in proteomics/metaproteomics and metabolomics experiments.

Liquid chromatography

Samples for proteomics and metabolomics analysis for plant-microbiome research are inherently complex containing thousands of biomolecules. Given the complexity of the samples utilized in this dissertation research, direct analysis without separation would lead to incorrect identification and quantification of individual analytes in the sample mixture due to the overlapping signals and hence there is need for analytical instrumentation that can effectively separate the complex sample mixture. The separation of analytes helps in detecting the low abundance molecules in the presence of high abundant ones. Liquid chromatography (LC) is an analytical technique used to separate the analytes in the sample mixture by passing a liquid mobile phase through a solid stationary phase. The analytes are separated based on their physicochemical properties such as hydrophobicity (reverse-phase), size (size exclusion chromatography), charge (ion exchange chromatography)⁸⁵. There can be wide distribution of

physicochemical properties of the analytes in biological samples. A combination of one or more chromatographic separations may be required for a better resolution of the complex sample mixtures.

One-dimensional liquid chromatography

One-dimensional liquid chromatography (1D-LC) is a standard chromatographic technique where the analyte mixture is separated using a single chromatographic column and this is the most employed technique. 1D-LC is easy to set up as it uses a specific stationary phase and mobile phase to separate the analytes based on their interactions with stationary phase. In 1D-LC, different stationary phase materials can be used such as reversed-phase (RP), strong-cation exchange (SCX), and hydrophilic interaction liquid chromatography (HILIC). The selectivity of chromatographic separation depends on the interactions of the analyte with the stationary phase and therefore an appropriate stationary phase should be selected based on the analytes. Other factors that can influence the separation and needs consideration are column length, particle sizes of stationary phase and mobile phase solvents⁸⁶.

Reversed-phase (RP) is the most used stationary phase in LC-MS analysis of biomolecules⁸⁶⁻⁸⁸. It uses a non-polar stationary phase, and the separation is based on the hydrophobic interactions of the analyte molecules with the stationary phase. The sample containing the analyte mixture is loaded onto the RP column using mobile phase solvent that contains low percentage of organic solvent. RP has retention power for polar analytes also which makes it a popular choice⁸⁸. After the analytes interact and bind to the non-polar stationary phase, they are separated using gradient elution, where the proportion of organic solvent is gradually increased in the mobile phase over the chromatographic run time. This allows the analytes with different polarities to elute from the column sequentially, with the more hydrophobic analytes eluting last as the organic solvent increases in the mobile phase⁸⁹. A octadecyl or C18-bonded silica column is most commonly used stationary phase material in RP-LC⁸⁸.

The chromatographic performance is significantly improved by using UPLC (ultra performance liquid chromatography) with columns with smaller particle size such as 1.7 μm compared to HPLC (high performance liquid chromatography) as UPLC can operate at much higher pressures such as 10,000 psi. UPLC offers significant advantages in resolution, speed, and

sensitivity of measurements⁹⁰. Further, nanoflow LC (nano-LC), which uses capillary column with smaller inner diameter results in smaller droplets during the electrospray ionization and enables relatively more ions to convert to gas phase compared to conventional LC. Using nano-LC, improved ionization efficiency can be obtained for both peptides and small molecules. As nano-LC operates at lower flow rate compared to microflow LC, there is less dilution of sample and therefore requires minimal sample amount^{91, 92}. In this dissertation, nano flow UPLC was employed with nanospray emitter with 75 μm ID and 250 nL/min flow rate.

Multidimensional chromatography

Although 1D-LC is effective in separating most analytes, it can be challenging to effectively separate very complex samples such as metaproteomes and may result in coelution of peptides due to limited peak capacity. Therefore, for more complex sample mixtures, a multidimensional chromatography approach is applied for increased resolution of analytes⁹³. Typically, a two-dimensional (2D-LC) approach which uses two sequential columns with different separation mechanisms is utilized for complex samples^{94, 95}. This allows significantly higher peak capacity and better resolution by separating analytes based on first separation property in first dimension, followed by second separation based on another separation property in the second dimension of 2D-LC⁹⁴. Multidimensional chromatographic separation can be achieved by several approaches to achieve deeper and orthogonal separation, by using RP in combination with other stationary phase materials such as strong cation exchange (SCX), strong anion exchange (SAX), or RP at different pH⁹³. Typically, in a 2D-LC approach employed for proteomic analysis, orthogonal separation is achieved by both hydrophobicity and charge by using SCX as first dimension and RP stationary phase as second dimension (**Figure 2.1**)^{58, 96}. For this a triphasic column called a back column is utilized, with first RP segment, followed by SCX and then RP segments. First, the sample is loaded onto RP segment, washed and desalted by flowing solvent with low organic content, then solvent with high organic content is flowed to elute the analytes from RP onto the SCX where the analytes are separated based on their charge. For analyte separation using SCX, step elution using increasing amounts of ammonium acetate is used to elute the analytes onto the RP segment where the analytes are separated based on their hydrophobicity using a gradient elution with increasing concentrations of solvent with higher organic content. The eluted analytes

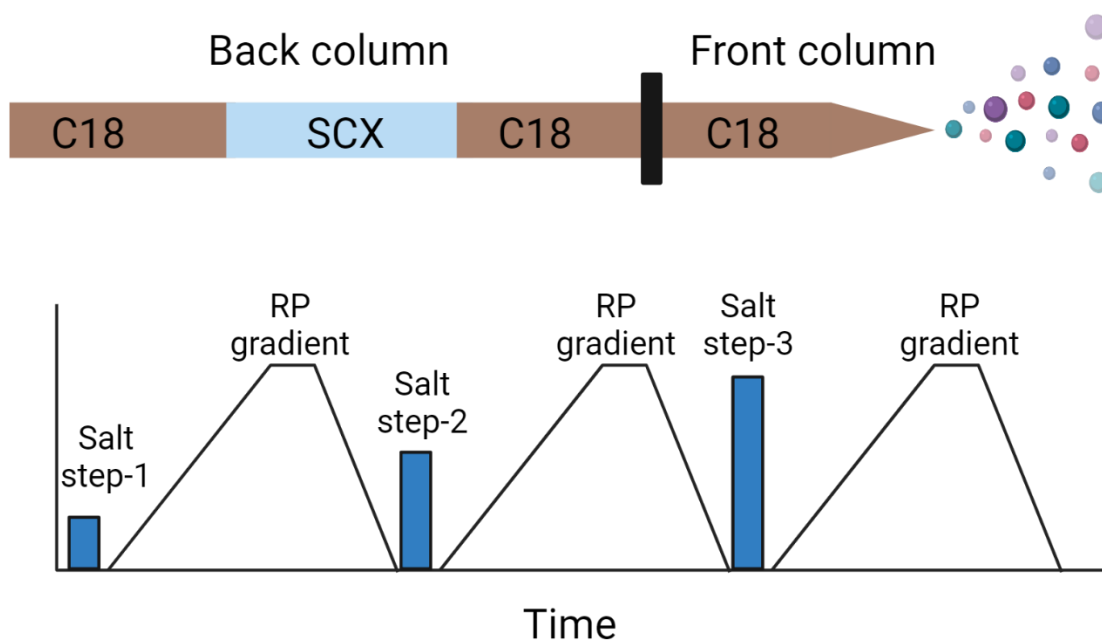


Figure 2.1: Multidimensional liquid chromatography column. A triphasic column using reverse phase (RP) C-18 for clean-up, followed by strong cation exchange (SCX) resin and C-18 for orthogonal separation based on charge and hydrophobicity. A series of salt cuts are used to elute analytes from SCX and then eluted from RP by using a gradient with increasing organic solvent content. The eluted peptides are then separated on front RP column and ionized using ESI.

are passed onto the front column with RP for separation and then directed into electrospray ionization source. In this dissertation, 2D-LC approach was employed for proteomic and metaproteomic analysis.

Mass spectrometry

Mass spectrometry (MS) is a powerful analytical technique used to measure the mass-to-charge (m/z) ratio of ions. MS can provide qualitative information (m/z on x-axis) and quantitative information (intensity on y-axis) of the measured ions. Mass spectrometers consist of three main components (1) an ionization source to convert the analytes into gas phase ions (2) a mass analyzer to measure the m/z of ions and (3) a detector to qualitatively and quantitatively detect the ions. Although mass spectrometry was originally developed to measure isotope abundances and elemental masses for applications in physics, it is now widely used across biological research⁹⁷. The ability to analyze biomolecules using mass spectrometry is due to the advances development of soft ionization techniques such as electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) that can convert the biomolecules into ions⁹⁸.

Ionization sources

The first step in mass spectrometry analysis is the ionization of analytes i.e., conversion of analytes from liquid or solid phase into gas-phased ions. Several ionization methods including electron ionization (EI), chemical ionization (CI), atmospheric pressure chemical ionization (APCI), electrospray ionization (ESI), matrix-assisted laser desorption/ionization (MALDI), desorption electrospray ionization (DESI) are currently available for ionization, each with their own advantages and applications⁹⁹. The choice of the ionization method depends on the type of analytes that are being measured. Two ionization methods MALDI and ESI are extensively used in biological research, because these are soft ionization techniques that imparts minimal energy to the analyte molecules during the ionization process. This helps to maintain the structural integrity of the fragile biomolecules during conversion to gas phase¹⁰⁰⁻¹⁰². MALDI utilizes a solid matrix, and ionization is facilitated via laser energy⁹⁸. MALDI produces singly charged ions of analytes which minimizes the spectral complexity and is widely applied for MS

imaging studies¹⁰³. ESI depends on the formation of charged droplets that release ions from the surface called the ion evaporation model or via complete solvent evaporation called charge residual model⁹⁹. In this dissertation research ESI has been used and is described in detail in the following section.

Electrospray ionization (ESI)

ESI is the ionization method which converts the analytes in liquid phase to gas-phase ions at the source of a mass spectrometer for subsequent mass analysis (**Figure 2.2**). ESI is the most used ionization source for LC-MS analysis because of its compatibility and sensitivity. During ESI there is minimal to no fragmentation of the analytes due to its soft ionization nature. It allows analysis of both positively charged and negatively charged analyte ions as it generates protonated ($[M + H]^+$) or deprotonated ions ($[M - H]^-$) depending on the solvents and ionization mode. In ESI-MS based peptide analysis for bottom-up proteomics, positive ionization mode is preferred as most peptides are expected to ionize in positive ion mode compared to negative ion mode¹⁰⁴. The peptides, particularly those generated using trypsin digestion have multiple proton-accepting sites and this makes them to exist as singly, doubly or multiply charged ions¹⁰⁵. However, acidic peptides that contain high amounts of acidic amino acid residues such as aspartic acid and glutamic acid are difficult to ionize using positive mode and may require negative ion modes, but this will result in complex MS/MS spectra that is difficult to interpret by using traditional methods¹⁰⁶.

To convert the analytes from liquid phase to gas phase ions, ESI uses electrical energy and involves three steps (1) fine spray of charge droplets are dispersed, (2) solvent evaporation and (3) ion ejection from the charged droplets¹⁰⁷. The analyte sample eluting from LC column is directed through an emitter tip (1-5 μm for nanoESI), during which a high voltage ($\sim 2.5\text{-}6\text{kV}$) is applied. The applied voltage can be either negative or positive based on the nature of the analyte being measured and it produces charged droplets that are sprayed out from the emitter tip into the atmosphere forming a Taylor cone, this process is called nebulization⁹⁹. The presence of a sheath gas, usually nitrogen flowing around the emitter during the nebulization process assists in formation of spray droplets from the tip. These charged droplets pass through a heated capillary on the mass spectrometer for complete solvent evaporation from the spray droplets to form ions.

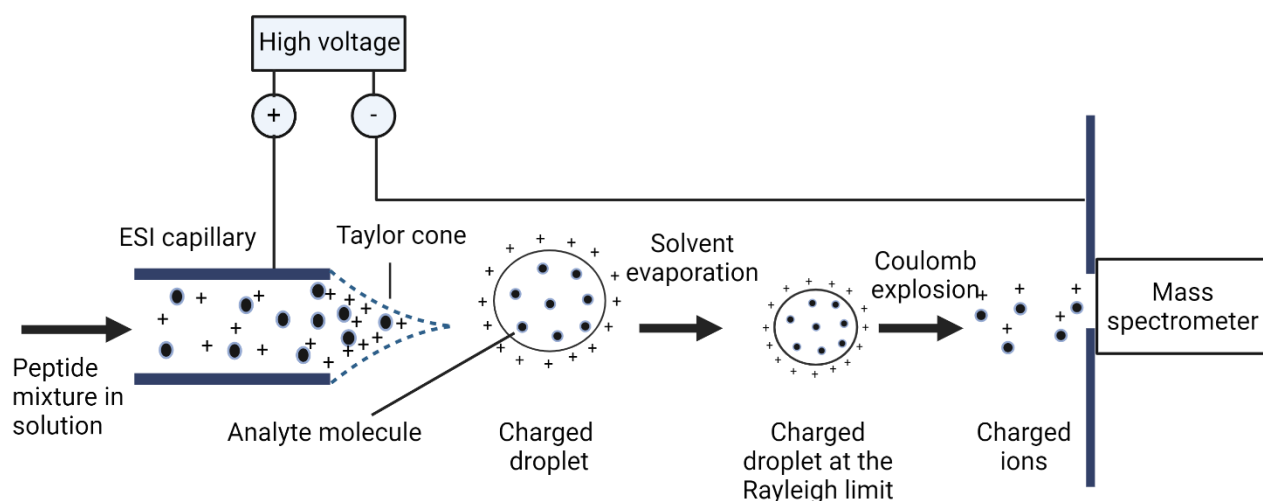


Figure 2.2: Schematic illustration of electrospray ionization (ESI) process. The charged analytes in sample solution are separated by liquid chromatography and large solvent droplets containing charged analytes are sprayed towards the heated capillary of the mass spectrometer from the chromatographic column emitter. The solvent evaporates as the large solvent droplets move towards the heated capillary and forces the ions together until charge repulsions causes the ions to leave the droplet. The charged ions are then pulled into the mass spectrometer due to the high voltage between the emitter and mass spectrometer. Figure generated using BioRender.com.

The ion formation involves solvent evaporation and Coulomb fissions or Coulomb explosion of the charged droplets, which results in formation of smaller droplets. As the solvent evaporates, the repulsions between charges at the surface of the droplets overcomes the cohesive force of the surface tension and the charged droplets become unstable. The gas-phase ions are directly released from the large, charged droplets, this occurs at Rayleigh limit. This results in the formation of small, charged droplets. In the charge residual model, the analyte ions will be released by complete evaporation of the solvent. The resulting ions formed by ESI at atmospheric pressure then pass through an orifice into the mass spectrometer for mass analysis^{99, 107}.

ESI is widely used as it is suitable for a wide range of compounds with broad range of polarity and varying molecular weights. However, one of the major drawbacks of ESI is the ion suppression that can occur because of the presence of a high concentration of nonvolatile analytes as it can change the efficiency of droplet formation which then inhibits release of ions into the gas phase. In samples with high concentration of analyte molecules, there could be competition for limited charges or space available on the surface of the droplet and this can also decrease the efficiency with which the ions are formed. There are several strategies to overcome ion suppression during ESI including sample purification such as solid-phase extraction (SPE), liquid-liquid extraction (LLE), or protein precipitation, modification of the chromatographic conditions by changing the mobile phase additives to shift the retention times of the analytes away from the eluting region that is affected by ion suppression, switching to the negative ion mode which results in fewer analytes to be ionized and therefore the matrix effect is lower, reducing the amount of sample being introduced to the ESI chamber by diluting the sample, reducing the injection volume, or reducing the flow rate. Nano-ESI used substantially lower flow rates, which allows fewer analyte and nonvolatile molecules to be injected into the ESI source and therefore enhances the desolvation process of the sample droplets⁹⁹. For these reasons, nanoflow-ESI has been employed in this dissertation research.

Mass analyzer and detector

Following the ionization of analytes, the next critical step in mass spectrometry analysis is qualitative (m/z) and quantitative (intensity) detection of ions, and this function is performed by

a mass analyzer and detector. Mass analyzers selectively separate the ions based on their m/z for detection by manipulating the ion motion in electric and/or magnetic fields. There are several types of mass analyzers available for the sorting the ions for detection and measurement of analyte ions including quadrupole, ion trap, time-of-flight (TOF) and Fourier transform ion cyclotron (FT-ICR) analyzers, however, the results obtained are greatly impacted by the instrument capabilities and complexity of the sample being studied¹⁰⁸. The combination of the mass analyzer with the detector is the ultimate determining factor of the quality and reliability of the results. Therefore, it is crucial to carefully select a mass analyzer and analytical figures of merit, the factors that are used to evaluate the analytical performance of the mass spectrometer are useful in evaluating and selecting a mass analyzer.

Analytical figures of merit

The analytical figures of merit need to be compared to determine the MS instrumentation that is appropriate for a given research study. Of the several figures of merit that can be used to evaluate the MS instrumentation, mass resolution, mass accuracy, dynamic range, mass range, detection limits and scan speed are important in determining the instruments suitability for specific research¹⁰⁹.

The mass resolution is defined as $m/\Delta m$, where m is the mass (m/z) of the ion and Δm is the peak width measured at full-width half-maximum, FWHM. There is another method also used for calculating mass resolution, called the 10% valley method. According to this method, the resolution is calculated by $m/\Delta m$ for two peaks with equal intensities in a mass spectrum at masses m and $m-\Delta m$ that are separated by a valley of 10% height of either peak. The resolving power is the ability to distinguish or resolve two ions with a small difference in m/z values. Based on the resolution, mass analyzers can be classified as unit-mass-resolution or high-resolution instruments¹⁰⁹. Mass accuracy is defined as how accurately an instrument measures the mass (m/z) of an ion. The exact mass (m/z) of an ion is its theoretical calculated mass. The accurate mass of an ion is experimentally determined m/z , that is measured with an appropriate degree of accuracy and precision. The mass error is the absolute (in μ) or relative difference (in ppm) between the exact mass and the experimentally determined mass of ion and is used as a measure of mass accuracy of the instrument¹⁰⁹. Mass accuracy is influenced by mass resolution

of an instrument, low resolution instruments has a mass accuracy of 0.1 mu, whereas the high resolution instruments enable high accuracy measurements by distinguishing between ions with similar nominal masses (0.01Da)¹⁰⁹. Dynamic range of a mass analyzer can be defined as the range between the least and highest abundant ions that can be detected in each measurement. Dynamic range is an important figure of merit when analyzing complex sample mixtures containing a broad of analyte molecules with varying abundances. Mass range is the difference between the smallest and largest mass (m/z) of ions that can be measured. Scan speed is the time required to collect a complete mass spectrum and faster scan speeds are desirable for high-throughput measurements¹¹⁰. The detection limit of the mass analyzer is the lowest concentration of the analytes that can be detected by the instrument and is above the signal to noise ratio.

Several mass analyzers are currently available, and a mass analyzer is selected based on the research question at hand and performance characteristics of the instrument to fulfill that requirement. Most current mass spectrometry systems have high sensitivity and acquisition speeds that can be applied for a wide range of research applications. For highly complex biological samples such as metaproteome samples used in this dissertation, fast scanning instruments with high resolution, high mass accuracy (~5ppm range), wide dynamic range (few orders of magnitude), and high sensitivity instruments are preferred. To combine the performance characteristics from different mass analyzers, hybrid instruments are developed by using different mass analyzers in the same instrument.

Q Exactive plus mass spectrometer

Q Exactive Plus is a hybrid quadrupole-Orbitrap mass spectrometry system and has significantly improved figures of merit (**Figure 2.3**). The orbitrap is an ion trap mass analyzer, that uses an oscillating electric field to capture and separate the ions and analyze their masses based on their oscillation frequencies¹¹¹. Q Exactive plus mass spectrometer (Thermo Scientific) was used for all the research presented in this dissertation. The Q Exactive plus consist of six main components, they are radio-frequency (RF lens), Advanced Active Beam Guide (AABG) for ion transmission, quadrupole mass filter for selection of precursor ions, c-trap for intermediate storage of ions, collision cell for high energy collisional dissociations (HCD) and orbitrap analyzer for mass analysis of ions. The analyte sample is introduced into the ion source

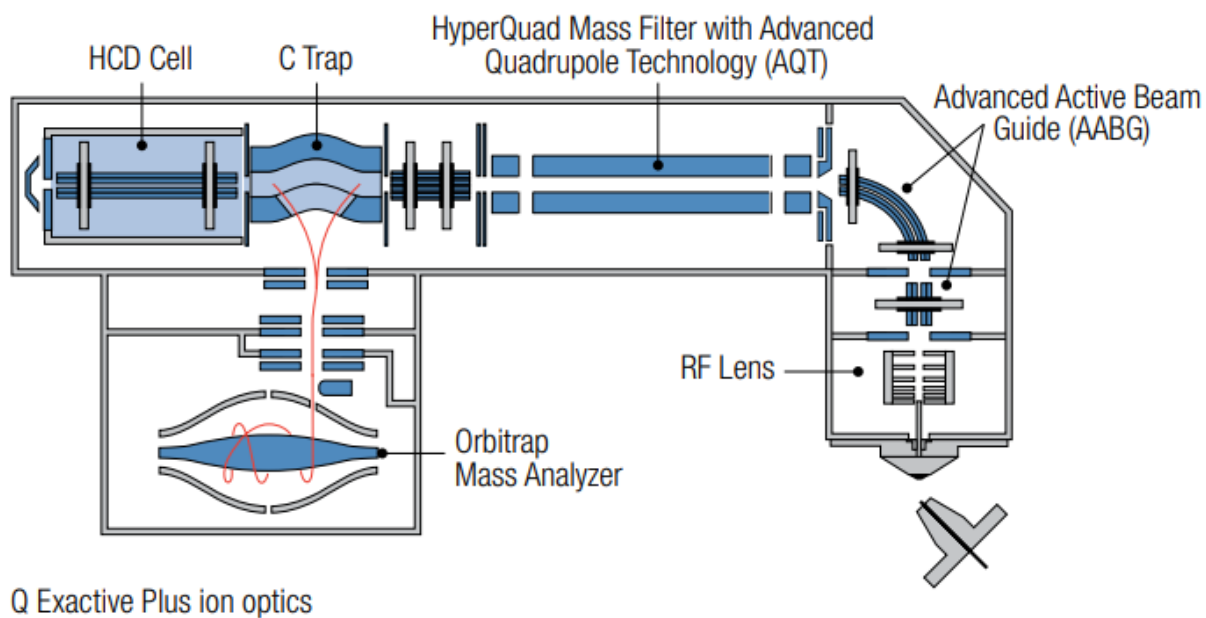


Figure 2.3: Schematic of Q Exactive plus hybrid quadrupole-orbitrap mass spectrometer. The figure was taken from the Thermo Scientific manual.

at the front-end using nano-LC system. The solvated ions pass through the RF lens, it focuses the ions into a tight beam and transmit them through advanced active beam guide that filters out uncharged neutral ions from entering the quadrupole. The quadrupole rod assembly functions as an ion transmission device, it can also filter the transmitted ions according to its mass-to-charge ratios, the filter characteristic of quadrupole is configured by applying RF and DC voltages to the rods. The transmitted ions are then transferred into C-trap, where the ions are accumulated, and their energy is decreased using nitrogen gas. The ions are then injected into orbitrap analyzer, where image current detection is used to acquire mass spectra. For MS/MS, the MS1 precursor ions are transferred into the HCD cell for fragmentation into fragment ions, these are then collected into C-trap as ion packets and sent to orbitrap for detection and mass analysis. This is repeated for all the selected most abundant precursor ions based on the top-N method selected to generate fragment ions and obtain MS/MS spectra¹¹². The Q Exactive plus performance characteristics include those that are desirable for the analysis of analyte mixtures from complex proteome and metabolome samples. These include mass range of orbitrap of 50-6000 m/z, fast scan speed upto 12 Hz, resolving power of 140,000, mass accuracy <1ppm, dynamic range of >5000:1 allows results with high throughput, high confidence, femtogram sensitivity, detection of trace-level and highly abundant ions in same scan.

Tandem mass spectrometry MS/MS and data acquisition

In mass spectrometry analysis, tandem mass spectrometry or MS/MS uses the two stages of mass analysis where parent/precursor ions are fragmented in a collision cell to generate fragment ions and further analyzed. MS/MS analysis provides increased confidence in the analyte identification by detection of both precursor and fragment ions¹¹³. There are several different methods available for ion fragmentation, of which collisional induced dissociation (CID) and higher energy collisional dissociation (HCD) are most routinely used for precursor fragmentation because of their speed and sensitivity¹¹⁴. The precursor ions are isolated and fragmented by inducing collisions with inert gases such as nitrogen, helium at a specific energy. In CID, ions are fragmented in an ion-trap whereas in HCD the ions are fragmented in a collision cell instead of an ion trap and then transferred back to orbitrap through the C-trap for analysis at high

resolution^{114, 115}. HCD employs higher energy dissociations than CID and enables a wider range of fragmentation. In this dissertation, Q Exactive plus mass spectrometer with HCD is used.

Data can be collected in different modes in tandem mass spectrometry namely data dependent acquisition (DDA) and data independent acquisition (DIA) (**Figure 2.4**). In DDA mode, first all precursor ions are detected in the survey MS spectrum, followed by selection of the precursor ions that have the highest intensity based on predefined top-N method for fragmentation in the collision cell, these are then analyzed and detected as the MS/MS spectrum. In DIA, all precursor ions are analyzed in MS mode simultaneously and then fragmented in the collision cell using high-collision energy to generate multiple fragment ions from multiple precursor ions which are then detected in one MS/MS spectrum¹¹³. Although DIA is becoming popular in recent years, DDA is used primarily in biological research. DDA with top-N where N could be 5 or 10 was used in this dissertation research.

LC-MS/MS analysis of biomolecules can be performed using targeted or untargeted approaches. Using the targeted approach, a small number of specific known analytes of interest are measured, while the untargeted approach analyzes all molecules that are measurable in a given sample. While targeted approach has several advantages including specificity and quantification of specific analytes, untargeted approach enables measuring all analytes including chemical unknowns and untargeted approach was employed throughout this dissertation work.

General considerations for LC-MS/MS experiments

A typical LC-MS/MS experiment involves several important steps including experimental design, sample collection, sample preparation, measurements, data processing and analysis. Careful attention must be given to each of these steps and optimized based on the research question to be addressed as they can significantly influence the quality of the results¹¹⁶.

Experimental design is a critical step of any experiment and requires careful planning to ensure it can address the research question at hand. Determining appropriate controls and replicates should be one of the important considerations in designing an experiment. Including both biological and technical replicates help in identifying variability and ensures statistical power for analysis of data generated. Biological samples such as plants and bacteria used in this

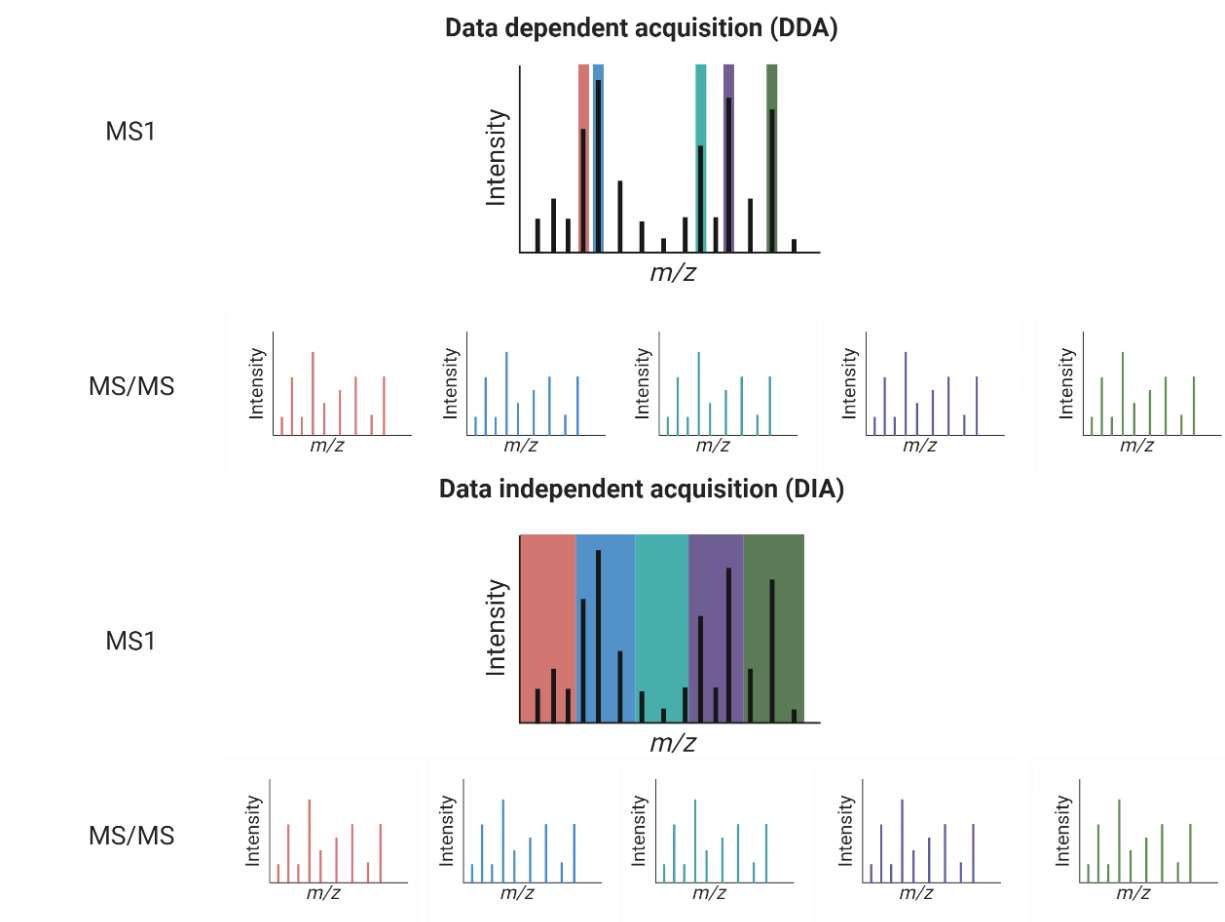


Figure 2.4: Overview of data acquisition strategies in tandem mass spectrometry (MS/MS). In data dependent acquisition (DDA), a predefined number of most abundant precursor ions are selected and fragmented for MS/MS analysis. In data independent acquisition (DIA), the full scan range is divided into several windows and selects all ions in each window for fragmentation and subsequent MS/MS analysis. Figure generated using BioRender.com.

dissertation have inherent variation and there is a need for including biological replicates in experimental design to determine whether the differences observed are an effect of experimental treatment condition or the inherent sample to sample variation. More number of replicates is better but a minimum of three biological replicates is recommended. Additionally technical replicates, which are repeated measurements of the same sample will help in determining the precision of the measurements by identifying any variation in results that can occur due to instrument performance. However, many studies do not include technical replicates due to cost or lack of sample material.

Sample collection and storage is another important consideration for biological samples. Generally, the plant and bacterial samples used in this dissertation research are collected quickly and stored at low temperatures at -80°C for long-term stability of samples. For metabolomic analysis, the samples are frozen immediately after collection using liquid nitrogen to reduce the metabolic activity. After sample collection, sample preparation is one of the critical steps during which proteins/metabolites are extracted and processed for LC-MS measurements. The choice of reagents and solvents used in the extractions greatly influences the extraction efficiency and therefore there is a need to carefully consider these choices and compatibility with MS should also be ensured.

Chromatographic separation is another critical factor as effective separation of analytes is required for analyzing complex biological samples⁸⁴. Selection of the chromatographic parameters including the mobile phase solvents, column, elution gradient, flow rate and run time needs to be optimized for efficient separation of analytes. The mobile phase solvents are selected based on the chromatographic stationary phase being used and the analyte samples that need to be separated.

Mass spectrometry parameters play a vital role in the sensitivity and specificity of detection. Mass accuracy, collision energies, and ion mode (positive or negative) are essential for the identification of analytes reliably. Lastly, data analysis also plays a major role in the quality of the results obtained. Data processing and analysis involves using spectral libraries, database searches, and statistical analysis to analyze and interpret the generated datasets. To ensure robust

downstream analysis, careful considerations for various data processing steps such as false discovery rates (FDR), data normalization, and imputation are necessary.

The choice of extraction protocols, types of liquid chromatography and mass spectrometry instrumentation, data acquisition and data analysis methods depend on the type of samples being analyzed to answer a specific research question. The following sections in this chapter provide details about the methods applied for the research presented in Chapters 3-6 of this dissertation.

LC-MS/MS based proteomics and metaproteomics

Proteomics is a large-scale characterization of proteins of an organism. It provides valuable insights into protein expression and functions which makes it a powerful tool to understand biological responses to experimental conditions¹¹⁷. Metaproteomics extends the proteomic analysis to complex microbial communities at a given point in time¹¹⁸. Metaproteomics enables the study of functional activity and interactions in diverse ecosystems by comprehensively analyzing the proteins produced by all the microbes within a community. It is useful in understanding community's functions in response to environment, metabolic processes, interactions with host etc⁵⁷. In this dissertation, for both proteomics and metaproteomics analysis advanced mass spectrometry techniques (LC-MS/MS) were employed to gain insights into protein expression at both the organismal and community levels.

LC-MS/MS based (meta)proteomic measurements can be conducted using three approaches namely bottom-up, middle-down and top-down approaches. In the bottom-up proteomics approach proteins are digested with an enzyme such as trypsin, and the resulting small tryptic peptides (0.8-3 kDa) are analyzed by mass spectrometry^{119, 120}. Similar to bottom-up approach, the middle-down approach also involves digestion of proteins into peptides, but in a restricted manner that produces longer peptides (2.5 – 8 or 10 kDa) than the peptides generated using bottom-up approach. Middle-down approach uses proteases such as OmpT, Sap9, IdeS that can cleave proteins to produce longer peptides¹²⁰. In top-down proteomics, the protein is ionized directly without digestion, it allows for improved sequence coverage and detection of post-translational modifications (PTMs)¹²¹. While each of these approaches have their own advantages, bottom-up approach is traditionally used in proteomics studies due to inherent

advantages of this approach including easy and better separation using LC, uniform ionization and better fragmentation in MS, proteome coverage¹²². In this dissertation, bottom-up approach was employed for proteomic and metaproteomics analysis described in chapters 3-5.

Bottom-up proteomics workflow

In general, bottom-up proteomics experiment consists of several steps including sample collection and sample preparation, peptide analysis using LC-MS/MS and data analysis (**Figure 2.5**). Each of these steps needs careful consideration depending on the samples being analyzed and their complexity. The various steps involved in bottom-up proteomics in this dissertation are described in the following sections.

Sample collection and sample preparation for proteomics

It is important to consider the sample type being analyzed and the complexity as it can dictate the amount of starting material required and the sample preparations steps for efficient extraction of proteins. For proteomic analysis of microbes typically a few micrograms of sample is required, whereas for metaproteomics of plant associated microbiome the amount of starting material required is much higher due to the limited microbial biomass that can be extracted from plant tissues⁵⁹. Care should be taken to collect the samples at the same time and preserved properly until further processing to eliminate any biases that could arise due to differences in sample collection and storage. The collected samples must be flash frozen immediately at -80°C to prevent proteolysis due to cellular proteases.

After sample collection, the first step in a bottom-up proteomics workflow is sample preparation, which includes cell lysis, protein extraction, protein denaturation and protein digestion. For a comprehensive analysis of proteins in a given sample, sample preparation is the first critical step as the quality and quantity of proteins extracted plays a major role in the quality of the obtained results^{105, 123}. Both physical and chemical methods can be used for cell lysis. The physical methods of cell lysis include mechanical disruption methods such as sonication or bead beating¹²⁴, while the chemical lysis methods use detergents such as sodium dodecyl sulfate (SDS) or Sodium deoxycholate (SDC) that destabilizes the cell membranes of the cells and solubilize proteins¹²⁵. In some protocols, a combination of addition of detergent, heating to high

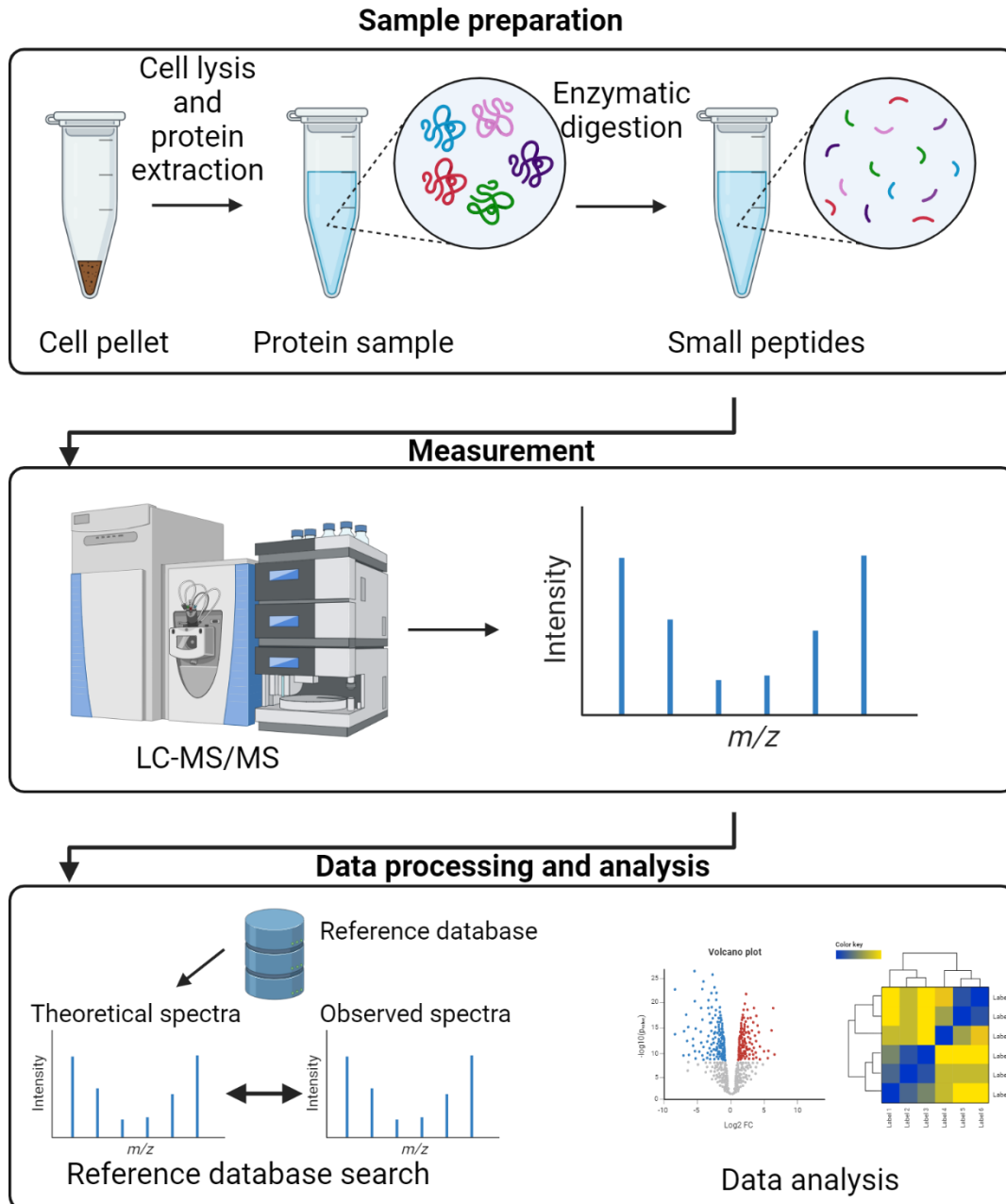


Figure 2.5: Overview of a typical bottom-up proteomics workflow. A typical bottom-up proteomics workflow includes (A) sample preparation that includes cell lysis, protein extraction and enzymatic digestion to peptides (B) Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) measurements for peptide separation and mass spectrometry analysis (C) data processing to identify peptides from acquired MS/MS spectra by matching to reference databases, followed by protein inference and downstream data analysis for biological interpretation of data. Figure generated using BioRender.com.

temperatures such as 90°C and mechanical disruption methods are used for cell lysis, and protein denaturation. Several methods exist for protein extraction including trichloroacetic acid (TCA)/acetone extraction, chloroform/methanol extraction, protein aggregation capture which efficiently separate the proteins from other biomolecules in the cell lysate¹²⁶⁻¹²⁸. Depending on the complexity of the sample matrix and interfering compounds such as in soil, extraction of proteins for proteomics can be challenging and may require optimization of existing methods or development of new methods^{129, 130}. In this dissertation, a novel method for peptide extraction from agar matrix was developed, the details of which are described in chapter 4.

Following cell lysis and protein denaturation, free thiol groups are blocked to prevent disulfide bond formation. Chemical reducing agents such as Dithiothreitol (DTT), 2-mercaptoethanol, tris(2-carboxyethyl)phosphine, or tris(3-hydroxypropyl)phosphine, are generally used for reducing the disulfide bonds and iodoacetamide (IAA), acrylamide, N-ethylmaleimide, or 4-vinylpyridine are used for alkylation¹³¹.

After protein extraction, reduction and alkylation, proteolytic digestion of proteins is performed by using a protease to generate peptides. The choice of protease depends on the research question as most proteases cleave peptides at specific cleavage sites¹³². Trypsin, a serine protease is a preferred protease for most LC-MS/MS based bottom-up proteomic experiments due to its specificity in cleaving proteins at carboxyl side of arginine and lysine and generates peptides with average length of 14 amino acids, which are amenable to both chromatographic separation and mass spectrometric analysis¹³³. The generated tryptic peptides are cleaned and desalted to remove any salts, detergents and buffers that could contaminate the samples. It is crucial to remove these contaminants from samples as they could interfere with the MS measurements. Reversed-phase C18 columns are commonly used for peptide cleanup and desalting¹³⁴.

In this dissertation, a combination of SDS, heating to 90-95°C and sonication or bead beating were used for cell lysis and denaturation, followed by reduction using 10mM Dithiothreitol (DTT) and alkylation using 30mM iodoacetamide (IAA). For protein extraction, the protein aggregation capture method¹²⁸ that uses magnetic beads to capture the precipitated proteins in

solution is used, followed by two sequential trypsin digestions using 1:75 trypsin to protein ratio and incubate at 37°C with constant shaking at 1000rpm for overnight and 3 hours.

Peptide separation using liquid chromatography followed by tandem mass spectrometry analysis

The sample mixture containing tryptic peptides for (meta)proteomic analysis are first subjected to liquid chromatography separation to reduce co-elution of peptides for enhanced identification during MS analysis¹³⁵. RP-LC described previously in this chapter is the preferred method used in proteomics to separate peptides based on their hydrophobicity¹³⁶. For efficient separation of complex peptide mixtures resulted from microbiome samples used in this dissertation research, two-dimensional LC using SCX and RP was employed for orthogonal separation based on charge and hydrophobicity. For the back column, a triphasic column using RP-SCX-RP was constructed as described previously and the back column was connected to front RP column which is 75 µm fused silica capillary column packed with C18, 1.7 µm particle size, 30cm in length. After the peptides are loaded and cleaned on the first RP segment of back column, they are passed on to SCX segment where they are separated based on their charge and eluted using salt pulses. The number of salt pulses and concentrations depend on the complexity of the sample being analyzed. In this dissertation, three salt pulses of 35mM, 50mM, and 1M ammonium acetate were utilized for peptide separation using SCX. For separation using RP, a 180minute gradient was utilized. The eluted peptides are ionized using ESI and mass spectrometry data was acquired using a Q Exactive plus (Thermo Scientific) mass spectrometer in data dependent acquisition mode using the top-10 method, where the top 10 most abundant precursor ions are selected from the MS1 scan and fragmented into fragment ions in MS/MS. The ions generated during peptide fragmentation during MS/MS are called "b" and "y" ions, where "b" ions represent fragments with the charge on the N-terminus of the peptide and "y" ions represent fragments with the charge on the C-terminus. The b and y ions are crucial for peptide sequencing as they enable identification of the amino acid sequence by analyzing the mass-to-charge ratio of these fragments^{137, 138}.

Data processing and downstream analysis of proteomics data

The acquired MS/MS raw spectral data needs to be processed for peptide identification and quantification. For this, different approaches including database search or de novo sequencing approaches can be utilized^{139, 140}. In this dissertation, peptide identification was performed using the reference database search approach for all proteomic analysis described in chapters 3-5.

Reference database search for peptide identification and quantification

The reference database search approach for peptide identification and quantification relies on availability of genome sequenced data which can be used to generate the reference databases of proteins and computer algorithms that enable matching the MS/MS spectral data to peptide sequences¹⁴⁰. The protein database that has complete representation of organisms in the sample needs to be selected for such analysis. In this approach, in-silico digestion of the proteins present in the reference database is performed to generate theoretical mass spectra based on the enzyme (e.g. trypsin) and number of allowed missed cleavages. The experimentally obtained MS/MS spectra are compared with the theoretical spectra generated from reference database and selects the peptide-spectrum matches (PSM) that have best match scores and filters the PSMs based on applied false discovery rate (FDR). Several computational algorithms such as are SEQUEST, MASCOT, MS Amanda are available¹⁴⁰⁻¹⁴², however they all follow similar principle for peptide identification. In this dissertation, SEQUEST-HT search engine was used within proteome discoverer (Thermo Scientific) software. The main scoring function of SEQUEST is XCorr, it measures the closeness of fit between the experimental and theoretical spectra. XCorr score is calculated by using a correlation factor which assesses the background correlation for each MS/MS spectra. XCorr is high if the direct comparison is significantly greater than background. The XCorr is a robust method for estimating the accuracy of the match. The sequence of the candidate peptide that has highest XCorr is then used to identify the protein through a reference protein database search. During the reference database search, decoy sequences are generated from reversed or random protein sequences and are appended to the reference database. The 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 rates (FDR). Percolator, a software tool

developed for fast and accurate calculate the FDR was used within proteome discoverer workflow¹⁴³.

After the peptides sequences are identified, the next step is protein inference¹⁴⁴. Protein inference can be challenging due to the shared peptide sequences between the proteins. When these shared peptide sequences are identified in the database search, it is difficult to map them to the original protein accurately and may lead to misassignments. There are several approaches to overcome this challenge with protein inference. Of these, one approach is to only consider the peptides that are uniquely mapping to a protein; however, this approach comes with a limitation of significantly reduced identifications. In another approach, protein groups are used by clustering proteins in the database with a sequence similarity cutoff and peptides unique to the protein group are used for the analysis. There is another approach for peptide matching to protein groups, this approach uses Occam's razor principle to identify the smallest set of proteins that can explain the observed peptides and master protein, a top-ranked protein based on number of peptide sequences, sequence coverage and number of PSMs is identified in a group¹⁴⁵. After the proteins are identified by any of the described protein inference methods, protein quantification is done to obtain relative abundance values. In this dissertation, label-free quantification, a quantification technique that does not require isotope labeling of proteins is performed. Using label free quantification, the samples are independently analyzed and compared across the conditions for relative quantification. Label-free quantification of proteins can be performed using peptide peak areas or area under the curve, ion intensities or spectral counts and then sum them to obtain protein quantification values¹⁴⁶. In this dissertation, the area under the curve (AUC) approach which uses the peptide signal intensities measured through the extracted ion chromatograms is utilized for relative protein quantification.

Data processing ana statistical analysis

The resulting data with their abundance values from the reference database search needs to be further processed before performing statistical analysis for biological interpretation of the data. The data processing steps include normalization to adjust for any technical bias or variations during the analysis, imputation of missing abundance values and data filtering for robust downstream analysis. Several data normalization methods are used for proteomics data

including LOESS, variance stabilization (Vsn), mean or median normalization¹⁴⁷. The data is log-transformed as it facilitates comparison of abundance fold changes between sample groups. Data imputation of missing values helps in improving the data coverage prior to statistical analysis¹⁴⁸. In this dissertation, data is normalized using a combination of LOESS and mean centering methods and log-2 transformed using InfernoRDN software¹⁴⁹, followed by data filtering and imputation using perseus software¹⁵⁰.

In proteomics studies a large number of proteins are identified, and of these identifying the proteins that display changes in their abundances between experimental conditions is useful in answering specific research questions. To identify which proteins are displaying changes in their abundances significantly between conditions, statistical analysis such as t-test or ANOVA, followed by post-hoc tests are performed to gain biological insights. Functional analysis of the significant differentially abundant proteins is performed to identify the pathways and biological processes that are changing. Gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Clusters of Orthologous (COG) groups are some of the functional annotation tools that are used to assign functional information to proteins. Following the statistical testing and functional analysis, data is visualized using various plots such as heatmaps, volcano plots to identify trends in the proteomics data.

LC-MS/MS based metabolomics

Metabolomics is the study used to identify and quantify small molecules (<1500 Da) in biological samples¹⁵¹. With advancements in mass spectrometry, thousands of metabolites can be rapidly measured simultaneously from biological samples¹⁵². Given the complexity of metabolites in a biological sample, a separation technique is employed prior to detection using MS¹⁵³. LC-ESI-MS is a preferred analytical technique in metabolomics analysis due to the chromatographic separation that reduces sample complexity, high throughput, soft ionization capability of electrospray ionization, and enables good coverage of metabolites^{83, 154}.

Depending on the objective of the metabolomics experiment, LC-MS/MS analysis could be performed using targeted or untargeted approaches. Typically, in targeted analysis a small number of known metabolites of interest are specifically targeted for identification and

quantification¹⁵⁵. The methods for metabolite extraction, data acquisition and analysis need to be carefully selected to detect the metabolites of interest in targeted approaches. Targeted metabolomics enables validation and absolute quantification of selected metabolites of interest. Untargeted metabolomics approach allows comprehensive analysis of all the measurable analytes in a sample including chemical unknowns¹⁵². In plant-microbiome research, one of the major focus is to uncover the interactions between the plant and associated microbiome and untargeted metabolomics enables comprehensive profiling of metabolites produced in plant-microbiome systems without prior knowledge of their identity and it captures the full diversity of chemical signals and metabolites both known and novel compounds involved in these interactions¹⁵⁶. In this dissertation, an untargeted metabolomics approach was employed to study the *Populus* root and exudate profiles.

Untargeted metabolomics workflow

Metabolomics analysis involves several important steps including sample collection, metabolite extraction, LC-MS/MS measurements, data processing and analysis (**Figure 2.6**). Root exudates are incredibly complex mixtures of chemical compounds and have wide chemical diversity^{157, 158}. Therefore, considerations should be given to every step involved in the analysis for unbiased metabolomics analysis of root exudates.

Sample preparation for metabolomics

Metabolite extraction is a crucial step in sample preparation for metabolomics experiments as it determines the range and quality of metabolites extracted. Several extraction methods including solid phase extraction, liquid extraction are used for extracting metabolites based on the sample and LC-MS compatibility¹⁵⁹. Each of these extraction techniques has their own set of advantages and limitations. Liquid extraction is a commonly used extraction technique for LC-MS metabolomics where metabolites are extracted using solvents. Extractions can be performed using a single solvent, called monophasic extraction or two solvents, called biphasic extraction. The choice between these extraction methods can significantly impact the metabolite coverage. Using monophasic extractions, metabolites are extracted in a single phase. The single solvent used for monophasic extractions could be a mixture of aqueous and organic solvents that are

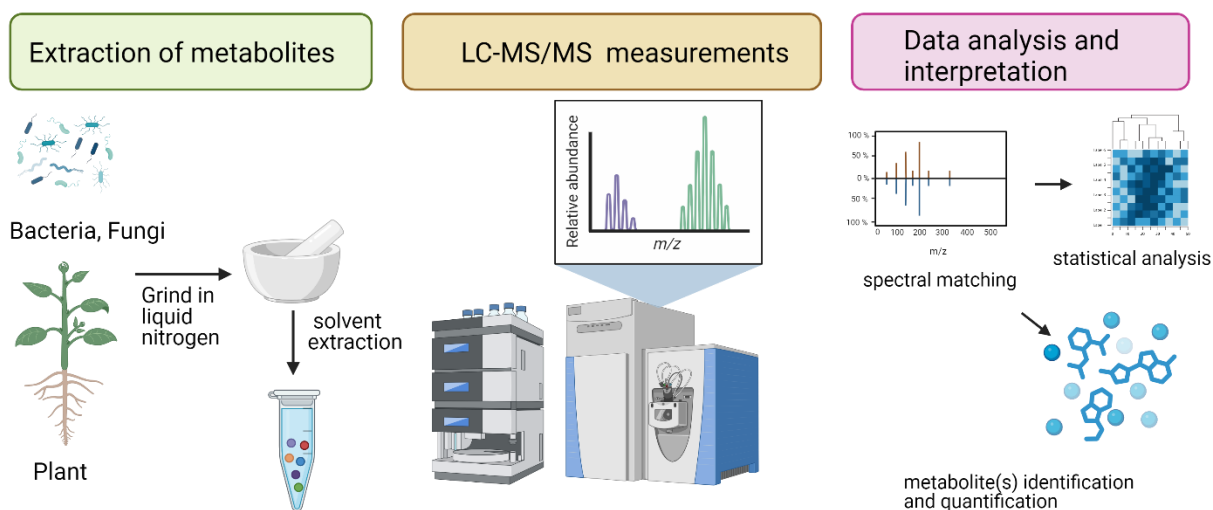


Figure 2.6: Overview of untargeted metabolomics workflow. Metabolites are extracted from samples using solvents, followed by liquid chromatography for separation and tandem mass spectrometry analysis. The raw mass spectra are processed for metabolite identification and downstream data analysis is performed for biological interpretation. Figure generated using BioRender.com.

miscible with each other. This method is simple to use, however selective extraction of certain classes of metabolites may be challenging with monophasic extractions. Monophasic extraction methods are typically used for untargeted metabolomics analysis aimed for broad metabolite coverage. Biphasic extractions use two immiscible solvents, usually an aqueous phase (e.g., water) and an organic phase (e.g., chloroform or ethyl acetate) and results in the separation of polar and non-polar metabolites into two distinct phases. This method allows selective partitioning of polar and non-polar metabolites into their respective phases, thereby facilitating efficient recovery of these compounds separately. However, biphasic extractions could be more labor-intensive and time-consuming compared to monophasic extractions. Therefore, the selection of extraction method depends on the metabolites of interest and the study goals. Root and exudate metabolites contain wide range of polar and non-polar compounds and a biphasic extraction method using equal volumes of LC-MS grade water and ethyl acetate was utilized in this dissertation to efficiently extract the wide range of metabolites produced by plants¹⁶⁰. After extraction, the aqueous extract is lyophilized to concentrate the metabolites using lyophilizer and the ethyl acetate extract is evaporated at room temperature. The metabolites are solubilized using solvents compatible for LC-MS/MS analysis.

LC-MS/MS measurements for Untargeted metabolomics

Plant root metabolites and exudate samples are complex mixtures consisting of thousands of metabolites and requires efficient separation prior to MS measurements¹⁵⁷. LC-MS/MS is one of the most widely used techniques for plant metabolomics. LC-MS analysis detects a high number of metabolites and allows coverage of thousands of most organic and some inorganic metabolites including the secondary metabolites produced by plants. It is sensitive for thermo-labile and polar metabolites^{161, 162}. The choice of various chromatographic parameters including chromatographic column, solvents used influences the data quality and completeness. The plant root exudates can vary considerably in terms of polarity. This makes it important to use proper separation technique with desired hydrophobicity or hydrophilicity to gradually elute the metabolites present in samples. The two widely used stationary phases for separation of metabolites includes the reverse phase liquid chromatography and hydrophilic interaction chromatography¹⁶³. HILIC is used for separating metabolites based on hydrophilic interactions of

metabolites with stationary phase while RP separation is based on hydrophobic interactions. In this dissertation, reversed-phase (RP) chromatography is used because it can separate a wide range of compounds, including some polar primary metabolites¹⁶⁰. A 30-minute gradient elution was employed for separation of metabolites using solvent with high aqueous content (95% water, 5% acetonitrile, 0.1% formic acid) and gradually increase the flow of solvent with high organic (70% acetonitrile, 30 % water, 0.1% formic acid). This gradient elution allows elution of both polar and non-polar metabolites. The metabolites after LC separation are analyzed by mass spectrometry. To obtain broad metabolite coverage, usually both positive and negative ion modes are utilized for untargeted metabolomics analysis, because different compounds have different ionization behaviors which lead to varying detection in across ion modes^{164, 165}. In this dissertation, the metabolites separated by RP-LC were analyzed with Q Exactive plus mass spectrometer operated in positive ionization mode only due to the large number of samples in this study. The scan range for the measurement was 135 to 2000 m/z with MS resolution of 70,000 and MS2 at a resolution of 17,500. A stepped collision energy of 10, 20 and 40 eV was utilized for fragmentation of 5 most intense precursor ions in each scan cycle. All data was acquired in profile mode.

Untargeted metabolomics data processing and analysis

Untargeted metabolomics data processing workflow includes several steps that convert the raw MS spectral data into lists of metabolite features detected, estimate their abundances and then assign chemical annotations. There are several computational programs and softwares available that are designed to accurately identify the precursor and fragment m/z of ions measured in a given experiment and compare them across all sample groups for identification and reliable quantification¹⁶⁶. Open source softwares such as MZmine, XCMS, MS-DIAL, MetaboAnalyst and vendor-based softwares such as Compound discoverer, MassHunter are available for processing of raw mass spectrometry files¹⁶⁷⁻¹⁷⁰. The open source softwares are equipped with GUI that is user friendly. All of these softwares in principle would take raw MS files as an input and carry out subsequent preprocessing steps including peak detection, extracted ion chromatogram building and resolving of different features to obtain relative quantification, followed by peak alignment of all the detected features across the entire experiment often termed

as feature alignment¹⁶⁶. Following this, for the compound annotations the softwares rely on matching the fragment spectra to reference library spectra¹⁷¹. The final desired output from these softwares would be in a tabular format with rows comprised of metabolite features detected and columns providing information for their putative identification, observed m/z, retention time, peak area in each sample group, difference between observed and measured m/z. In this dissertation, compound discoverer software (Thermo Scientific) was used to process the untargeted metabolomics data and for metabolite annotation, NIST-2020 spectral library and GNPS spectral libraries were used.

The untargeted metabolomics data is further processed by filtration, imputation and normalization. Data filtering is crucial to remove any unwanted noise in the data or to filter out metabolite features present in experimental blanks¹⁷². Further untargeted metabolomics data has missing values that need to be imputed prior to performing any downstream analysis. Various imputation methods exist, including replacing missing values with zero, k-nearest neighbors (kNN), singular value decomposition (SVD)¹⁷³. Following the data pre-processing, metabolomics data need to be normalized to remove any variation in data due to technical reasons for which several normalization methods are available¹⁷⁴. The resulting data is analyzed using various statistical analysis depending on the research goals. ANOVA or t-test are performed depending on the number of sample groups to compare the means between the groups, and significantly changing metabolites are identified between conditions. Multivariate analysis such as Principal Component Analysis (PCA), Partial Least-Squares Discriminant Analysis (PLS-DA) are used to visualize the data and explore the relationships between various sample groups in a given experiment¹⁷⁵. Various machine learning approaches including random forest classification are used in metabolomics data analysis for classifying samples based on metabolite profiles and identifying key metabolites that drive the classification¹⁷⁶. Data can be visualized using several plots such as heatmaps, bar charts, box plots etc. In this dissertation, data filtration methods applied for metabolomics data are described in detail in chapter 6. Data imputation, normalization and various statistical analysis are performed using MetaboAnalyst version 6.0¹⁷⁰ and the details are described in chapter 6.

Summary

This chapter describes the general workflows for LC-MS/MS based metaproteomics and untargeted metabolomics analysis including sample preparation, measurements, data acquisition and data analysis steps utilized in conducting the research presented in this dissertation. Briefly, proteomic and metaproteomic analysis of bacterial samples from monocultures and constructed microbial communities was performed and described in chapters 3-5. The *Populus trichocarpa* root and exudate metabolite samples were analyzed by an untargeted metabolomics approach, which is described in chapter 6. The methodologies described in this chapter were adopted or customized based on the research question to be addressed and the details are provided in individual chapters of this dissertation.

CHAPTER 3: A METAPROTEOMIC APPROACH TO STUDY MICROBIAL COMMUNITY STRUCTURE AND INTER-MICROBIAL INTERACTIONS IN CONSTRUCTED MICROBIAL COMMUNITIES

Author contributions

Text and figures were adapted from the following published journal articles:

Shrestha, H.K., **Appidi, M.R.**, Villalobos Solis, M.I. et al. Metaproteomics reveals insights into microbial structure, interactions, and dynamic regulation in defined communities as they respond to environmental disturbance. BMC Microbiology 21, 308 (2021).

<https://doi.org/10.1186/s12866-021-02370-4>

Manasa R. Appidi's contributions include metaproteomic sample preparation, metaproteomics measurements, data analysis and interpretation, generated figures, writing and editing of original manuscript.

Wang J, **Appidi MR**, Burdick LH, Abraham PE, Hettich RL, Pelletier DA, Doktycz MJ. 2024. Formation of a constructed microbial community in a nutrient-rich environment indicates bacterial interspecific competition. mSystems 9:e00006-24.

<https://doi.org/10.1128/msystems.00006-24>

Manasa R. Appidi's contributions include metaproteomic sample preparation, metaproteomics measurements, data analysis and interpretation, writing and editing of original manuscript.

Note: All the supplemental material including supplementary figures and tables mentioned in text are available in the online version of the published manuscript

<https://doi.org/10.1128/msystems.00006-24> and their numbering reflects that of the online material.

Introduction

Plants have complex microbiomes that are associated with them and impacts plant's fitness and productivity. The intra and interspecies microbial interactions within or outside plant tissues

are closely connected to essential processes such as water and nutrient acquisition, stress response, and enhancing plant resilience against diseases or herbivores by priming the host defense mechanisms¹⁷⁷⁻¹⁸⁰. Although the co-evolution of binary interactions between plants and their associated microbiomes is supported by extensive research, the complex ecological interactions between microbes and the evolution of microbe-microbe interaction mechanisms occurring in nature remains poorly understood. Additionally, to what extent does the inter-microbial interactions shape the microbial assemblages and function in nature and the molecular mechanisms used by microbes to cooperate, compete, and persist in complex microbial consortia remains unclear.

Various forms of inter-microbial interactions exist in nature such as competition for resources, synergism, antagonism, and these interactions can be modified by the influence of the environment. It is crucial to unravel these diverse and dynamic interactions between microbes and predict their competitive and cooperative potentials to fully understand the microbial dynamics in its environment. The functional contributions of individual microbes in a community including their metabolic functions, inter-species interactions, and community dynamics can be predicted using several computational and mathematical modeling approaches. In general, the modeling approaches utilize ordinary differential equations (ODEs), annotated genomes and sequence read abundances¹⁸¹. The temporal dynamics of microbial interactions can be studied using the generalized Lotka–Volterra (gLV) models based on ordinary differential equations. A recent non-linear modeling approach called compositional Lotka-Volterra (cLV) approach, based on relative abundances demonstrated its ability to predict microbial trajectories accurately over time in a community¹⁸². The extent of resource competition and microbial metabolic interactions can be predicted using constraint-based genome-scale metabolic models such as flux balance analysis (FBA)¹⁸³. Besides these computational and mathematical modeling approaches, ecology-guided models can also predict the metabolite cross-feeding interactions based on metagenomic and metabolomic data and are valuable in understanding the community dynamics¹⁸⁴. Overall, these microbial modeling approaches provide a guide to begin understanding the functional state of microbial consortia and their competitive and cooperative metabolic interactions. However, to fully understand the dynamics of microbial communities and

the underlying principles influencing them, integrating the predictive models with experimental data for example those obtained from experiments using constructed microbial consortia.

Generally, two complementary methodological frameworks can be applied to study microbial communities: holistically in natural environments or in laboratory-controlled conditions. Both approaches have contributed significantly in advancing the plant microbiome research³⁴. Studying the total rhizosphere proteome using the holistic approach can provide fundamental knowledge on interactions between plant, microbes and their surrounding environment. However, protein extraction from soil is extremely challenging because of the complexity of soil matrix which leads to low quantity and quality of the extracted proteins¹⁸⁵. Alternatively, we can achieve a detailed understanding of relatively simple subsets of a microbial community and the emergent properties that dictate microbial composition and function first and then test these in more adding additional building blocks of complexity (such as hosts, community members) for different environmental conditions (e.g., liquid media, soil). With the increasing complexity of the system, certain principles are expected to become less relevant, and these findings will begin to uncover higher-order interactions.

The defined microbial communities of cultivated microbes provide a clear experimental framework to explore microbe-microbe interactions and their mechanisms. The defined community assemblies coupled to omics-based technologies offers a powerful approach to achieve higher predictive power for understanding the genotype-to-phenotype associations. A crucial step to understand microbial community dynamics is acquiring high-resolution, quantitative data that reflects in situ conditions. The microbial community dynamics have been estimated mostly using data generated by 16 S rRNA gene amplicon sequencing. However, there are inherent limitations to these approaches¹⁸⁶. The members in microbial consortia experience active vs. dormancy dynamics and at any given point in time, only a subset of microbial population is active, while others are in a dormant state with reduced metabolic rates^{187, 188}. To understand the functional contributions of microbes to ecosystems, it is crucial to distinguish between active and inactive microbial taxa. To address this, both 16 S rRNA transcript and 16 S rRNA gene sequencing data are paired together by normalizing the measured rRNA levels by the abundances of individual members¹⁸⁸. However, due to the poor correlation between rRNA content or RNA/DNA ratios and growth rates, 16 S rRNA/rRNA gene sequencing can be best

interpreted as potential microbial activity¹⁸⁸. Metaproteomics is a powerful approach for accurate estimation of biomass from viable and functioning cells^{189, 190} and allows the large-scale identification and quantification of proteins from microbial communities which facilitates characterization of microbial membership, their functional roles and interspecies interactions in the community¹⁹⁰.

In addition to direct associations with plants, members of the rhizosphere microbiome adapt to local changes in the environmental factors including resource availability, which impacts the growth of individual microbial populations and the inter-species interactions. Stochastic changes in the microbiome assemblies have been demonstrated to have significant downstream effects and leading to formation of alternative communities that can change the ecological dynamics¹⁹¹. First, to study microbe-microbe interactions during the microbial assembly of *Populus* rhizosphere microbiome, a reductionist approach using a defined microbial community consisting of ten taxonomically diverse bacterial strains previously isolated from roots and rhizosphere of *Populus* species^{192, 193} was utilized. The assembly process was interrogated for the 10-member microbial consortia co-cultured in a defined minimal (MOPS + glucose) or a complex (R2A) growth medium by performing serial passaging using a 1:10 dilution for 15 passages to form stable communities¹⁹³. The changes in relative microbial population sizes for the community when passaged in different growth environments were analyzed using 16S rRNA gene amplicon Illumina sequencing and metaproteomic analysis^{192, 193}. The community composition and organism relative abundances stabilized after around 5 passages and resulted in a few dominant strains depending on the growth medium¹⁹³. The distinct stable microbial community compositions observed when grown in different growth media could be due to the differences in the chemical environment that largely affects the growth rates of individual microbial members. The relative abundances assessed using metaproteomics data identified *Pseudomonas* sp. GM17 to be the dominant member of the microbial community regardless of the growth media and environmental disturbances due to passaging¹⁹². Discrete transition phases of community composition were observed across the serial passages in R2A media where shifts in community membership and relative abundances of individual strains suggest the dynamic inter-species interactions during the community assembly and selection process for stable microbiome formation. The proteomic analysis showed that *Pseudomonas* sp.

expressed several proteins involved in the biosynthesis of siderophores, which provide fitness advantage to microbes. Additionally, proteins related to antibiotics and secondary metabolites, known to have biocontrol activity and facilitate antagonistic behavior were identified in proteomic analysis of *Pseudomonas* sp.¹⁹². Its high growth rate and antagonistic behavior helped *Pseudomonas* sp. GM17 to dominate in the community by outcompeting or preventing the growth of other microbial species^{192, 193}. To overcome the unfavorable conditions such as antibiotics produced by other microbes, limited nutrients or reduced oxygen levels in their surrounding environment, microbes express certain coping mechanisms to persist in the community. The analysis of proteins expressed by *Pantoea* sp. YR343 showed higher proportion of proteins related to stress responses in minimal MOPS media, while in complex media increased abundance of proteins related to carotenoid biosynthesis, motility, and antibiotics resistance were observed suggesting its adaptive mechanisms to overcome the unfavorable conditions¹⁹². This metaproteomic analysis also revealed *Bacillus* sp. formed spores as an adaptive response to the unfavorable conditions in the community¹⁹². Both 16S rRNA gene amplicon sequencing and metaproteomics results showed three members *Pseudomonas* sp. GM17, *Pantoea* sp. YR343 and *Sphingobium* sp. AP49 persisted in the stable community after 15 passages in R2A media^{192, 193}.

To further investigate how microbes form stable communities in nutrient-rich environments, without obligate metabolic interdependencies for growth and the effects of initial inoculum ratios on the community structure, the three bacterial strains *Pseudomonas* sp. GM17, *Pantoea* sp. YR343 and *Sphingobium* sp. AP49 that persisted in the stable consortium emerged from the ten-member mixture of bacteria grown under nutrient-rich conditions were selected to construct simplified 3-member communities. The three bacterial strains were co-cultured using varying initial inoculation ratios and were serially transferred every 48 hours for five passages in R2A media to determine the effects of initial inoculum ratios on the resulting community structure. The use of different inoculum ratios should provide insights into priority effects in defining community composition^{194, 195}. Plating counting and qPCR analysis results showed that the three species co-exist throughout the five passages¹⁹⁶. The relative abundances of the species calculated at different timepoints during the serial transfers using qPCR and plate counting showed that *Pseudomonas* sp. GM17 dominates the community 48 hour after inoculation in

communities with varying initial inoculum ratios¹⁹⁶. To understand the competitive interactions influencing the community composition, the initial passage was temporally probed and metaproteomic analysis was performed to dissect the functional behaviors of individual bacterial species during the assembly process of communities with varying initial inoculum ratios.

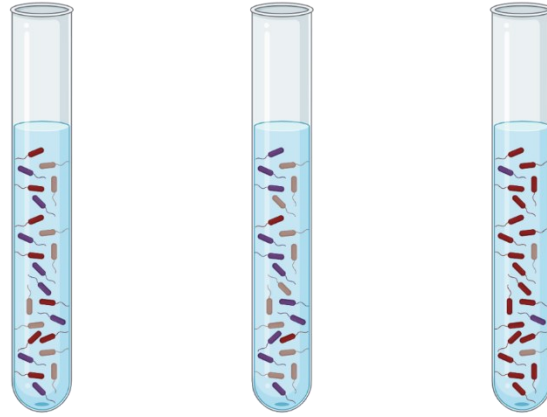
Materials and methods

Text and figures adapted from Wang et al., 2024. mSystems.

Preparation of 3-member communities with varying initial ratios

The bacterial strains *Pantoea* sp. YR343, *Pseudomonas* sp. GM17 and *Sphingobium* sp. AP49 were selected for the three-species community experiment. Bacterial monocultures were inoculated from R2A agar plates and grown in 10 mL of R2A medium overnight at 30°C with shaking at 200 rpm for preparing seed cultures. The three bacterial strains were also grown in isolation. The three-species communities were constructed under three different initial inoculum ratio conditions including 1:1:1, 1:0.001:1 and 0.001:1:0.001 of *Pantoea* sp. YR343, *Pseudomonas* sp. GM17, and *Sphingobium* sp. AP49 (**Figure 3.1**). To prepare initial inoculum for communities, an overnight culture of each bacterial strain was normalized to 10⁷ CFU/mL by calibration curve between OD₆₀₀ and CFU/mL. For community with equal initial ratios (1:1:1), one hundred microliters of each strain were combined with 9.7-mL R2A liquid medium to attain a final volume of 10 mL. For communities with 1:0.001:1 and 0.001:1:0.001 initial inoculum ratios of *Pantoea* sp. YR343, *Pseudomonas* sp. GM17, and *Sphingobium* sp. AP49, one or two of the corresponding strains were diluted 1,000-fold to 10⁴ CFU/mL and then one hundred microliters of each strain were combined with 9.7-mL R2A liquid medium to attain a final volume of 10 mL. All cultures were grown in triplicate in R2A media at 30°C with shaking at 200 rpm. The monocultures were collected at an early stationary phase, and community samples were collected at eight time points 0 hr, 4 hr, 8hr, 12 hr, 16 hr, 24 hr, 36 hr and 48 hr. At each timepoint, complete 10mL sample was collected and centrifuged to pellet the cells. Cell pellets were stored at -80°C until processed for proteomic analysis.

Initial inoculum ratios



1 : 1 : 1

1 : 0.001 : 1

0.001 : 1 : 0.001

(*Pantoea* sp. YR343 : *Pseudomonas* sp. GM17: *Sphingobium* sp. AP49)



Cell pellets collected at multiple timepoints

0, 4, 8, 12, 16, 24, 36, 48h

3 replicates for each timepoint

Figure 3.1: Three-member communities with varying initial inoculum ratios. Cell pellets were collected for metaproteomics analysis at eight timepoints 0 hr, 4 hr, 8 hr, 12 hr, 16 hr, 24 hr, 36 hr and 48 hr (n=3 for each timepoint). Figure generated using BioRender.com.

Cellular protein extraction for proteomics

All cell pellets were solubilized in 300 μ L of lysis buffer [4% sodium dodecyl sulfate (SDS) in 100-mM Tris, pH 8.0], followed by sonication (30% amplitude, 10-s pulse with 10-s rest, 2-min total pulse time using a Branson digital sonifier; Branson Ultrasonics Corporation, CT) and incubation at 90°C for 10 min. The cell lysates were centrifuged at $21,000 \times g$ for 10 min at room temperature to remove any cell debris, and proteins were quantified with a Nanodrop One spectrophotometer (Thermo Scientific). Samples were reduced and alkylated, followed by protein extraction by protein aggregation capture method and enzymatic tryptic digestion as previously described¹⁹². The samples were adjusted to 0.1% formic acid (FA) (Fisher Chemical, Fair Lawn, NJ), vortexed, and allowed to sit for 10 min. The samples were placed on a magnetic rack, and supernatants containing the peptides were added to prewet Vivaspin 10-kDa molecular weight cut off (MWCO) filters (Sartorius, Bohemia, NY) and centrifuged at $12,000 \times g$ for 15 min at room temperature. Tryptic peptide flow throughs were collected, and peptide concentrations were measured using the Nanodrop One spectrophotometer (Thermo Scientific) and transferred to the autosampler vials for liquid chromatography-tandem mass spectrometry (LC-MS/MS) measurements.

Protein identification and quantification

All peptide samples were analyzed as previously described using two-dimensional LC-MS/MS with a few modifications in separation¹⁹². Community sample peptide mixtures were separated and analyzed across three successive strong cation exchange (SCX) fractions of increasing concentrations of ammonium acetate (35mM, 50mM, and 1M). The bacterial isolate peptide mixtures were separated using the same separation regime, albeit with only one SCX fraction (1M). All MS raw data files were analyzed using the Proteome Discoverer software v.2.5 (Thermo Fisher Scientific, Cleveland, OH). Each raw data file was processed by SEQUEST HT database search algorithm¹⁴⁰ and percolator¹⁹⁷ was used to evaluate confidence in peptide-to-spectrum (PSM) matching. Peptide and PSMs with $q < 0.01$ were considered identified and at least one unique peptide sequence was required for proteins. Protein relative abundances were calculated by summing together peptide extracted ion chromatograms.

Statistical analysis

Protein abundance values were log₂-transformed, normalized using LOESS and mean-centered across entire dataset using InfernoRDN software¹⁴⁹. Data subsets for each bacterial strain were extracted and further mean-centered using InfernoRDN. For a robust downstream analysis, the analysis was limited to proteins that were identified in at least two out of three biological replicates in at least one sample group. Imputation was performed for the proteins with missing abundance values using random values drawn from the normal distribution (width 0.3, downshift 2.8) using Perseus software v.1.6.12.0¹⁵⁰. Pairwise comparisons using Student's *t*-tests were performed between different timepoints to identify the differences in protein abundance values. Proteins are characterized as significantly differentially abundant if they pass the significance threshold of *p*-value ≤ 0.05 and absolute log₂ fold-change difference > 1 . Pairwise comparisons were performed for each bacterial strain to identify the differences between the community with equal initial ratios at 48-hour timepoint and monoculture datasets.

Results

Differential proteomic analysis reveals enhanced metabolic activity in synthetic microbial community compared to monoculture in nutrient-rich conditions

The three bacterial strains *Pantoea* sp. YR343, *Pseudomonas* sp. GM17 and *Sphingobium* sp. AP49 were grown in isolation and in three-member communities with different initial inoculum ratio including 1:1:1, 1:0.001:1 and 0.001:1:0.001 of *Pantoea* sp. YR343, *Pseudomonas* sp. GM17, and *Sphingobium* sp. AP49. Proteomic analysis identified thousands of proteins for each bacterial strain grown as monocultures and collected at stationary phase (**Figure 3.2**). The metaproteomic analysis identified tens to thousands of proteins for the bacterial strains across the entire experiment including eight timepoints for the three different communities constructed with varying initial inoculum ratios (**Figure 3.3**).

To understand the microbial functions as a result of the community dynamics and interspecies interactions in a community, the proteomes of each of the three bacterial monocultures at stationary phase were compared with proteomes in the 3-member community

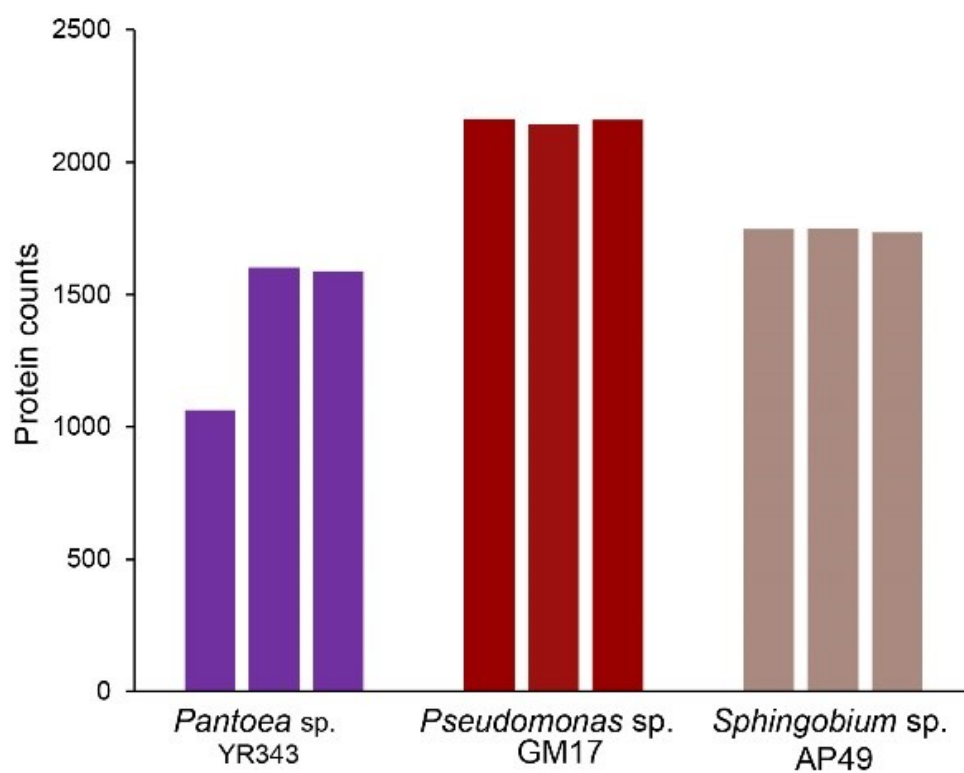


Figure 3.2: Qualitative assessment of proteomes of bacterial monocultures. Number of proteins identified for each bacterial monoculture at early-stationary phase. Each bar represents a biological replicate.

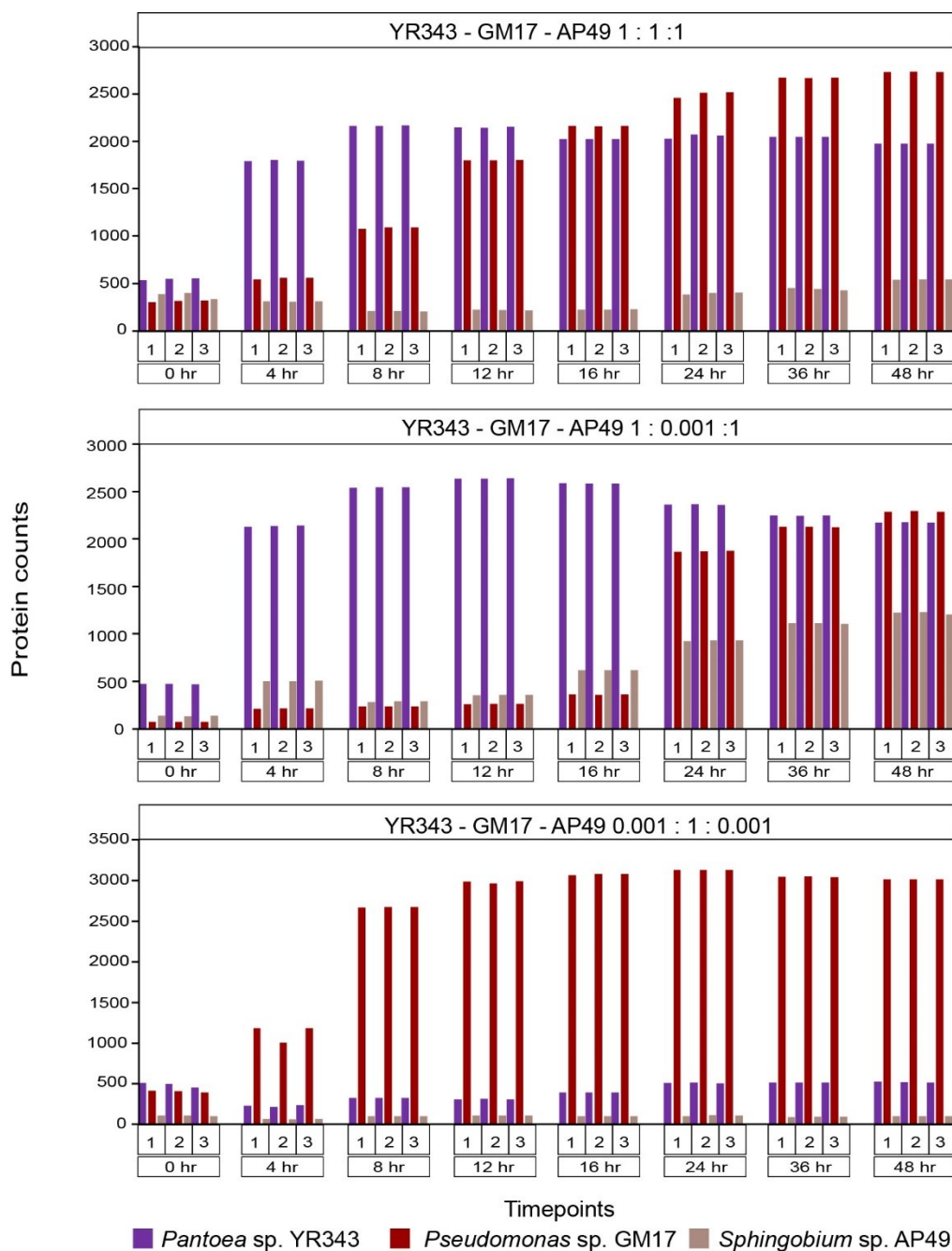


Figure 3.3: Qualitative assessment of proteomes of bacterial strains in community. Number of proteins identified for each bacterial strain in 3-member synthetic communities with varying initial inoculum ratios collected at eight timepoints. Each timepoint has 3 replicate samples represented by 1,2,3 in the figure.

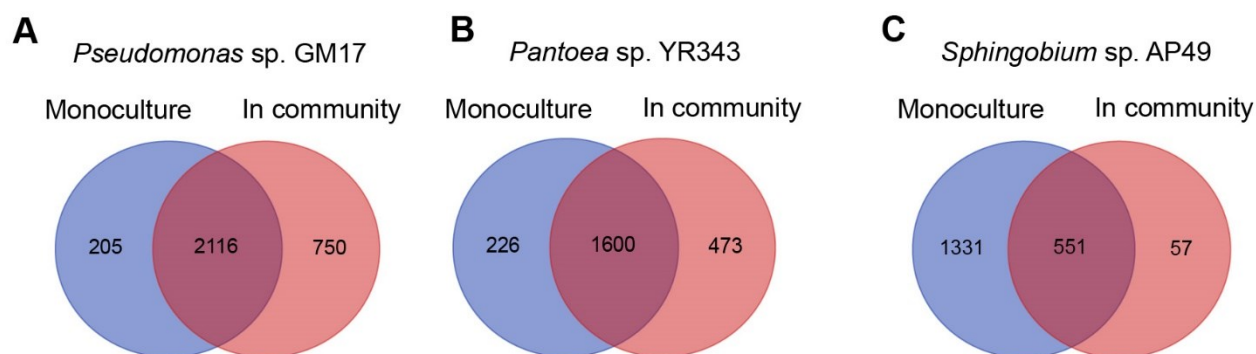


Figure 3.4: Overlap of proteins between monocultures and 1:1:1 community at 48 hr timepoint. Venn diagrams showing the number of quantifiable proteins that are uniquely identified or shared between the three bacterial strains in the three-member community with equal initial inoculum ratios (YR343-GM17-AP49 1:1:1) at 48 hr and their monocultures at early stationary phase in R2A medium (A) *Pseudomonas* sp. GM17 (B) *Pantoea* sp. YR343 (C) *Sphingobium* sp. AP49.

with equal initial ratios (1:1:1) at 48 h timepoint (Table S4, refer online version). Majority of the identified proteins were found to overlap between the monoculture and community proteomes for *Pseudomonas* sp. GM17 and *Pantoea* sp. YR343, but still had a large number of proteins that were uniquely identified when grown in the community (**Figure 3.4**). An opposite trend was observed for *Sphingobium* sp. AP49, where a higher proportion of proteins were uniquely identified when grown as a monoculture (**Figure 3.4**), which could be due to the relatively lower number of proteins detected for *Sphingobium* sp. AP49 than *Pseudomonas* sp. GM17 and *Pantoea* sp. YR343 when grown in the mixed community (**Figure 3.2 and 3.3**).

Statistical analysis using Student's t-tests was performed between proteomic data obtained from community with equal initial inoculum ratios at 48 hr and monoculture at stationary phase samples to identify the differentially abundant proteins (**Figure 3.5**). The proteomes of these bacterial strains exhibited significant differences when grown in community compared to monocultures. In *Pantoea* sp. YR343 41% of proteins were significantly differentially abundant in community, 64% in *Pseudomonas* sp. GM17 and 91% in *Sphingobium* sp. AP49. Of the significantly changing proteins, a higher proportion displayed increase in abundance in community compared to their corresponding monocultures in both *Pantoea* sp. YR343 (85%) and *Pseudomonas* sp. GM17 (81%). However, in *Sphingobium* sp. AP49, 95% of the significantly changing proteins displayed decrease in abundance in community, this could be due to fewer proteins identified in community samples for *Sphingobium*.

The proteins that showed significant increase in abundance when grown in the community compared to the monocultures for *Pseudomonas* sp. GM17 shows increased metabolic activity as indicated by the KEGG pathway analysis performed using Interactive Pathways Explorer (iPath v3, <https://pathways.embl.de/>)¹⁹⁸. Several metabolic processes including the carbohydrate metabolism, amino acid metabolism, energy metabolism, lipid metabolism displayed increased activity in *Pseudomonas* grown in community compared to monoculture (**Figure 3.6**). Similarly, *Pantoea* sp. YR343 also exhibited increased metabolic activity when grown in community (**Figure 3.7**).

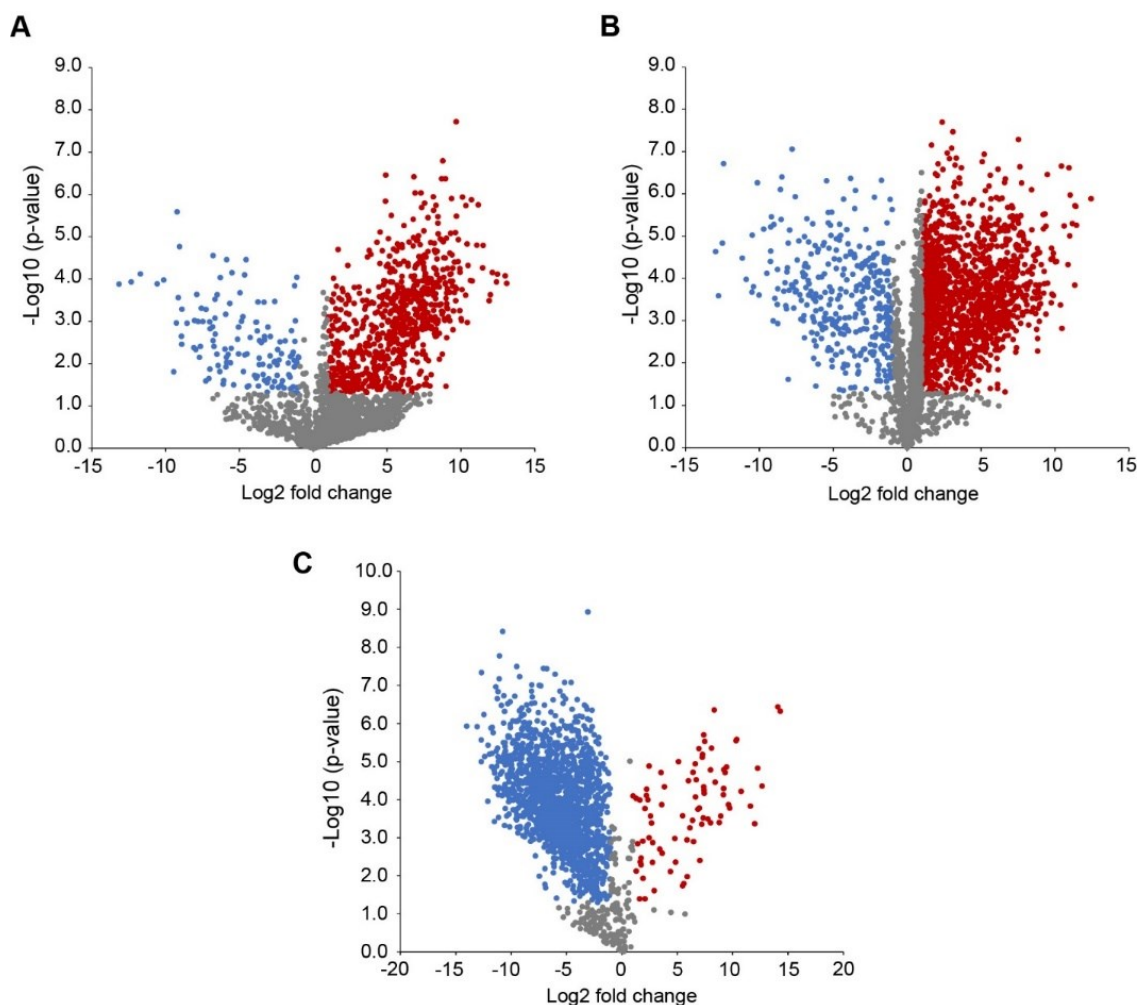


Figure 3.5: Volcano plots showing the significant differentially abundant proteins for each bacterial species at 48 hr timepoint in the three-member community with equal initial inoculum ratios (GM17-YR343-AP49 1-1-1) compared to their monocultures at early stationary phase in R2A medium (A) *Pantoea* sp. YR343 (B) *Pseudomonas* sp. GM17 (C) *Sphingobium* sp. AP49.

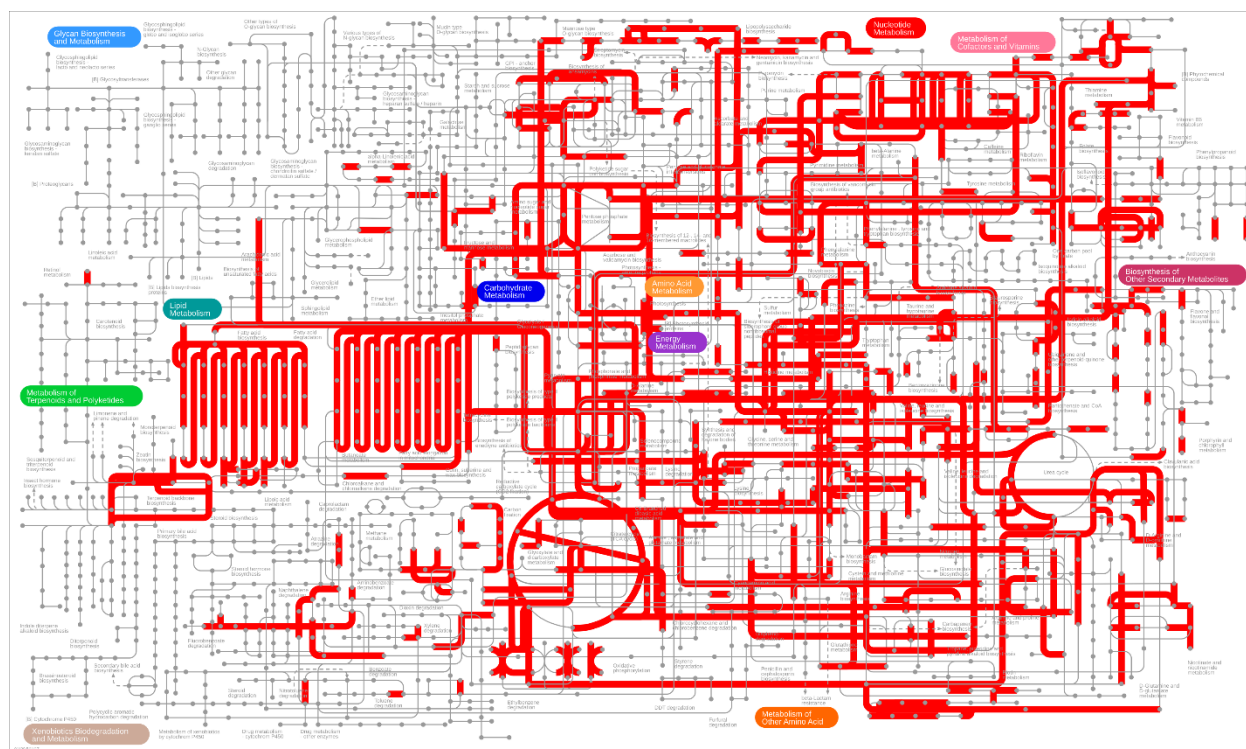


Figure 3.6: Global KEGG map of metabolic pathways for *Pseudomonas* sp. GM17. KEGG KO terms associated with *Pseudomonas* sp. GM17 proteins that displayed significant increase in abundance in community with equal initial inoculum ratios (GM17-YR343-AP49 1-1-1) compared to *Pseudomonas* sp. GM17 monoculture at early stationary phase in R2A medium.

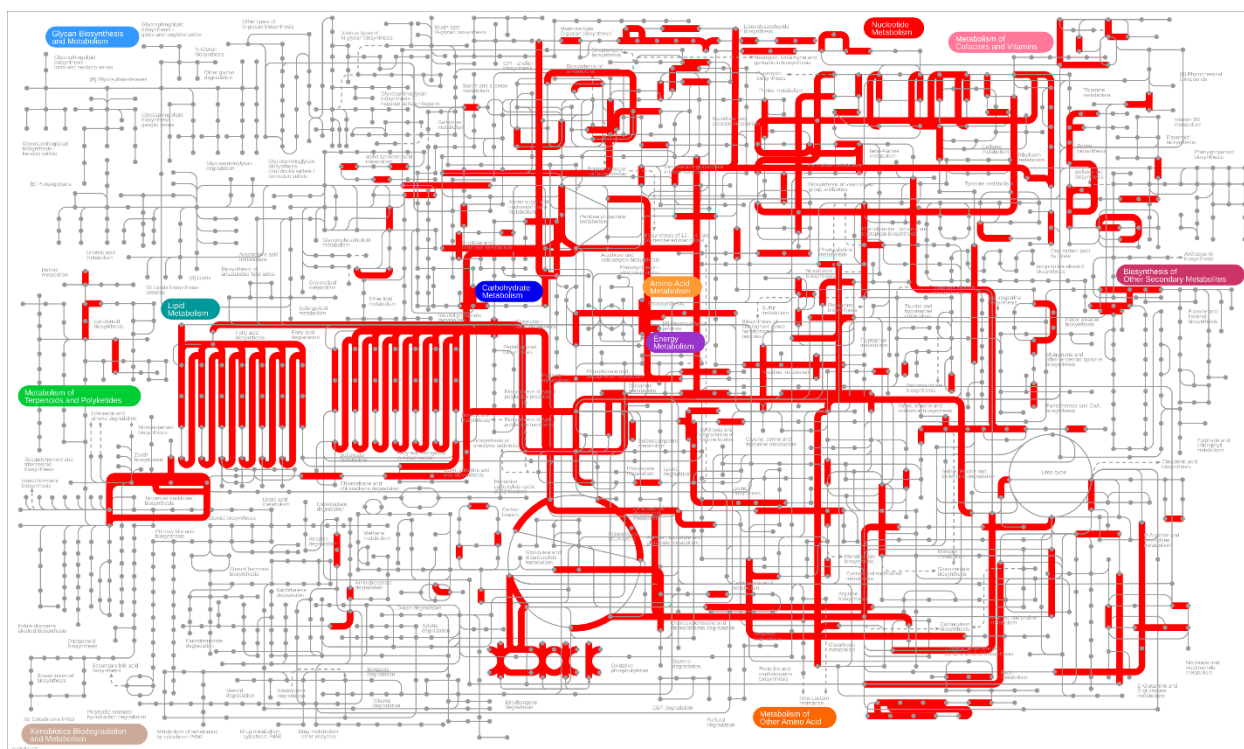


Figure 3.7: Global KEGG map of metabolic pathways for *Pantoea* sp. YR343. KEGG KO terms associated with *Pantoea* sp. YR343 proteins that displayed significant increase in abundance in community with equal initial inoculum ratios (GM17-YR343-AP49 1-1-1) compared to *Pantoea* sp. YR343 monoculture at early stationary phase in R2A medium.

The influence of various initial inoculum ratios on the resulting community structures and temporal dynamics assessed by metaproteomics

A time-series experiment across 48 hrs was performed for metaproteomics analysis to decipher the functional temporal dynamics of the three bacterial strains in 3-member communities with various initial inoculum ratios. Relative organism population size estimates were determined by summed protein abundances for the eight sampling timepoints (0, 4, 8, 12, 16, 24, 36 and 48 hr) of the 3-member communities and temporal shifts were observed in population sizes of bacterial strains in the communities with varying initial inoculum ratios. In the community with equal initial inoculum ratios, *Pseudomonas* sp. GM17 was the most abundant member (52.0%) at the end of 48 hr, whereas *Pantoea* sp. YR343 was the second most abundant member (45.2%). Although, the relative abundance of *Sphingobium* sp. AP49 (2.8%) was much less, it still maintained its presence in the community. In the community where the initial inoculum had lower amount of *Pseudomonas* (YR343- GM17-AP49 1: 0.001:1), *Pantoea* sp. YR343 was the most abundant member (57.4%) at 48 hr, however surprisingly *Pseudomonas* sp. GM17 became the second abundant strain with 32.6% relative abundance and *Sphingobium* sp. AP49 which was started with a higher amount equal to *Pantoea* was the least abundant member (10.0%). As expected, in community with higher initial inoculum of *Pseudomonas* (YR343- GM17-AP49 0.001: 1: 0.001), *Pseudomonas* sp. GM17 dominated the community at 48 h with 91.6% relative abundance. Although the initial inoculum amount of *Pantoea* was much lower, it was able to make up to 6.8% of the community at 48 hr (**Figure 3.8**).

Interestingly, the time points at which the community structure has undergone major shifts were observed to be dependent on the initial amount of *Pseudomonas* sp. GM17 in the mixture. In the community with equal initial inoculum ratios (YR343- GM17-AP49 1: 1: 1), shifts in population sizes occurred gradually between 4 hr to 16 hr, and in community where it was a minority member in the initial inoculum (YR343- GM17-AP49 1: 0.001:1), the community shifts were observed at later time points between 16 hr and 24 hr, where the relative abundance of *Pseudomonas* sp. GM17 was increased and *Pantoea*'s relative abundance was decreased (**Figure 3.8**). Interestingly, in all three initial inoculum ratio conditions at 48 hr timepoint, all three bacterial strains were able to maintain their presence although *Sphingobium* was the least abundant to other two community members. The results indicate that although initial inoculum

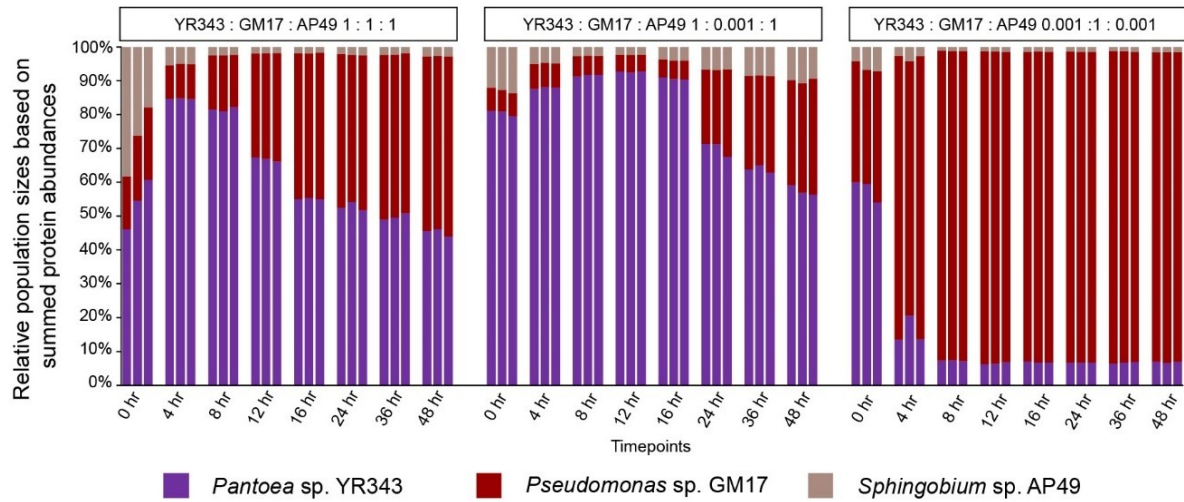


Figure 3.8: Assessment of population sizes of bacterial strains in three-member communities. Community composition assessed using summed protein abundances obtained from metaproteomic analysis of three communities with varying initial inoculum ratios of bacterial strains collected at eight timepoints, with 3 biological replicates per timepoint. Individual bar represents a replicate for a given timepoint.

ratios of the bacterial strains had a minimal effect on final community composition, it significantly affected the temporal dynamics that lead to the final community structure.

Increased metabolic activity of *Pseudomonas* sp. GM17 contributes to its dominance in the community

To understand the functions of *Pseudomonas* contributing to increase in its population size that occurred between 4hr and 16hr timepoints in equal initial inoculum community, one-way ANOVA was performed followed by hierarchical clustering analysis. The majority of the proteins that are significantly differentially abundant (90.5%, 2084 of 2302 proteins) were clustered together and displayed increasing abundance between these timepoints (**Figure 3.9A**). The gene ontology (GO) enrichment analysis was performed for the proteins belonging to this cluster using ClueGO v2.5.6, in cytoscape¹⁹⁹ and the significantly enriched GO terms belonged to several metabolic processes indicating increasing metabolic activity of *Pseudomonas* across the timepoints (**Figure 3.9B**).

Similarly, to identify the differentially abundant proteins in *Pseudomonas* sp. GM17 in GM17- YR343- GM17-AP49 1: 0.001:1 community, Student's t-tests was performed between the 16 hr and 24 hr timepoints. The proteome of *Pseudomonas* has undergone significant changes during these 8 hrs with about 89% of identified proteins displayed increase in abundance. The GO enrichment analysis of these proteins revealed the biological processes that were highly abundant at 24 hr compared to 16 hr, these include several metabolism related pathways including carbohydrate and amino acid metabolic processes, cellular macromolecule biosynthetic process, and cellular component organization (**Figure 3.10**). These results provide insights into *Pseudomonas* functional state in community, highlighting activation of several metabolic processes which support bacterial growth by providing energy and building blocks necessary for survival and growth.

Overall, the analysis of community composition at different timepoints in communities that started with varying inoculum ratios and the differential proteomic analysis for *Pseudomonas* between the timepoints that showed maximum effect on population size shifts highlighted that the temporal dynamics of these proteome changes is dependent on the organism's starting inoculation ratio. Interrogating these temporal dynamics provided detailed insights into the

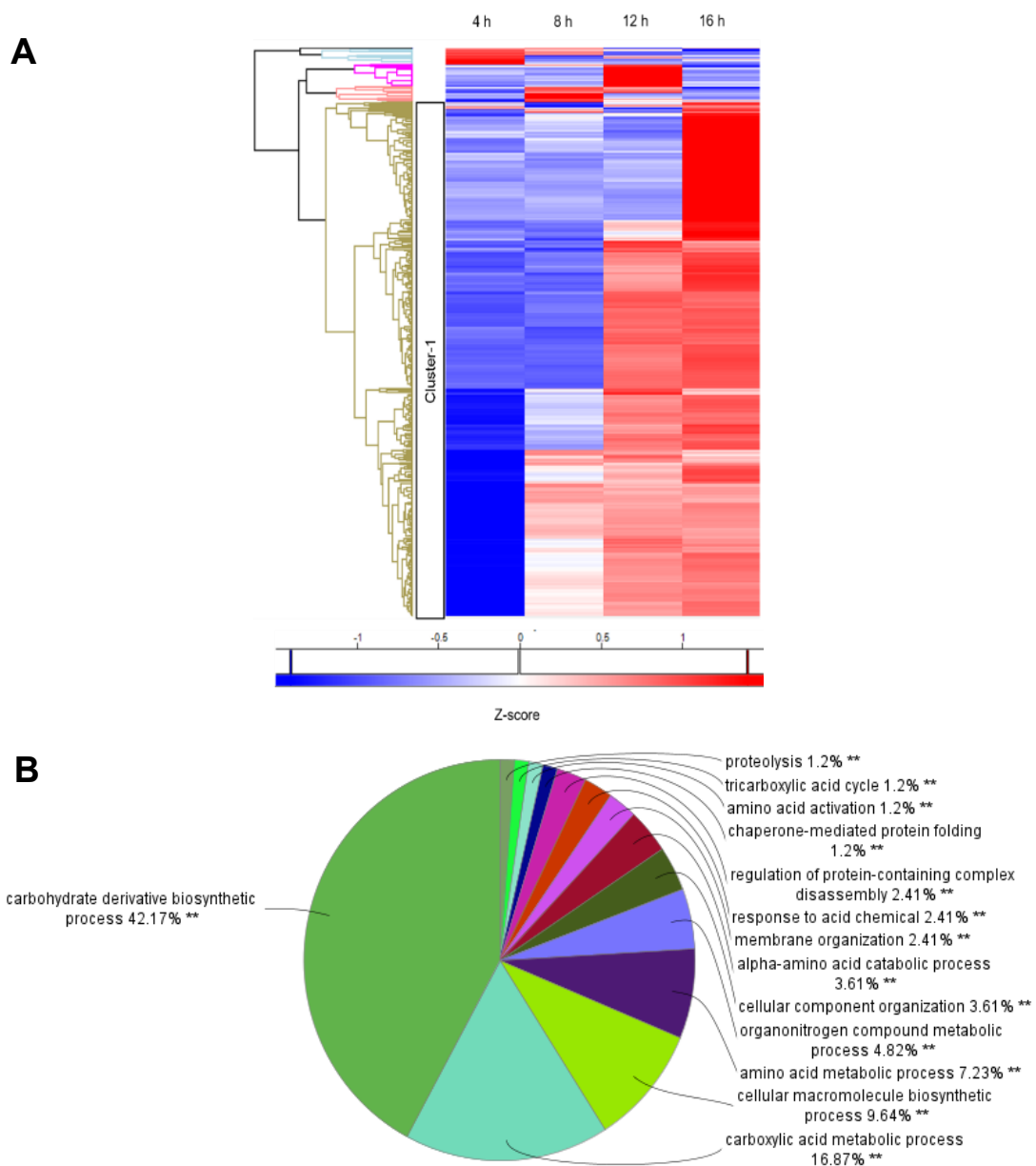


Figure 3.9: Functional analysis of *Pseudomonas* sp.GM17 in community with equal ratios. (A) One-way hierarchical clustering (rows=Euclidean distances) of Z-scores of log₂-transformed protein abundances for significant differentially abundant proteins (multiple-sample test analysis of variance was performed in Perseus software, permutation-based false discovery rate <0.05; number of randomizations=250) across 4 h to 16 h timepoints, revealed a major cluster (cluster-1) with highest membership of proteins. (B) The piechart representing the significant GO terms resulted from GO enrichment analysis of biological process category (right-sided hypergeometric test, p-value ≤0.05, Bonferroni correction) of proteins from cluster-1. The % represents the percentage of terms per group.

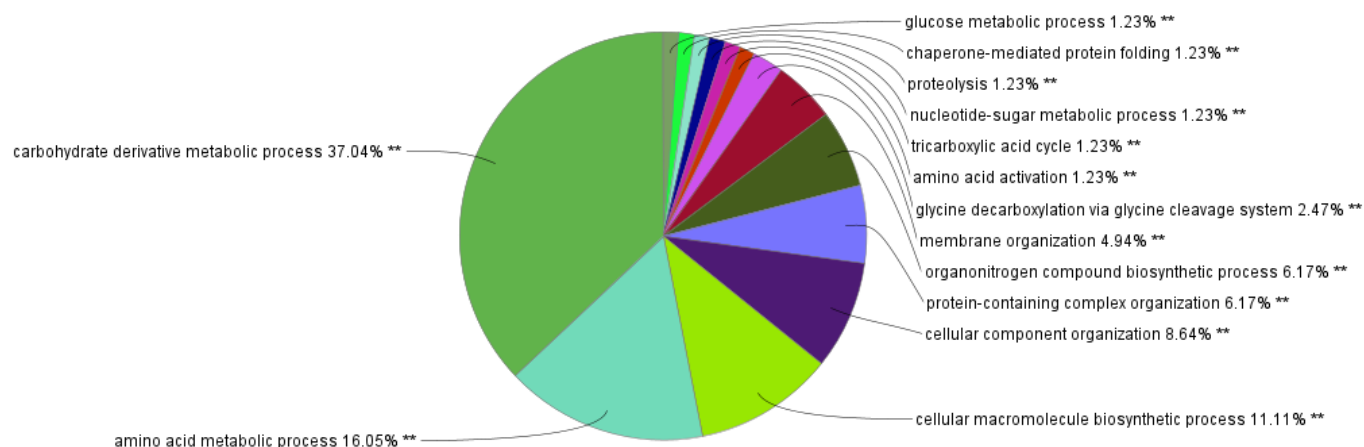


Figure 3.10: Gene ontology enrichment results for *Pseudomonas* sp. in YR343- GM17-AP49 1: 0.001:1 community. The piechart representing the significant GO terms resulted from GO enrichment analysis of biological process category (right-sided hypergeometric test, p-value ≤ 0.05 , Bonferroni correction) of proteins that showed significant increase in abundance at 24 hr compared to 16 hr timepoint. The % represents the percentage of terms per group.

increased metabolic shifts and how that correlates with changing community compositions across time. *Pseudomonas* exhibited its remarkable metabolic flexibility that allows it to grow regardless of the initial disadvantage of lower starting inoculum in this study.

Besides the metaproteomic analysis described here, the community structures were simulated by generalized Lotka-Volterra and dynamic flux balance analysis to gain mechanistic predictions into the community emergence when grown in a nutrient-rich environment¹⁹⁶. Based on the significantly decreased growth rates of these three strains when grown in community when compared to monocultures, it was predicted that the interspecies interactions in the community are competitive interactions. Both metabolic flux balance analysis and metaproteomics results showed that the competitive relationships within the community were not altered by different starting inoculum ratios¹⁹⁶.

Conclusions

Comprehensive understanding of intricate ecological interactions between microbes, including the mechanisms driving the inter-species interactions and the extent to which these interactions shape the microbial community are crucial in understanding the microbial ecosystem, yet they are not well understood. The reductionist approach presented in this chapter provided a tractable opportunity to study naturally occurring complex microbial interactions in a controlled laboratory environment. Together with high-resolution metaproteomics, an accurate assessment of microbial population sizes was obtained for the constructed microbial community while simultaneously providing valuable insights into temporal dynamics in the proteomes of bacterial strains that helped in elucidating the biological processes underpinning the final community composition across three distinct communities with varying initial inoculum ratios. The proteomic comparisons of bacterial strains when grown as a monoculture and in a community highlights the extensive proteome remodeling that occurs when an organism grows in a multispecies community under nutrient-rich conditions. Simple constructed communities similar to those described in this chapter coupled with metaproteomics can enhance our fundamental knowledge of microbial interactions and community dynamics that can further be tested by integrating more building blocks into experimental design to increase the complexity of

the system. These synthetic community studies also provide a starting point for designing and probing microbial consortia for desired biological functions in nature.

CHAPTER 4: DEVELOPMENT OF AN EXPERIMENTAL APPROACH TO ACHIEVE SPATIALLY RESOLVED PLANT ROOT-ASSOCIATED METAPROTEOMICS USING AN AGAR-PLATE SYSTEM

Author contributions

Text and figures were adapted from the following published journal article:

Appidi, M.R., Bible, A.N., Carper, D.L., Jawdy, S.S., Giannone, R.J., Hettich, R.L., Morrell-Falvey, J. and Abraham, P.E., 2022. Development of an experimental approach to achieve spatially resolved plant root-associated metaproteomics using an agar-plate system. Molecular Plant-Microbe Interactions, 35(8), pp.639-649.

Manasa R. Appidi's contributions include study design, metaproteomic sample preparation, data collection and data analysis, writing and revisions of the manuscript.

Note: All the supplemental material including supplementary figures and tables mentioned in text are available at <https://doi.org/10.1094/MPMI-01-22-0011-TA> and their numbering reflects that of the online material.

Introduction

Plants are continually interacting with microbes, some of which benefit the host while others thrive at the host's expense¹. Microbes can be found across several regions of the plant, including the phyllosphere (above the ground), endosphere (inside plant tissues both above- and belowground), and rhizosphere (belowground region surrounding the roots)⁵. There are many interacting factors that ultimately shape the composition of the associated microbiomes, with the relative success of microbial establishment and persistence largely dependent on how well an individual competes for host-mediated resources. Across the plant root system, the deposition of a complex mixture of secondary metabolites, amino acids, secreted enzymes, and cell lysates has been shown to play a role in shaping microbial communities²⁰⁰⁻²⁰² yet how microbiomes emerge and function along the plant root axis remains poorly understood.

In general, complex plant–microbe systems can be studied using either a holistic or reductionist approach. Characterizing biological systems with a holistic approach generates a more realistic assessment but with less control over the many biotic and abiotic factors influencing experimental observations. Reductionist approaches, on the other hand, offer more control but observed outcomes may be less applicable to our understanding of natural systems because nature's many complexities are not fully considered. As such, research efforts working to bridge the holistic-reductionist divide are crucial to advancing our understanding of these complex biological systems^{34, 203}. By first building a detailed understanding of relatively simple biological systems, high predictive power can eventually be achieved through combining different abiotic or biotic factors to recapitulate natural environments. As more complex systems are studied, certain principles governing plant–microbe interactions are expected to become more or less relevant, and these realizations will begin to uncover higher-order interactions influenced by experimental designs.

In the absence of plant hosts, experimental manipulations of nutrient availability on microbial community assemblage and compositional trajectory are often performed via serial dilution^{192, 193, 204}. The outcome of these experimental manipulations can uncover how different compositions of microbes may coexist or competitively exclude one another based on resource availability^{192, 205}. Understanding the fundamental implications of resource competition on microbial community assembly and functions along a plant root is important but is currently limited by methodological challenges.

Plant root-associated microbiomes are most effectively characterized by extraction methods that aim to maximize depth and coverage of the microbial biomass without disrupting the plant host tissue to minimize measurement interference⁵⁵. Of course, studies aiming to interrogate endophytic microbes within plant tissue becomes significantly more difficult due to competing plant biomass and, therefore, require more elaborate extraction procedures not described here²⁰⁶. Current methods to extract and subsequently characterize root-associated microbes are most often achieved via physical removal of microbial cells from the surface of the root via washing with or without sonication procedures²⁰⁷. It should be expected that the success of these approaches will be influenced by the overall amount of microbial biomass available for

characterization; therefore, overall root size and microbial growth and activity are important factors that must be considered.

It is now well known that different plant species select for distinct microbiome communities along their root systems through differences in rhizodeposition of resources²⁰⁸. The dynamic nature of root growth and development leads to a continuous temporal and spatial change in the amount and composition of rhizodeposits and, in turn, dynamic changes in the local microbiome compositions and functional activity²⁰⁹. For example, at root tips, plants may slough off cells and secrete mucilage whereas, further up the root, exudates are passively released via root hair cells²¹⁰⁻²¹². Currently, extraction methodologies that facilitate spatiotemporal characterization of root-associated microbiomes are yet to be realized.

Plant agar-plate culture systems are widely used²¹³ and would offer an appropriate platform for laboratory-controlled manipulation of root-associated microbiome assemblage to understand the influence of nutrient availability along the root axis. The semisolid agar matrix facilitates microbial movement and allows microbes to form communities based on the variable chemical environments surrounding the root. However, a method to spatially extract and characterize root-associated microbiomes at a molecular level using this agar-based platform has yet to be developed. Here, we describe and ground-truth a method by which microbiomes can be extracted from an agar matrix. In particular, we focus on an extraction method designed for mass spectrometry-based metaproteomics because it provides accurate estimates of both microbial relative abundances and functional activity^{189, 214}. We utilized a synthetic community (SynCom) consisting of 10 taxonomically diverse bacterial strains to characterize the spontaneous assembly and function of a *Populus trichocarpa* root-associated microbiome. The method described herein is easy to implement and facilitates spatially resolved metaproteomics of microbial communities along the plant root axis. More broadly, this method can be used to characterize a myriad of biological systems employing agar-based culture systems.

Materials and methods

Preparation of plants, bacterial cultures, and a SynCom

Populus trichocarpa (genotype B819) shoot tips were cut from greenhouse-grown plants and sterilized for 4 h by rotating in a 50-ml solution of 1× Murashige-Skoog (MS) basal media with vitamins and glycine (Caisson) and 5% plant-preservative mixture (PPM) (Plant Cell Technology). Once sterilized, the shoot tips were rinsed in sterile H₂O and rooted in a growth chamber (16-h daylength, 15.6 μmol full-spectrum LED) in 1× MS basal medium with vitamins and glycine (Caisson), 0.05% morpholineethanesulfonic acid hydrate (Sigma-Aldrich), 3% sucrose, 0.1% PPM (Plant Cell Technology), 0.5% activated charcoal (Sigma-Aldrich), and 0.18% Gelzan (Plant Media) to establish stock plants. After rooting and growing to an adequate size, axenic shoot tips were subcultured from stock plants and placed into new enclosed containers of the tissue culture media described above and grown for 3 weeks in the same growth chamber until rooted.

A previously described microbial SynCom consisting of 10 bacterial strains previously isolated from *P. deltoides* root systems were used^{192, 193}. Bacterial strains used in this SynCom were *Bacillus* sp. BC15, *Caulobacter* sp. AP07, *Duganella* sp. CF402, *Pantoea* sp. strain YR343, *Paraburkholderia* sp. strain BT03, *Pseudomonas* sp. strain GM17, *Rhizobium* sp. strain CF142, *Sphingobium* sp. strain AP49, *Streptomyces* sp. strain YR139, and *Variovorax* sp. strain CF313^{16, 215-222}. Overnight cultures of all 10 bacterial strains were grown in R2A liquid media (Teknova)²²³ at 28°C and 250 rpm. For the SynCom, the 10 bacterial cultures were normalized to an optical density at 600 nm of 2.40 and equal volumes of the 10 cultures were combined to generate the bacterial inoculum used in these studies.

Plant-agar culture growth conditions

Rooted *P. trichocarpa* plants were removed from initial growth media (which leaves the plant roots mostly clean with no debris) and roots were submerged in 25 ml of 1/4× Hoagland's media (Hoagland modified basal salt mixture; PhytoTech Labs) supplemented with 0.1% sucrose, 0.3% agarose, and the SynCom community at a concentration of approximately 10⁷ cells/ml in a sterile

Petri dish. Each plant-agar culture was incubated for 24 h ($n = 3$) or 48 h ($n = 3$) in a sterile plastic container to maintain humidity at 28°C in growth chambers with 16-h days alongside agar control plates ($n = 3$ for 24 and 48 h) that were similarly produced but lacked plant hosts (**Figure 4.1A**).

Extracting microbiomes directly associated with the plant root system for metaproteomics

To characterize microbiomes directly associated with the root surface, plant roots (biological triplicates) were excised with a sterile scalpel and transferred to Eppendorf tubes and washed in 300 μ l of sodium dodecyl sulfate (SDS) lysis buffer (4% SDS [Sigma-Aldrich] in 100 mM Tris, pH 8.0) by vortexing gently to liberate microbial cells bound to roots. The roots were removed, and crude lysate suspensions were sonicated (30% amplitude, 10-s pulse with 10-s rest, 2-min total pulse time) (Branson digital sonifier) and boiled at 90°C for 10 min to complete microbial lysis. The cell lysates were centrifuged at $21,000 \times g$ for 10 min at room temperature to remove any cell debris. Samples were reduced by adjusting to 10 mM dithiothreitol (Sigma Life Science) by incubating at 90°C for 10 min, followed by alkylation with iodoacetamide (Sigma Life Science) added to a final concentration of 30 mM and incubation for 15 min in darkness. Proteins were then extracted from the precleared lysates by protein aggregation capture¹²⁸. Briefly, Sera-Mag beads (5 μ l at 50 μ g/ μ l) (GE Healthcare Life Sciences) were washed with liquid chromatography-mass spectrometry (LC-MS)-grade acetonitrile (ACN) (EMD Millipore Corp.) on a magnetic rack. Tubes were removed from the magnetic rack, and lysate was added and then adjusted to 70 % ACN to induce protein aggregation onto beads. Samples were allowed to settle for 10 min, gently resuspended, and allowed to settle for another 10 min. The samples were placed on a magnetic rack and the supernatant was removed, followed by washing with 1 ml of ACN and 1 ml of 70 % ethanol (EMD Millipore Corp.) while on a magnetic rack. The samples were removed from the magnetic rack and 50 μ l of Tris buffer was added. Proteins bound to beads were processed with two sequential aliquots of sequencing-grade trypsin (Thermo-Scientific) for an overnight and 3-h enzymatic digest using a 1:75 (wt/wt) protein/trypsin ratio at 37°C with constant agitation (1,000 rpm). The samples were then adjusted to 0.1 % formic acid (FA) (Fisher Chemical), vortexed, and allowed to sit for 10 min. The samples were placed on a magnetic rack and supernatants containing the peptides were added to prewet Vivaspin 10 kDa MWCO filters (Sartorius) and centrifuged at $12,000 \times g$ for 15 min at

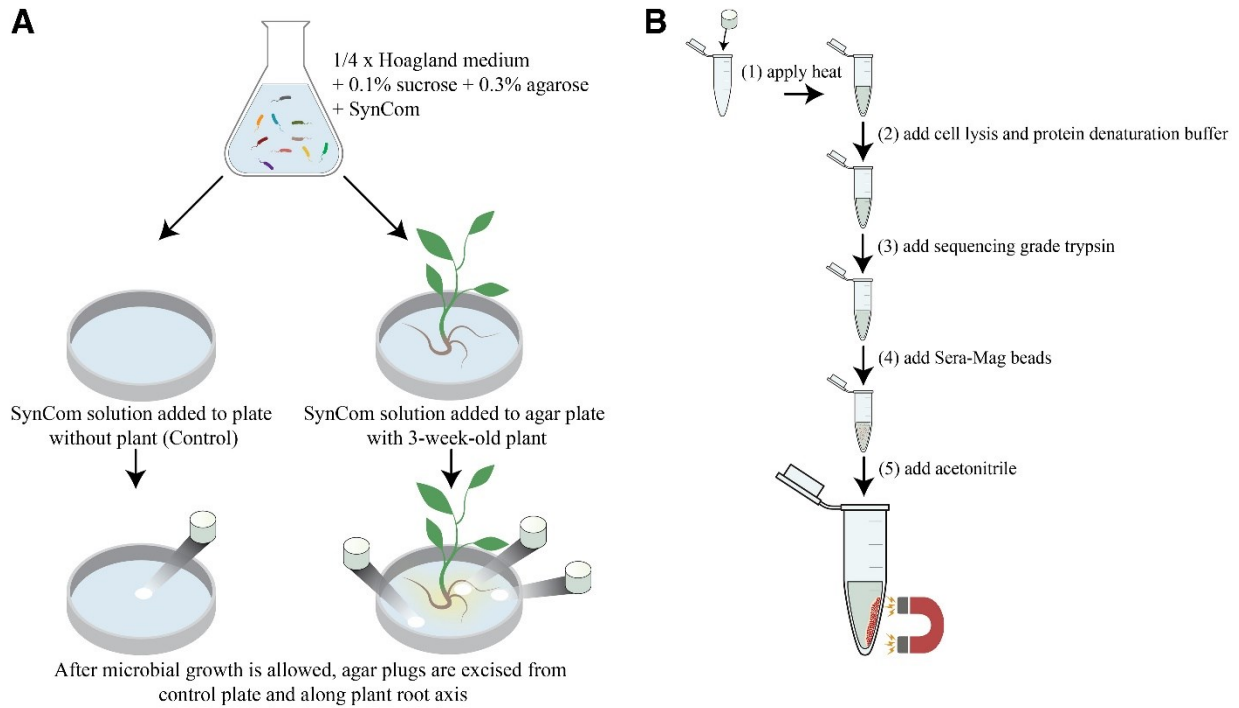


Figure 4.1: Experimental design to achieve spatially resolved plant root-associated microbes using an agar system. (A) A synthetic community (SynCom) containing 10 bacterial strains in equal numbers inoculated in 1/4× Hoagland's media + 0.1% sucrose + 0.3% agarose; control agar plates containing SynCom without a plant; plant (*Populus trichocarpa* B819) in agar containing SynCom; agar plugs excised from control agar plates containing SynCom without plant at 0, 24, and 48 h (n = 3); and agar plugs excised from root base, root tip, and distant regions for each plant in agar containing SynCom at 24 and 48 h (n = 3). (B) In-solution cell lysis, enzymatic protein digestion, and peptide extraction using Sera-Mag beads.

room temperature. Tryptic peptide flowthroughs were collected, and peptide concentrations were measured using the Nanodrop One spectrophotometer (Thermo-Scientific) and transferred to the autosampler vials for LC tandem MS (MS/MS) measurement.

Note that the root wash extraction described above was performed on a sample collected 48 h post inoculation ($n = 3$) with the 10-member SynCom and this measurement was paired with a “whole root plug” extract that is described below. These paired samples facilitated a metaproteomics comparison between microbiomes adhered to the root (i.e., root wash) versus those in close proximity (i.e., agar plug) (discussed above).

Extracting microbiomes from agar for metaproteomics

Microbial cell populations in agar were selectively excised using a sterilized scalpel to assess microbiome changes in composition and function across time and in response to the plant root system. To characterize how the SynCom microbiome changes across time in the absence of a plant root system, agar plugs were obtained from culture plates that lacked a plant root system at 0 h (control plug 0 h), 24 h (control plug 24 h), and 48 h (control plug 48 h) in biological triplicate. Note, because the amount of microbial cells surrounding a plant root system are expected to change for different plant hosts and starting inocula, the size (i.e., volume) of the agar plug should be evaluated per biological system. In this study, agar plugs equivalent to 0.5 ml were sufficient for each sample and, thus, used throughout this study. The agar plugs obtained at 0 h were excised immediately after the media turned into a semisolid state in order to assess whether extractions from agar alter microbiome measurements when compared with those obtained from the starting SynCom inoculum. The SynCom liquid control stock sample was prepared in biological triplicate by washing pelleted cells with phosphate-buffered saline to remove media and resuspending cells in 100 mM ammonium bicarbonate.

For plant-agar cultures, agar plugs were excised to obtain four spatial measurements: whole root, root tip, root base, and a region distant from the plant root system (approximately 3 to 4 cm distant from root, near the edges of the Petri dish). Whole-root agar plug samples were collected at 48 h and the other locations were collected at 24 and 48 h in biological triplicate. To ensure sufficient material for subsequent extractions, multiple agar plugs were pooled together; for

example, agar plugs from all root tips for each plant were pooled. Prior to additional sample processing, all agar plugs (0.5 ml) were liquified by heating for 10 min at 90°C in a heat block.

Samples were processed by adding an equal volume of lysis buffer (4% sodium deoxycholate [SDC] in 100 mM ammonium bicarbonate) (Sigma-Aldrich) to bring the final concentration to 2% SDC and kept warm to avoid solidification of agar. Note, SDC was selected as the detergent because, unlike SDS, it permits trypsin activity and is easily removed²²⁴. The samples were sonicated (30% amplitude, 10-s pulse with 10-s rest, 2-min total pulse time) (Branson digital sonifier), then incubated at 90°C for 10 min. The cell lysates were centrifuged at $21,000 \times g$ for 10 min at room temperature to remove any cell debris and proteins were quantified with a Nanodrop One spectrophotometer (Thermo-Scientific). Samples were reduced and alkylated as described in the section above. Samples were heated again briefly to reliquify the agar and allowed to cool down; then, proteins were processed with two sequential aliquots of sequencing-grade trypsin for an overnight and 3-h enzymatic digest using a 1:75 (wt/wt) protein/trypsin ratio at 37°C with constant shaking (1,000 rpm). Each sample was acidified with 1% FA, vortexed, and centrifuged at maximum speed for 10 min at room temperature to precipitate and remove SDC. Samples were transferred to new Eppendorf tubes where Sera-Mag beads (GE Healthcare Life Sciences) were used to aggregate and remove the agar from the samples. For each sample, Sera-Mag beads (approximately 5 μ l at 50 μ g/ μ l, previously washed and resuspended in equal volume of ACN) were added and final volumes were adjusted to 70% LC-MS-grade acetonitrile (EMD Millipore Corp.). Twice, samples were vortexed gently and allowed to rest for 10 min to allow for aggregation of the agar matrix onto the magnetic beads. Eppendorf tubes were then transferred to a magnetic rack where the supernatant solution containing peptides was removed and subsequently transferred to new tubes (**Figure 4.1B**). Importantly, care was taken to not aspirate Sera-Mag beads. To remove residual SDC, hydrated ethyl acetate was added to each sample at a 1:1 (vol/vol) ratio, vortexed, and centrifuged for 10 min at maximum speed. The lower aqueous layer containing the peptides was collected and cleaned again with ethyl acetate. The final lower aqueous layer containing peptides was transferred into new Eppendorf tubes, then vacuum-dried to completion using a SpeedVac Concentrator (Thermo-Fischer Scientific). The peptides were resolubilized in 0.1% trifluoroacetic acid and cleaned on a Pierce peptide desalting spin column (Thermo Scientific) per the manufacturer's instructions. Concentrated

peptides were resuspended in a 25- μ l aqueous solution (0.1% FA) and subsequently quantified using a Nanodrop One spectrophotometer (Thermo Scientific). Peptides were transferred to auto-sampler vials for LC-MS/MS measurements.

Protein identification and quantification

All peptides were analyzed using two-dimensional LC on an Ultimate 3000 RSLCnano system (Thermo-Fisher Scientific) coupled with a Q Exactive Plus mass spectrometer (Thermo-Fisher Scientific). For each sample, a 2- μ g inject of digested peptide mixture was flowed across an in-house-built strong cation exchange (SCX) Luna trap column (5 μ m by 150 μ m by 50 mm) (Phenomenex) followed by a nanoEase symmetry reversed-phase (RP) C18 trap column (5 μ m by 300 μ m by 50 mm) (Waters) and an aqueous solvent (5% ACN and 0.1% FA), then washed for 10 min. The loaded peptide mixtures were separated and analyzed across three successive SCX fractions of increasing concentrations of ammonium acetate (35 mM, 50 mM, and 1 M), each followed by a 100-min linear organic gradient (250 nl/min flow rate) from 5% aqueous solvent (5% ACN and 0.1% FA) to 30% organic solvent (80% ACN and 0.1% FA) to separate peptides across an in-house-pulled nanospray emitter (75 μ m by 350 mm) packed with C18 Kinetex RP C18 resin (1.7 μ m) (Phenomenex). All MS data were acquired with Thermo Xcalibur (version 4.2.47) using the top N method, where N could be up to 10. Target values for the full-scan MS spectra were 3×10^6 charges in the 300 to 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 was performed by higher-energy C-trap dissociation 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 50-ms maximum injection time. 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 2.5) (Thermo-Fisher Scientific)²²⁵. Each MS raw data file was processed by the SEQUEST HT database search algorithm¹⁴⁰ and confidence in peptide-to-spectrum (PSM) matching was evaluated by Percolator¹⁹⁷. Peptide 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. Organism relative

abundances were calculated by dividing their summed protein abundances from the sum total of all quantified protein abundances^{189, 192}.

Statistical analysis

Proteins with at least one unique peptide were exported from Proteome Discoverer. Log₂-transformation of protein abundances was performed followed by local regression (LOESS) normalization and mean-centering across the entire dataset using InfernoRDN software (v1.1.7793)¹⁴⁹. To identify the differences in the proteome expression of the bacteria between different sample groups, PERMANOVA was performed across the control, distant from root, root tip, and root base samples at 24- and 48-h time points using the vegan package (v.2.5-7) in R²²⁶. Protein abundances subset for each bacterial strain of the SynCom were extracted from the normalized dataset and further mean-centered using InfernoRDN. To improve the robustness of downstream analysis, the analysis was limited to proteins that were identified in at least two of three replicates of at least one sample group. 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)¹⁵⁰. To identify the proteins that are differentially expressed across the measured proteomes, ANOVA was performed (permutation-based false discovery rate [FDR] < 0.05; number of randomizations = 250) across experimental conditions. Proteins that passed the significance threshold of P value ≤ 0.05 were characterized as significantly differentially abundant. One-way hierarchical clustering of standardized Z-scores using Euclidean distances for rows was performed for significantly changing proteins. For the clusters obtained, GO enrichment tests were performed using a Fischer's exact test using the OmicsBox software tool (v.1.4.11) (<https://www.biobam.com/omicsbox>). Biological process and molecular function GO terms that had an FDR-controlled P value < 0.05 were considered enriched.

Results and discussion

A novel approach to extract and characterize root-associated microbiomes using metaproteomics

Currently, extraction methods employing washing are used for metaproteomic characterization of root-associated microbiomes^{56, 227}. Here, we evaluated the performance of a detergent-based wash solution (i.e., root wash) and compared it with the novel peptide extraction method from agar (i.e., agar extraction) (discussed below). For both methods, high-resolution mass spectrometry-based metaproteomics was employed to characterize the microbiome directly associated with a 3-week-old *P. trichocarpa* root surface 48 h post inoculation with a 10-member SynCom.

We found large differences in the number of total proteins identified and the plant root protein contamination between the two methods. For the root wash method, we identified 5,776 proteins, and approximately 37% (**Figure 4.2A and 4.2B**) of these were from *Populus* roots. Although metaproteomics has been applied in several root-related studies^{56, 228, 229}, to our knowledge, there are no previous measurements that could be used to appropriately benchmark these results. Moreover, comparing these results to other studies is not trivial given that microbiome complexity (composition and function) is context specific. Nevertheless, qualitatively, these measurements achieved a similar depth in the number of organisms and proteins identified when compared with a measurement performed on a cell pellet obtained from the inoculum used for SynCom inoculation (**Figure 4.3A**). In general, this comparison suggests that the amount of biomass available for metaproteomic measurements in this plant–microbe root system was not a limitation.

The ability to selectively extract and characterize microbiomes from the agar surrounding the whole plant root system is expected to improve the depth and context of metaproteomic measurements for three major reasons: (i) avoiding an extraction directly from the root surface is expected to reduce the number of identified plant proteins; (ii) the amount of microbial biomass is not limited by root surface area; and (iii) spatial sampling provides locational context which, if varied across the root or nutritional gradients, enables the sampling of potential niche members

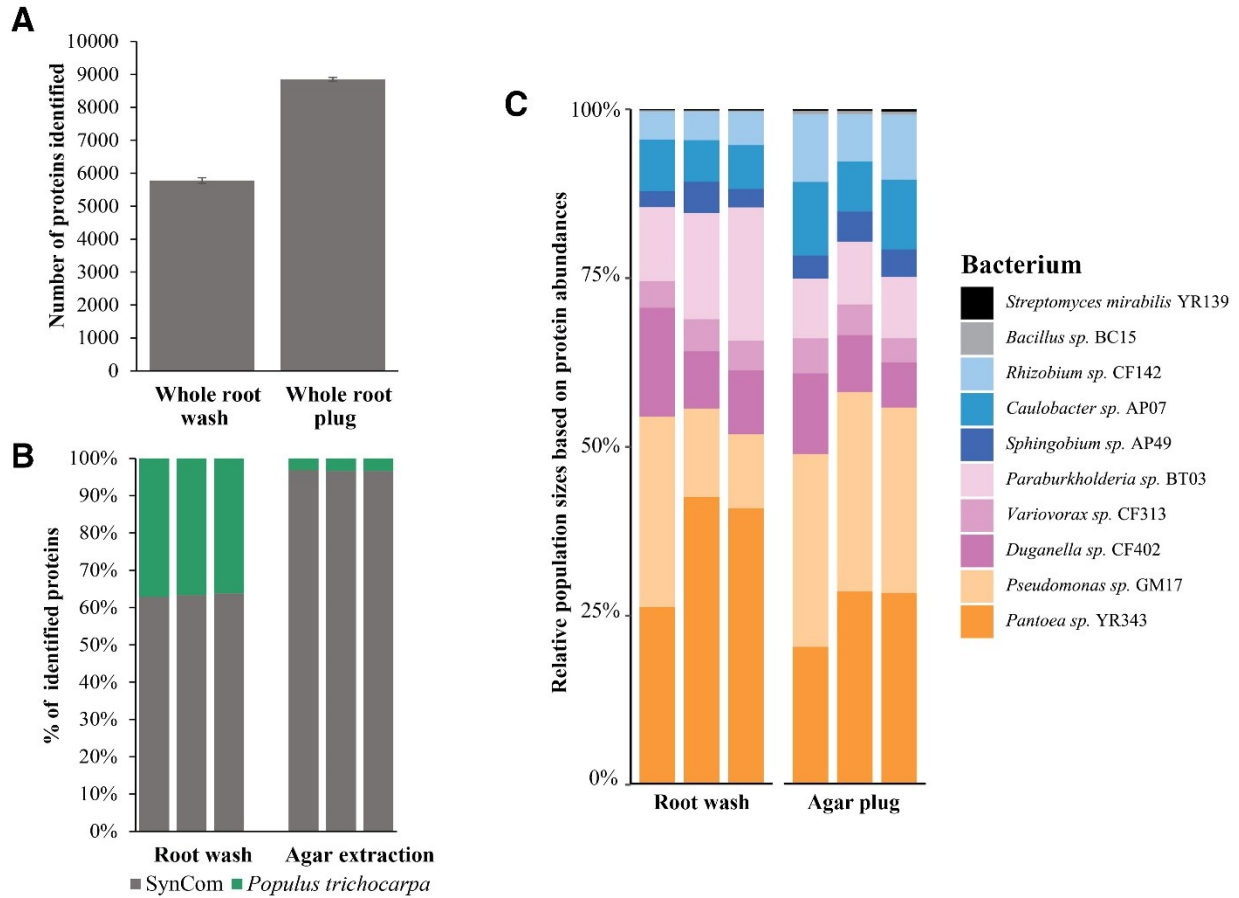


Figure 4.2: Proteins extracted from root-associated microbiomes using different extraction methods for plant-agar samples at 48 h. (A) Average number of proteins identified ($n = 3$). Error bars represent standard error of mean. (B) Percentage of proteins identified for synthetic community (SynCom) and *Populus trichocarpa*. (C) Relative organism abundances for the bacterial strains using their summed protein abundances.

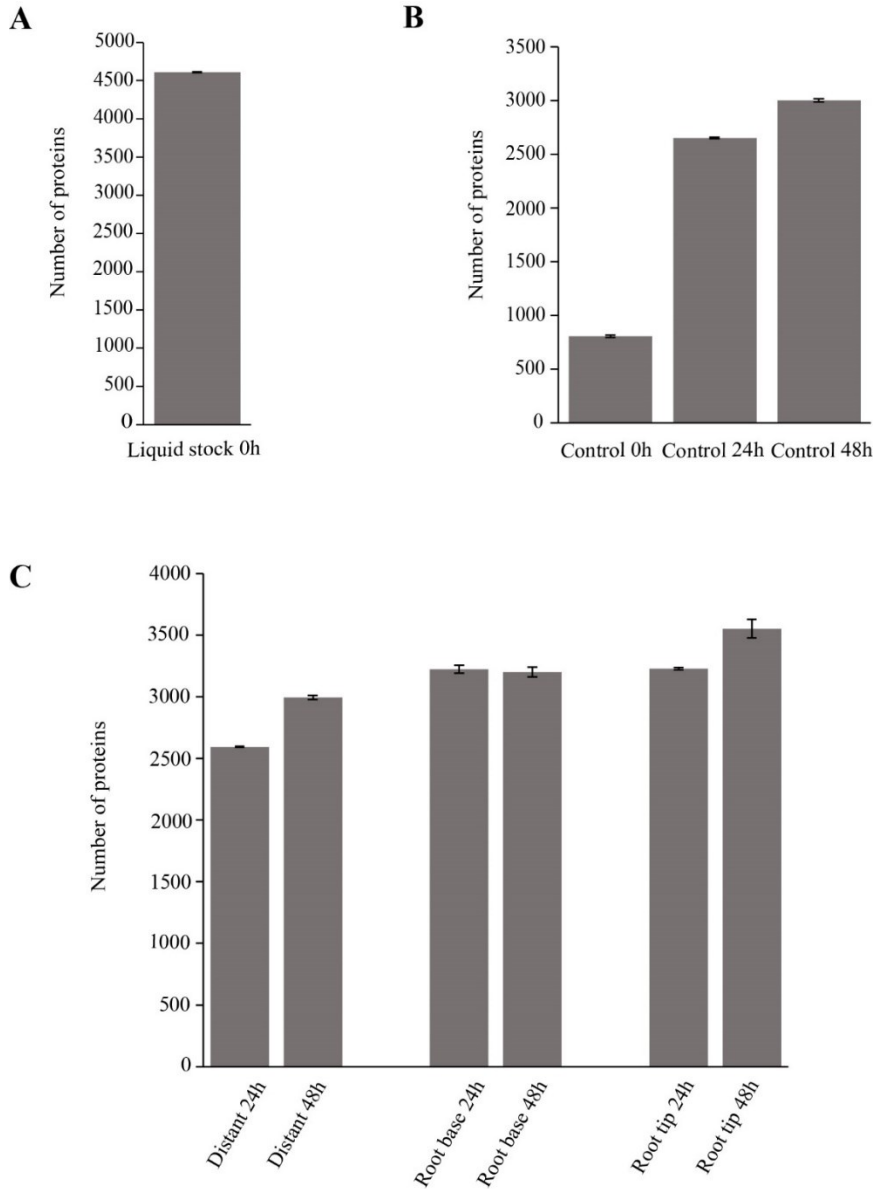


Figure 4.3: Global proteomics metrics using novel agar extraction method. (A) The average number of proteins identified from control liquid stock 0-h. (B) The average number of proteins identified from control plugs at 0-h, 24-h and 48-h (n=3). (C) The average number of proteins identified from root tip, root base and distant from root plugs at 24-h and 48-h (n=3). Error bars represent standard error of mean.

or functional signatures that would otherwise be missed. Therefore, we compared the results from the root wash (whole root wash, 48 h) to an experimentally paired agar plug excised from the surrounding root (whole root plug, 48 h). Indeed, the overall metaproteomics depth achieved by this novel extraction method far surpassed the washing method alone (approximately 53% gain in protein IDs). In total, the extract and measurement from the collected whole root agar plug resulted in a higher number of proteins identified (8,853 proteins) with a notably lower percentage of plant proteins (3.3%) observed (**Figure 4.2A and 4.2B**). Additionally, approximately 90% of the proteins identified by the root wash method overlapped with the associated whole root agar plug.

Although the overall depth of the metaproteome coverage improved, the microbiome composition is expected to be spatially influenced by the plant root system. Therefore, we used metaproteomics to estimate relative microbial biomass amounts¹⁸⁹ (**Figure 4.2C**) (Supplementary Table S1, refer online version) and performed an analysis of variance (ANOVA) test to assess the similarity of composition and relative abundance of organisms between the microbiome directly associated with the plant root system (root wash) and the microbiome in agar immediately surrounding the root (agar extraction). For this particular comparison, the overall microbiome composition was statistically unchanged ($P = 1$). Individually, however, a relatively low-abundance member (*Rhizobium* sp. strain CF142) did slightly vary in abundance between the two extractions (**Table 4.1**). The observed change in the abundance of this particular microbe could be indicative of a biological deviation from the averaged microbial population directly interacting with the root surface; however, an alternative explanation is that the increased metaproteomics depth resulting from reduced plant root protein “contamination” improved the accuracy of the estimated average composition of root-associated microbes. Furthermore, the agar-plug extraction method increased measurement reproducibility, as demonstrated by derived organism ratios across replicates (**Figure 4.2C**). Regardless, the estimated microbial compositions are nearly indistinguishable between the two approaches but, more importantly, the success of the agar-plug extraction method provides exciting new opportunities to achieve spatially resolved metaproteomics along the plant root axis.

Table 4.1: LIMMA results using the relative abundances of the individual taxa comparing the agar sample to the root wash. Positive log fold change indicates increase in agar relative to root. A Benjamini-Hocherberg correction for multiple tests was applied.

Organism	logFC	adj.P.Val
<i>Rhizobium sp.</i> CF142	0.0403	0.0396
<i>Bacillus sp.</i> BC15	0.0014	0.0622
<i>Pantoea sp.</i> YR343	-0.0968	0.3069
<i>Pseudomonas sp.</i> GM17	0.0973	0.3099
<i>Caulobacter sp.</i> AP07	0.0201	0.3099
<i>Paraburkholderia sp.</i> BT03	-0.0321	0.316
<i>Duganella sp.</i> CF402	-0.0319	0.3309
<i>Streptomyces mirabilis</i> YR139	0.0004	0.3309
<i>Sphingobium sp.</i> AP49	0.0017	0.8825
<i>Variovorax sp.</i> CF313	-0.0004	0.9401

Spatially resolved metaproteomics along the plant root axis

Spatiotemporal patterns of microbiome assembly, organization, and function are expected to exist along the plant root axis^{209, 230} yet it remains unknown how, why, when, and where particular microbes successfully compete for plant-derived nutrients. Therefore, the ability to obtain high-quality metaproteomic datasets from agar motivated us to characterize the temporal and spatial assembly of the SynCom community along the *P. trichocarpa* plant root system.

Prior to studying the assembly and function of the SynCom along the plant root axis, we employed the newly developed extraction method to assess community composition changes between 0-, 24-, and 48-h time points in the absence of the plant (control) when allowed to grow in a modified growth media (discussed below). To support early microbial growth and motility, the Hoagland's growth media commonly used for growing plants was supplemented with 0.1% sucrose and 0.3% agarose (**Figure 4.1A**). On average, the number of identified proteins increased across the time points (**Figure 4.3B**), and this is likely a result of increases in available biomass as well as changes in the metaproteome expression profiles. Importantly, the estimated starting SynCom composition was statistically unchanged when extracted from the inoculum or using the newly developed agar extraction method ($P = 1$) (**Figure 4.4**). Based on a permutational multivariate analysis of variance (PERMANOVA) ($P = 0.004$), the overall community composition changed significantly in control across the 48-h period (**Figure 4.5**). Individually, nine members exhibited a significant change in their relative abundance between the time points (**Table 4.2**). In this particular environment (i.e., media sans *P. trichocarpa*), *Pantoea* sp. strain YR343 increased in relative abundance to become a dominant member of the SynCom at the 24-h time point. However, this was no longer apparent at the 48-h time point, whereby *Caulobacter* sp. strain AP07 and *Rhizobium* sp. strain CF142 became the two most dominant members.

The same starting SynCom solution was then applied to a 3-week-old *P. trichocarpa* root system and allowed to grow for 48 h. To assess the influence of the plant root system on SynCom assembly, agar plugs were excised to obtain three spatial measurements: root tip, root base, and a region distant from the plant root system. Each discrete location was characterized at 24- and 48-h time points (Supplementary Table S1, refer online version). Similar to the no-plant SynCom

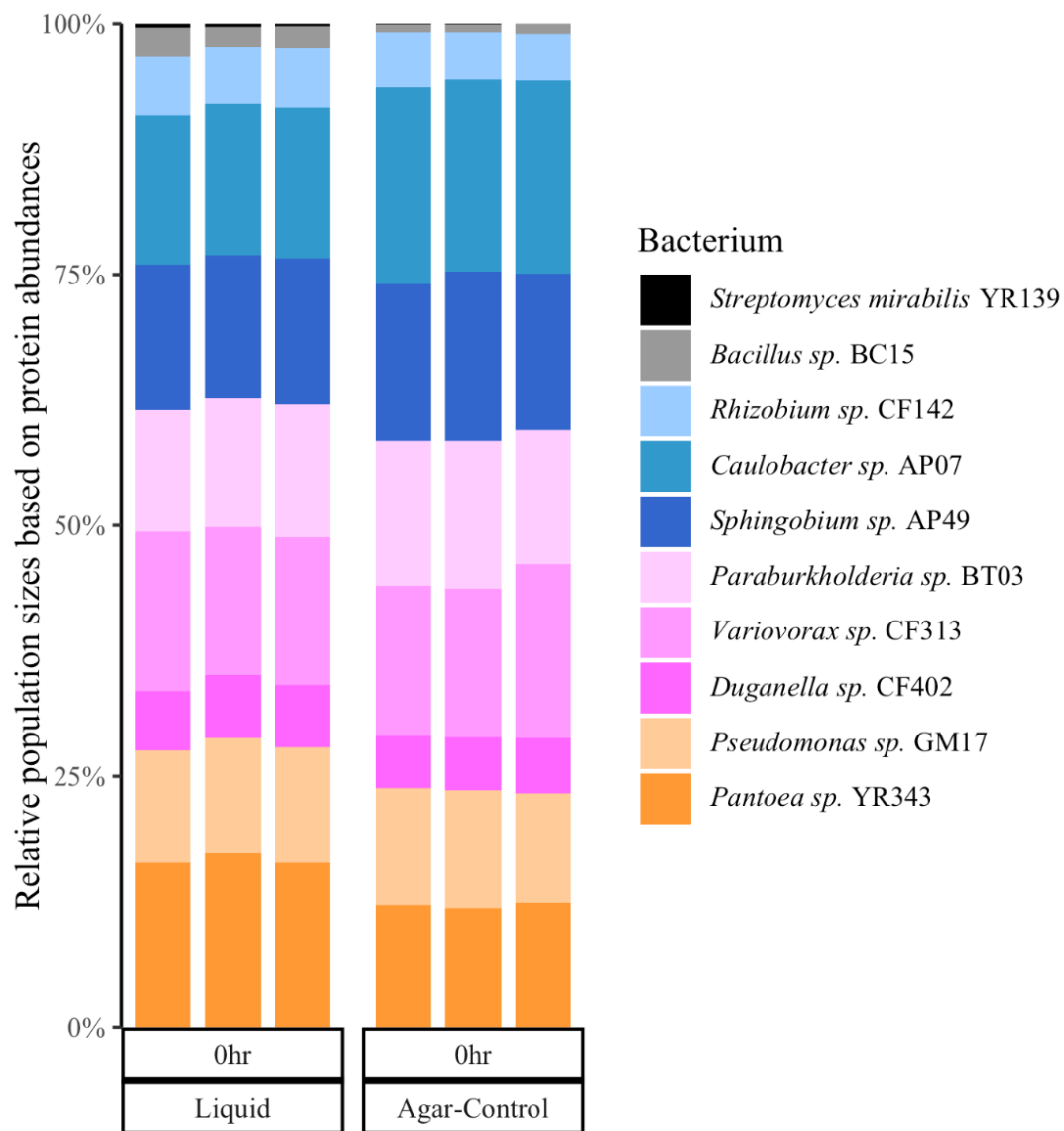


Figure 4.4: Assessment of microbial relative population sizes based on their summed protein abundances. Relative population sizes of microbes in control liquid stock 0-h and control plug 0-h, n=3.

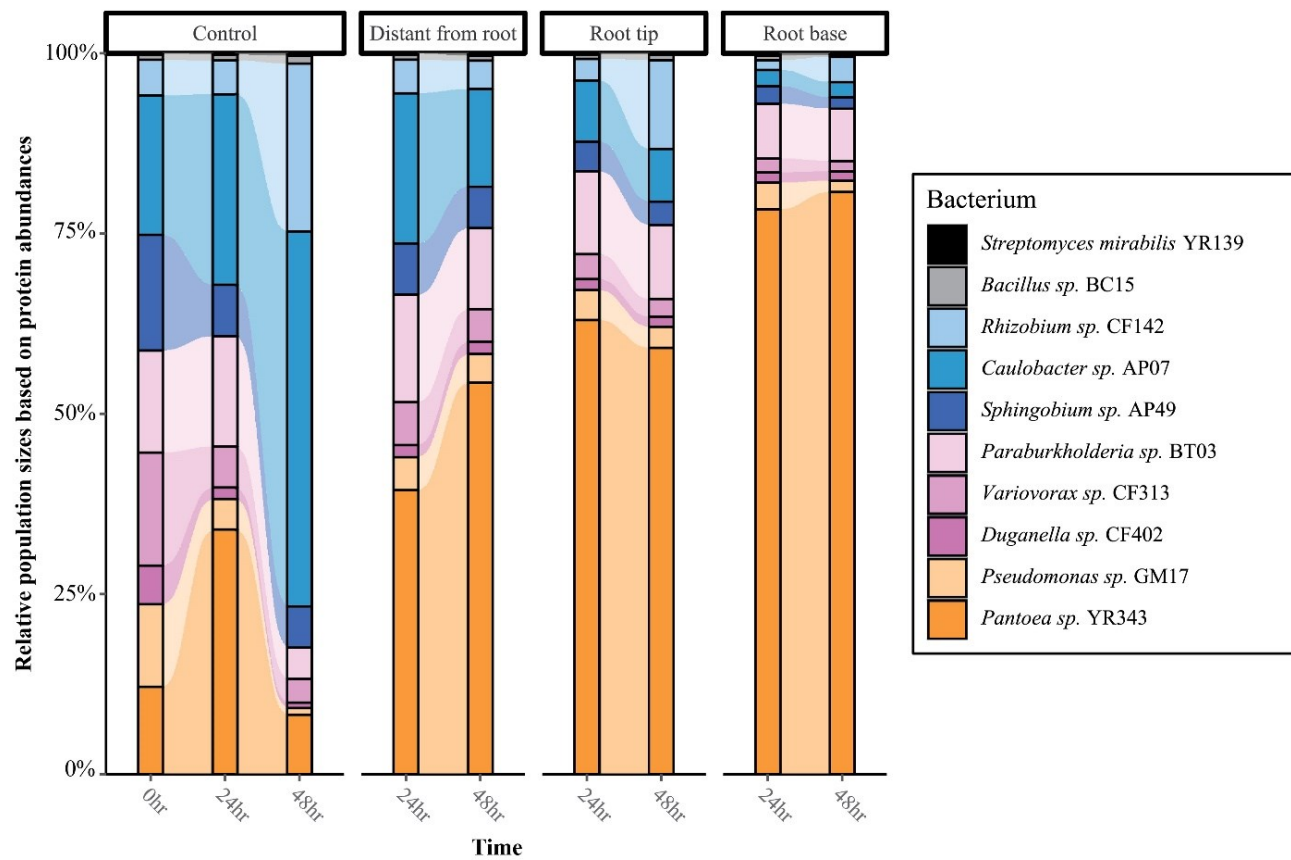


Figure 4.5: Assessment of microbial relative population sizes based on their summed protein abundances. Averaged organism relative abundances ($n = 3$) are illustrated for identified microbes in control plugs at 0, 24, and 48 h and three discrete locations surrounding the plant root system (root tip and base and a region distant from the root) at 24- and 48-h time points.

Table 4.2: ANOVA results for each member in the 10-member SynCom based on the relative abundances over time.

Organism	DF	p-value
<i>Rhizobium sp.</i> CF142	2	0.000865
<i>Bacillus sp.</i> BC15	2	0.07
<i>Pantoea sp.</i> YR343	2	0.00129
<i>Pseudomonas sp.</i> GM17	2	4.91E-08
<i>Caulobacter sp.</i> AP07	2	3.33E-05
<i>Paraburkholderia sp.</i> BT03	2	9.22E-07
<i>Duganella sp.</i> CF402	2	2.44E-08
<i>Streptomyces mirabilis</i> YR139	2	0.0017
<i>Sphingobium sp.</i> AP49	2	4.32E-06
<i>Variovorax sp.</i> CF313	2	4.67E-06

measurements, the average number of identified proteins generally increased from the 24- to 48-h time point and proteins identified per sample group ranged from 2,500 to 3,500 across the sample groups (**Figure 4.3C**). Overall, biological reproducibility between agar was high (average Pearson correlation coefficients of approximately 0.89).

To assess the effect of plant, spatial location, and time on the SynCom, PERMANOVA was performed on Euclidean distances using the metaproteomics data collected. Overall, the plant root system had the greatest influence on the measured microbiomes (50%), followed by the interaction between plant and time (24%), and time (14%) (**Figure 4.6A; Table 4.3**).

Intriguingly, there was a notable difference between microbiome compositions in the control agar plug at the 48-h time point and the region distant from the root at the 48-h time point. Both sample groups were expected to have similarly assembled communities at the 48-h time point because the influence of plant nutrients was absent or presumably lacking. Given the observed differences in microbiome composition, this result suggests that the plant root's zone of influence reaches significantly outward to the edges (approximately 3 to 4 cm from nearest root) of the plant-agar culture plates. Here, the root-associated microbiome composition largely favored the abundance of *Pantoea* sp. strain YR343, with its relative abundances increasing in the SynCom along the root system and the greatest abundance being observed at the root base. Additionally, four individual members of the SynCom had relative abundances that were significantly changing across the plant root axis (**Table 4.4**). For example, *Rhizobium* sp. strain CF142 became the second most dominant member at the root tip at the 48-h time point (**Figure 4.5**).

Detecting functional changes in microbial activity along the plant root axis

Beyond estimating microbiome compositions, metaproteomics can improve our understanding of how a distinct subset of SynCom members was able to grow and thrive along the plant root system in this particular environment^{54, 231}. To this end, we interrogated the proteome data of *Pantoea* sp. strain YR343, the most dominant plant-root-associated microbe, to assess its functional profile across the 24- and 48-h time points with and without the plant root system. Altogether, this analysis included 1,803 proteins representing approximately 37% of the encoded *Pantoea* sp. strain YR343 reference proteome. Importantly, this amount of proteome

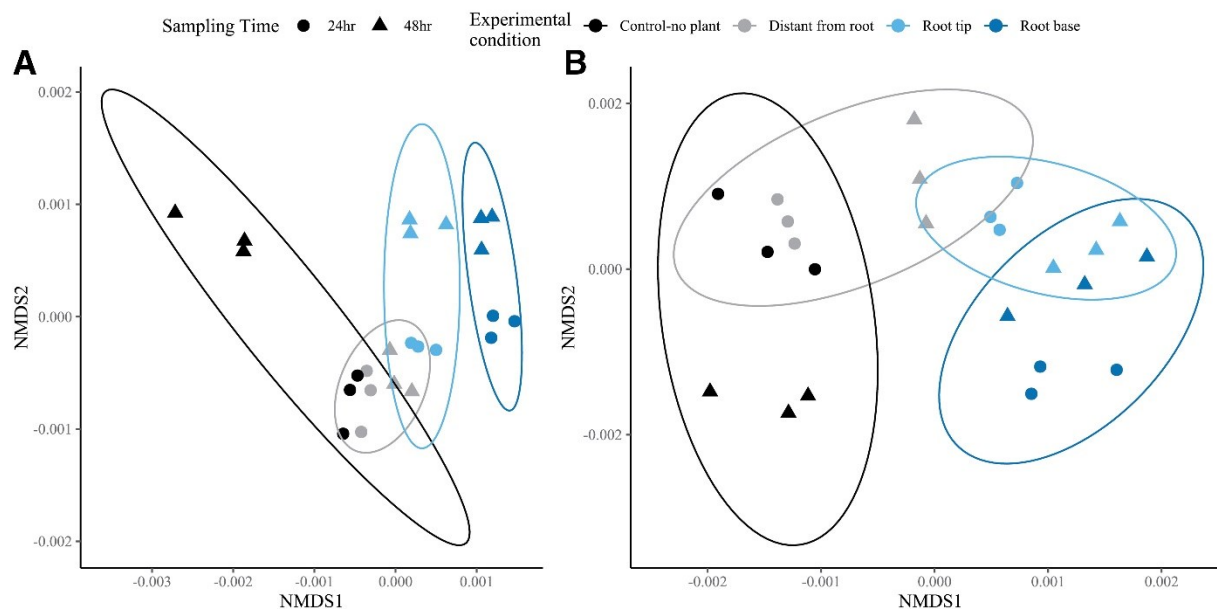


Figure 4.6: Association with plant shows higher variation in proteome expression of synthetic community. (A) Nonmetric multidimensional scaling (NMDS) plot based on Euclidean distances of 10-member proteomics data showed higher variation is observed in relation to association with the plant (permutational multivariate analysis of variance [PERMANOVA], $R^2 = 50.2\%$) followed by the interaction between the relation to plant and time (24%) and then time (13.9%). (B) NMDS plot based on Euclidean distances of *Pantoea* sp. strain YR343 proteomics data showed that higher variation is observed in relation to association with plant (PERMANOVA, $R^2 = 65.7\%$) followed by the interaction between the relation to plant and time (17.1%) and then time (8.4%).

Table 4.3: PERMANOVA results of Euclidean distances of proteins. (A) From all 10-member SynCom and (B) *Pantoea* sp. YR343 using the spatial location along the plant and time of sampling.

A.

ADONIS 10-member SynCom results			
Variable	Df	R2	P
Spatial location	3	0.502	0.001
Time	1	0.139	0.001
Spatial location: Time	3	0.24	0.001

B.

ADONIS <i>Pantoea</i> sp. YR343 results			
Variable	Df	R2	P
Spatial location	3	0.657	0.001
Time	1	0.084	0.001
Spatial location: Time	3	0.171	0.001

Table 4.4: ANOVA results for each member in the 10-member SynCom based on the relative abundances at the 48h time across spatial locations.

Organism	DF	p-value
<i>Rhizobium sp.</i> CF142	2	0.139
<i>Bacillus sp.</i> BC15	2	0.0517
<i>Pantoea sp.</i> YR343	2	0.0699
<i>Pseudomonas sp.</i> GM17	2	8.72E-04
<i>Caulobacter sp.</i> AP07	2	0.0513
<i>Paraburkholderia sp.</i> BT03	2	2.09E-01
<i>Duganella sp.</i> CF402	2	1.85E-01
<i>Streptomyces mirabilis</i> YR139	2	0.027
<i>Sphingobium sp.</i> AP49	2	1.45E-02
<i>Variovorax sp.</i> CF313	2	1.50E-03

coverage is similar to what was achieved in prior proteomic studies of this individual bacteria^{232, 233}.

Factors influencing microbiome composition were similar to those affecting proteome expression of *Pantoea* sp. strain YR343. PERMANOVA results indicate that the plant root system has a major influence (66%) on proteome expression, followed by the interaction of plant and time (17%) and time (8%) (**Figure 4.6B; Table 4.3**). To identify the differentially abundant proteins in *Pantoea* spp. across the measured proteomes, an ANOVA test was performed; 1,720 proteins were determined to have a significant difference in relative protein abundance across the sample groups (P value ≤ 0.05). One-way hierarchical clustering was performed providing a comprehensive picture of protein regulation across the experimental conditions (**Figure 4.7**). In total, four major clusters were observed. The cluster with the most membership (cluster 3; 566 proteins) consists of proteins with increased relative abundance in all plant root extractions when compared with the control samples at the 24- and 48-h time points. In other words, cluster 3 represents proteins upregulated in relative abundance in the presence of the 3-week-old *P. trichocarpa* root system. A gene ontology (GO) enrichment analysis on the proteins in cluster 3 revealed 36 overrepresented GO terms for biological processes and molecular functions (Supplementary Table S6, refer online version). These GO terms included but were not limited to ion binding, cell motility, transport, response to stress, membrane organization, generation of precursor metabolites, and energy. At a high functional level, these molecular processes explain how *Pantoea* sp. strain YR343 responds to the plant root chemical environment where it competes for available resources through overexpression of several motility-related proteins such as methyl-accepting chemotaxis proteins, response regulator proteins with CheY-like receiver domains, chemotaxis signal transduction proteins, and reprogramming its growth and metabolic activities.

Conclusions

Functional genomic measurements for plant-root-associated microbiomes are often challenged by the performance of available extraction methods¹⁸⁵. In nature, plant root systems are surrounded by complex soil chemistries and physical properties that will vary by biome (e.g.,

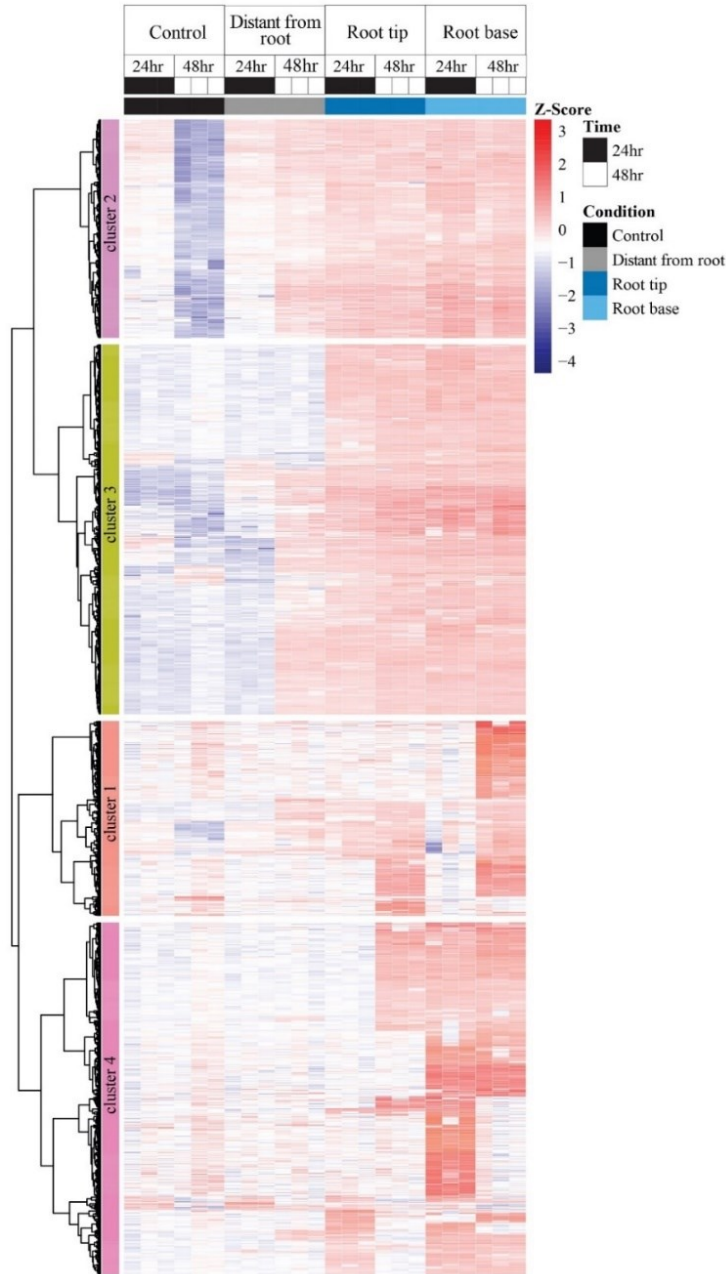


Figure 4.7: Hierarchical clustering results for *Pantoea* sp. strain YR343. One-way hierarchical clustering (rows = Euclidean distances) of Z-scores of log₂-transformed protein abundances for significantly regulated proteins (multiple-sample test analysis of variance was performed in Perseus software, permutation-based false discovery rate < 0.05; number of randomizations = 250) across experimental conditions for *Pantoea* sp. strain YR343 revealed four major clusters of proteins.

related to how microbiome communities and functions are altered by plant root rhizodeposition at certain points in space or time.

In order to address knowledge gaps precluded by too much complexity, experimental representation of biological systems is instead refined to key parameters and their interplay. For plant-microbe interactions, these reductionist approaches are performed using a variety of artificial components^{234, 235} or engineered laboratory systems and permit us to answer important questions currently unobtainable in natural ecosystems^{37, 236}. For example, the widely used plant-agar system offers a well-controlled platform, with low-to-medium complexity, that would permit us to answer several important questions. How do microbiome communities assemble in the presence of a plant root system? How does composition and function change along the plant root system at certain points in time and space? Does assembly and maintenance of root-associated microbiomes observed with a specific plant genotype remain the same for other host genotypes? Unfortunately, a method to assess spatiotemporal plant-microbe interactions along the root system in plant-agar systems at a molecular level is yet to be realized.

In order to address this technical limitation, we developed a novel peptide extraction method for high-resolution metaproteome characterization of plant-agar systems. After inoculating a 3-week-old *Populus* plant root system with a 10-member SynCom and allowing growth for a 48-h period, we compared the performance (number of proteins identified) and resulting estimates of organism relative abundances between a root wash method and the newly developed agar extraction method. When comparing the two methods, we observed statistically unchanged estimates of microbiome composition and deeper metaproteome measurements (approximately 53% gain in protein identifications) for the agar extraction method, supporting the utility of the agar extraction method for whole-root characterization of microbiomes. More importantly, we demonstrated that this method affords new capabilities in sampling the spatial composition and function of root-associated microbiomes along the plant root axis.

For this particular plant-microbe biological system, substantial changes in composition and function were observed from the root base to the root tip, and even at a region distant from the plant root system. These findings have significant implications in our understanding of how microbiomes respond to plant rhizodeposition and how nutrient availability can affect the nature

of microbiome assembly in different niche environments, particularly those associated with *Populus* spp. In fact, this study was designed to complement a previous work that used the same 10-member SynCom but assessed microbiome assembly in a liquid culture environment with rich or minimal media¹⁹². The high-level observation that the same 10-member SynCom forms different assemblies across varying methodological frameworks is quite intriguing because we can now work to identify rules of assembly, and understanding these rules will give us greater insight into predicting how microbiomes assemble in different environments. More broadly, we anticipate that this novel extraction approach will benefit other soft-agar plate studies that study nutrient exchange among microbes²³⁷.

Quite surprisingly, we recognized that the 10-member SynCom community assembled in this study is nearly identical in composition to that observed in a previously studied clay soil matrix that was measured using 16S ribosomal RNA gene amplicon sequencing¹⁶. The major difference between the two experiments is that it took 21 or 2 days to form nearly identical communities in potted soil or plant-agar, respectively. We posit that environmental matrix properties such as hydration levels, which affect the diffusion of plant rhizodeposits and microbial mobility, resulted in these temporal differences. Although we anticipate that community assembly will become more distinct between the plant-agar system and more natural soils, the relatedness with clay soil matrix addresses concerns about the plant-agar system being too artificial.

CHAPTER 5: UNRAVELING HOW BACTERIAL ADAPTATION STRATEGIES IN THE MICROBIOME ARE SHAPED BY THE CHEMICAL ENVIRONMENT OF THE PLANT RHIZOSPHERE

Author contributions

Text and figures were adapted from the following published journal article:

Appidi. Manasa, Bible. Amber, Abraham. Paul, Jawdy. Sara, Hettich. Robert & Morrell-Falvey. Jennifer. (2024). Unraveling bacterial adaptation strategies in the microbiome shaped by the chemical environment of the plant rhizosphere. Phytobiomes Journal. 10.1094/PBIOMES-08-24-0081-R.

Manasa R. Appidi's contributions include proteomic sample preparation, data collection and data analysis, writing and revisions of the manuscript.

Note: All the supplemental material including supplementary figures and tables mentioned in text are available at <https://doi.org/10.1094/PBIOMES-08-24-0081-R> and their numbering reflects that of the online material.

Introduction

The rhizosphere represents a dynamic and complex environment where chemical and physical factors play crucial roles in shaping microbial communities^{238, 239}. Plant roots release a variety of compounds, such as root exudates, which serve as a rich nutrient pool for microbial populations and signals for plant-microbe interactions^{240, 241}. These exudates, comprised of sugars, organic acids, amino acids, and other compounds, create a unique chemical milieu that attracts and sustains a diverse array of microorganisms^{212, 241, 242}. Simultaneously, the physical structure of the rhizosphere, influenced by root morphology and soil composition, further contributes to the niche-specific conditions^{243, 244}. The intricate interplay of chemical and physical parameters in the rhizosphere dictates the composition and functioning of associated microbial communities, impacting plant health and nutrient cycling²⁴⁵. Bacteria that are beneficial to the plant are collectively referred to as plant growth-promoting rhizobacteria (or

PGPR), and they influence plant growth and health by exerting a range of functions such as phosphate solubilization, nitrogen fixation, enhancing water availability, suppression of pathogenic activity, and synthesis of phytohormones²⁴⁶⁻²⁴⁸. The intricate interactions between plants and PGPR within the rhizosphere underscore the complexity of functional dynamics in this critical zone and emphasize the pivotal role these bacteria play in fostering and modulating plant health. Beyond their direct interactions with the plant host, plant-associated bacteria interact with other soil organisms, competing for space and resources in the complex soil ecosystem²⁴⁹. Comparative genomic analyses of diverse bacteria have uncovered specific genes that are enriched in plant-associated bacteria compared to their non-plant-associated counterparts, particularly in genes related to carbohydrate metabolism²²⁰. Similarly, meta-transcriptomic investigations of microbial communities in the rhizosphere have unveiled specific pathways and gene products crucial for survival in this dynamic environment^{250, 251}. Furthermore, the characterization of mutants in genetically tractable organisms has identified additional bacterial gene products contributing to plant association^{252, 253}. Despite significant strides in unraveling the drivers of plant association, there remain considerable knowledge gaps regarding how environmental conditions in the rhizosphere, along with interactions among diverse organisms, shape the composition and functionality of the plant microbiome. To this end, metaproteomics can provide important complementary information about the detailed metabolic activities of microbial communities, thereby helping to expand our understanding of these dynamics and provide insights into the complex interplay between plant-associated bacteria and the surrounding environment.

To gain a deeper understanding into the chemical and physical cues that drive plant-microbe interactions, researchers often employ synthetic microbial communities consisting of defined mixtures of microorganisms^{39, 254, 255}. These model systems help delineate the functions of specific microbial taxa in plant-microbe interactions, not only on an individual basis, but also within the larger context of the microbial community. Key mechanisms underlying the establishment of symbiotic relationships, nutrient cycling, and the overall resilience of plant-associated microbial ecosystems can be uncovered by manipulating synthetic communities in controlled experiments. This approach provides valuable insights into the fundamental principles governing plant-microbe interactions. In our previous studies, a metaproteomic analysis of a

synthetic microbial community consisting of ten bacterial strains isolated from *Populus deltoides* roots grown under varying nutrient conditions revealed a number of functions and adaptation strategies employed by the most abundant bacterial species in the community in response to interspecies interactions and nutrient competition in a simplified liquid media^{192, 193}. The bacterial strains in the ten-member constructed community were selected to be representative of the diversity found within the natural poplar microbiome^{192, 193, 256}. In the current study, we aimed to better understand how these ten bacterial species alter their proteome in the presence of a plant host which influences microbial behavior through chemical cues and physical interactions. Historically, these types of studies present a unique set of challenges, including limited bacterial biomass, inherent difficulty in separating bacteria from plant root tissue, and extraction of plant proteins that may influence measurements of the bacterial proteome²⁵⁷. Recent strides in methodologies, particularly those employing a plant in an agar system, have facilitated the extraction of microbial proteomes from the agar surrounding plant roots and provided insights into the spatial dynamics of microbial recruitment and community composition²⁵⁸. Despite these advances, there is a need for novel approaches that can effectively surmount the existing challenges associated with studying microbial functions that are most likely influenced by the chemical environment of the plant.

In this study, we employ a unique approach for analyzing bacterial responses to the plant rhizosphere chemical environment by applying bacteria to a filter that is in contact with the plant roots. The use of a filter allows for chemical exchange between the bacteria and the plant roots but prevents physical interactions between them, thereby facilitating efficient bacterial collection for proteomics analyses. While this reductionistic approach was designed as a practical solution to studying how bacteria respond to the chemical environment of a plant, we acknowledge the constraints associated with this experimental design, namely that the bacteria are not in direct contact with the plant root, which precludes measurement of contact-dependent responses and that the quantity and composition of the applied bacterial samples may not reflect the spatial and temporal organization of the natural community. For these reasons, we take care to qualify our results in the framework of our experimental approach and avoid over-reach about the more complicated *in vivo* system. Using this approach, we investigated the proteome profiles of ten bacterial strains originally isolated from the *Populus* rhizosphere. Our experimental design

involved assessing the bacterial proteomes when grown individually and in a synthetic community in both the presence and absence of a plant. Our findings demonstrate discernible differences in the behavior of individual bacterial strains in response to the chemical environment of a plant host and within a plant-associated bacterial community. This study contributes to our understanding of the dynamic interplay between bacteria and their surroundings and provides valuable insights into bacterial functional adaptations in response to the plant host and other species within the microbial community.

Materials and methods

Preparation of plants, bacterial cultures, and growth conditions

Sterile cuttings of *Populus trichocarpa* genotype BESC819 plants were generated from shoot tips of greenhouse-grown plants and grown in Murashige-Skoog (MS) media for three weeks for rooting in plant growth chambers as described previously²⁵³. Bacterial strains used in this study are listed in **Table 5.1**. All strains were grown in 100 mL liquid R2A media (Teknova) with shaking at 28°C. Most strains were grown overnight, but slower growing strains, including *Rhizobium* sp. CF142 and *Streptomyces* sp. YR139, were grown for two days. The *Pantoea* sp. YR343 transposon mutant MCP2190::Tn5 was described previously²⁵².

Expression of plant-responsive promoters on Populus roots using filter-based system

Pantoea sp. YR343 (pPROBE-DGC2884) was grown overnight in R2A media with shaking at 28°C²⁵². Cells were collected by centrifugation at 5,000 rpm for 5 minutes (Eppendorf 5424 microcentrifuge), then washed and resuspended in PBS at a final optical density of 2.11 (at 600 nm). Next, we placed 3-week old rooted *Populus* cuttings on the surface of agar plates consisting of 0.25 X Hoagland's modified basal salt mixture (PhytoTech Labs) with 0.1% sucrose and 1.5% agar. A sterilized 0.45 µm nylon filter (Millipore) was placed along the length of the plant root or on top of the media as a control. Washed cells were then spotted as 5 µL drops on the sterilized filter strips and incubated in a plant growth chamber at 24°C with a 12-hour light and 12-hour dark photoperiod. After 24 hours, cells were scraped from filters and

Table 5.1: Bacterial strains utilized in this study

Bacterial strain	Phylogeny
<i>Streptomyces mirabilis</i> YR139	<i>Actinobacteria</i>
<i>Bacillus</i> sp. BC15	<i>Firmicutes</i>
<i>Caulobacter</i> sp. AP07	α - <i>Proteobacteria</i>
<i>Rhizobium</i> sp. CF142	α - <i>Proteobacteria</i>
<i>Sphingobium</i> sp. AP49	α - <i>Proteobacteria</i>
<i>Duganella</i> sp. CF402	β - <i>Proteobacteria</i>
<i>Paraburkholderia</i> sp. BT03	β - <i>Proteobacteria</i>
<i>Variovorax</i> sp. CF313	β - <i>Proteobacteria</i>
<i>Pantoea</i> sp. YR343	γ - <i>Proteobacteria</i>
<i>Pseudomonas</i> sp. GM17	γ - <i>Proteobacteria</i>
Engineered Strains	Reference
YR343 (pPROBE-DGC2884)	252
MCP2190::Tn5 mutant (mCherry)	252
<i>Pantoea</i> sp. YR343:GFP	253

imaged for expression of GFP (green fluorescent protein) using a Zeiss LSM710 confocal microscope equipped with a 63X objective.

Bacterial inoculation on Populus roots using a filter-based system

We used 0.25 X Hoagland's modified basal salt mixture (PhytoTech Labs) with 0.1% sucrose and 1.5% agar in standard 100 mm diameter petri dishes for studying bacterial responses to plant root chemicals. Fresh cultures of bacterial cells were collected by centrifugation at 5,000 rpm for 5 minutes in a F13-14x50cy rotor. Cells were washed in sterile PBS, then adjusted to a volume of 10 mL at an optical density at 600 nm (OD₆₀₀) of 6.0. Next, we added individual plants to petri dishes filled with the plant media. On top of each plant root, we added a sterilized 0.45 µm nylon filter (Millipore) and took care to ensure the membrane contacted roots only, not stems or leaves. We then added 1 mL of cell culture to the filter. For control experiments that had no plants, we placed a sterilized 0.45 µm nylon filter on the plant media directly and then added 1 mL of cell culture. For the 10-member community, all ten strains were mixed in equal ratios and inoculated onto nylon filters both in the presence and absence (control) of a plant root as described above for individual strain inoculation. In order to avoid damaging the plants, lids were removed from the petri dishes for the duration of the experiment and replaced with a sterilized 8-ounce *PhytoCon* clear plastic container (PhytoTech labs) to maintain adequate humidity and a sterile environment. All containers were placed into a reach-in plant chamber and incubated at 24°C with a 12-hour light and 12-hour dark photoperiod for 48 hours prior to harvesting the bacteria. Bacterial cells were harvested for proteomics by rinsing filters with PBS and collecting cell pellets in microcentrifuge tubes. Cell densities were measured to allow for standardization during data analysis.

Cellular protein extraction for proteomic analyses

Cell pellets were resuspended in 300 µL of lysis buffer (4% sodium dodecyl sulfate (SDS) in 100 mM Tris, pH 8.0) by vortexing. Samples were heated at 90°C in a heat-block for 10 minutes followed by sonication for 2 mins (30% amplitude, 10-s pulse with 10-s rest, 2-min total pulse time) (Branson digital sonifier) and heated for additional 10 minutes at 90°C. Samples were centrifuged at maximum speed for 10 minutes to remove the cell debris and the cell lysates were

collected into a fresh tube. Proteins were quantified using a Nanodrop One spectrophotometer (Thermo Scientific). Samples were reduced with 10mM dithiothreitol (DTT) and incubated at 90°C for 10 minutes. Samples were then alkylated by incubating with 30 mM iodoacetamide (IAA) for 15 minutes in the dark to prevent the reformation of disulfide bonds. Proteins were agglutinated on Sera-Mag beads using Protein Agglutination Capture method as described previously¹²⁸. Proteins were digested with two separate and sequential aliquots of sequencing grade trypsin (Promega) at 1:75 (wt/wt) protein:trypsin ratio. Proteins were first digested overnight, followed by 3 hours. The digestions were done at 37°C at constant shaking. The samples were acidified with formic acid to 0.5% formic acid concentration and vortexed. Vivaspin 500 columns (10 kDa MW cutoff filter) were equilibrated with 500µL of 100mM tris, pH 8.0 and centrifuged at 12,000 x g for 15 minutes. Tryptic peptides were added to the equilibrated Vivaspin columns and centrifuged at 12,000 x g for 15 minutes and the flow-through was collected. Care was taken to not aspirate Sera-Mag beads. Peptides were quantified using a Nanodrop One spectrophotometer (Thermo Scientific) and transferred to the auto-sampler vials for LC-MS/MS measurement.

Protein identification and quantification

All peptides were analyzed using two-dimensional (2D) liquid chromatography (LC) on an Ultimate 3000 RSLCnano system (Thermo Fisher Scientific, USA) coupled with a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific, USA). For each sample, a 2 µg inject of digested peptide mixture was flowed across an in-house built strong cation exchange (SCX) Luna trap column (5µm, 150µm X 50mm; Phenomenex, USA) followed by a nanoEase symmetry reversed phase (RP) C18 trap column (5µm, 300µm X 50mm; Waters, USA) an aqueous solvent (5% ACN, 0.1% FA) and then wash for 10 min. The loaded peptide mixtures were separated and analyzed across three successive SCX fractions of increasing concentrations of ammonium acetate (35mM, 50mM, and 1M), each followed by a 100-minute linear organic gradient (250 nL/min flow rate) from 5% aqueous solvent (5% ACN, 0.1% FA) to 30% organic solvent (80% ACN, 0.1% FA) to separate peptides across an in-house pulled nanospray emitter (75 µm X 350 mm) packed with C18 Kinetex RP C18 resin (1.7 µm; Phenomenex, USA). All MS data were acquired with Thermo Xcalibur (version 4.2.47) using the top N 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 was 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 50 ms maximum injection time. 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 2.5, Thermo Fisher Scientific, USA)²²⁵. Each MS raw data file was processed by the SEQUEST HT database search algorithm¹⁴⁰ and confidence in peptide-to-spectrum (PSM) matching was evaluated by Percolator¹⁹⁷. Peptide 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. Organism relative abundances were calculated by dividing their summed protein abundances from the sum total of all quantified protein abundances¹⁸⁹.

Statistical analysis

Proteins with at least one unique peptide were exported from Proteome Discoverer. Log2-transformation of protein abundances was performed followed by LOESS normalization and mean-centering across the entire dataset using InfernoRDN software v1.1.7793¹⁴⁹. For the community samples, protein abundances subset for each bacterial species were extracted from the normalized dataset and further mean-centered using InfernoRDN. To improve the robustness of downstream analysis, the analysis was limited to proteins that were identified in at least two out of three replicates of at least one sample group. 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.1.6.12.0¹⁵⁰. To identify the proteins that are differentially abundant in the presence of a plant compared to the control for each bacterial species, a student's t-test was performed and proteins that pass the significance threshold of $p\text{-value} \leq 0.05$ and absolute Log2-fold change greater than 1 are characterized as significantly differentially abundant.

Root colonization assays

Colonization experiments were performed as described previously²⁵³. Briefly, wild type *Pantoea* sp. YR343 and the MCP2190::Tn5 mutant (affecting PMI39_02190) were grown overnight in R2A media as described above. We treated 3-week-old *P. trichocarpa* BESC819 plants that were transplanted into a clay soil supplemented with Hoagland's nutrient solution with approximately 10^7 cells per plant. Plants and bacteria were grown for three weeks as described above, then harvested for analysis of colony forming units and imaging of root colonization. To facilitate imaging, we used a GFP-labeled wild type strain and a mCherry-labeled MCP2190::Tn5 strain that was generated previously²⁵². Imaging was performed using a Zeiss LSM710 confocal microscope.

Results

A filter-based system enables the study of bacterial responses to the plant root chemical environment without physical contact with the plant

The ability to examine how plant root association influences bacterial physiology has been hampered by the difficulty of separating root-associated bacterial biomass from plant material to allow for robust proteomic analyses. As a compromise, we reasoned that bacterial responses to the plant host could be examined using a filter-based system that allows chemical exchange between the plant root and bacterial inoculum but maintains physical separation to facilitate bacterial sample collection (**Figure 5.1**). Control experiments were performed by growing bacteria on a filter in the absence of a plant (**Figure 5.1A**) and compared to bacteria on a filter placed over a plant root (**Figure 5.1B, 5.1C**). For these experiments, we utilized nylon filters with a 0.45 μm pore size due to the convenience of sterilization, overall durability, and common use in metabolomics studies²⁵⁹, suggesting minimal negative effects on chemical communication between the plants and bacteria. To test this experimental design, we utilized a previously characterized plant-responsive reporter strain of *Pantoea* sp. YR343 in which GFP expression is driven by the promoter of a gene encoding a diguanylate cyclase (PMI39_02884) that is active in and the presence of a plant host²⁵². A nylon filter was placed on the roots of *P. trichocarpa* BESC819 on an agar plate containing 0.25 x Hoagland's media supplemented with 0.1% sucrose

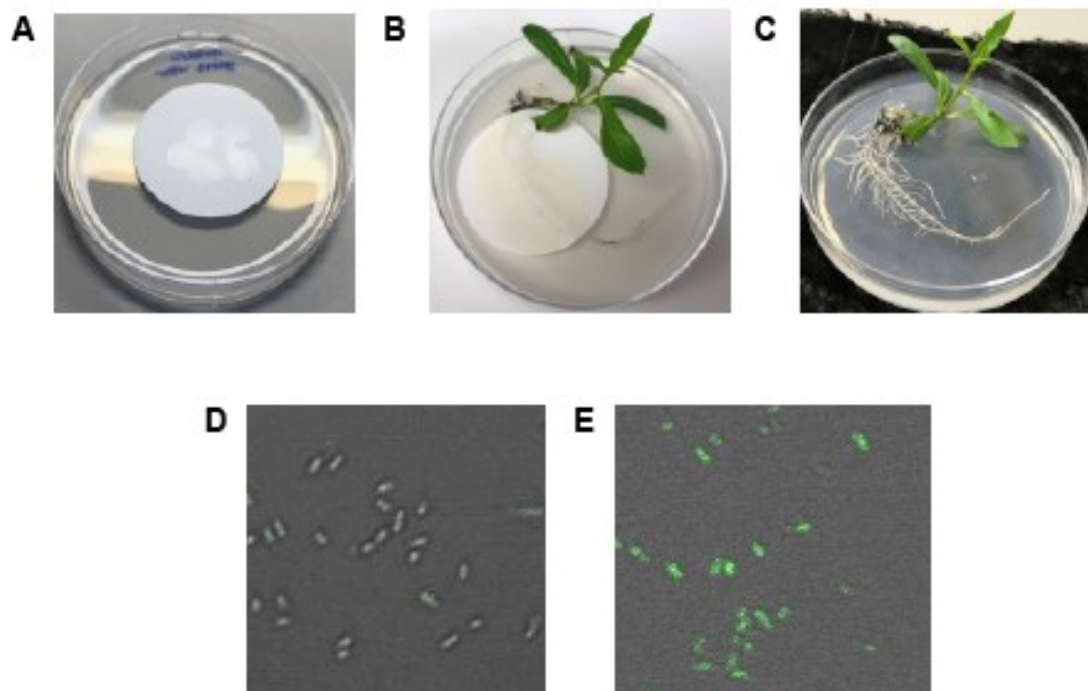


Figure 5.1: Overview of experimental design and proof-of-concept. The experimental design consisted of agar plates with nylon filters that were inoculated with bacteria and placed on agar plates as (A) a control or (B) on the roots of an axenic *Populus trichocarpa* BESC819 plant. (C) plant root after removal of the filter with bacteria. (D, E) GFP expression in *Pantoea* sp. YR343 (pPROBE-DGC2884) cells grown on filters in the (D) absence or (E) presence of a plant root. GFP expression is controlled by a plant responsive promoter that responds to chemical signals detected through the filter.

and *Pantoea* sp. YR343 reporter strain was applied to the filter and allowed to incubate for 24 hours. A control filter with the reporter strain was placed on a separate agar plate to examine the response in the absence of the plant. Comparison of GFP fluorescence levels from the *Pantoea* sp. YR343 reporter strain showed significantly higher fluorescence in the presence of the plant root compared to the control filter indicating that the promoter of gene PMI39_02884 can be activated under these experimental conditions (**Figure 5.1D, 5.1E**). This result is consistent with the assumption that bacteria can respond to the chemical environment of the plant through the filter.

Quantification of bacterial proteomes revealed varying responses to the presence of a plant by individual isolates

Based on the results above, we proceeded to investigate the impact of the plant-associated chemical environment on the responses of ten bacterial species previously isolated from *Populus deltoides* roots^{192, 193}. Each bacterial species was grown individually and inoculated onto nylon filters both in the presence and absence (control) of a plant root. Following incubation for 48 hours, the bacterial cells were collected from filters for proteomic analyses. In this global proteomic measurement, taxon-specific proteins were identified across the ten bacterial strains grown individually, with comparable measurement depths between strains grown in the absence or presence of plant roots (**Figure 5.2**, Supplemental Table 1, refer online version).

To identify the responses of these bacterial strains to the plant root environment, a differential proteomic analysis was performed for each strain. To assess changes in the abundances of individual proteins in the presence of plant compared to controls for each strain, a student's t-test was performed and proteins with significant differences ($p\text{-value} < 0.05$, and absolute $\log_2 \text{FC} > 1$) in protein abundances were identified (Supplemental Table 2, refer online version). In response to the plant root chemical environment, these bacterial strains displayed varying degrees of changes in their proteome regulation (**Figure 5.3**). For example, a large percentage of the *Bacillus* sp. BC15 proteome underwent significant changes (~54%); a similar trend was observed in *Duganella* sp. CF402 (~50%). However, the directionality of change is opposite in these two strains, with 87% of the differentially abundant proteins showing an increase in abundance in *Bacillus* sp. BC15, whereas 85% of the differentially abundant proteins showed a

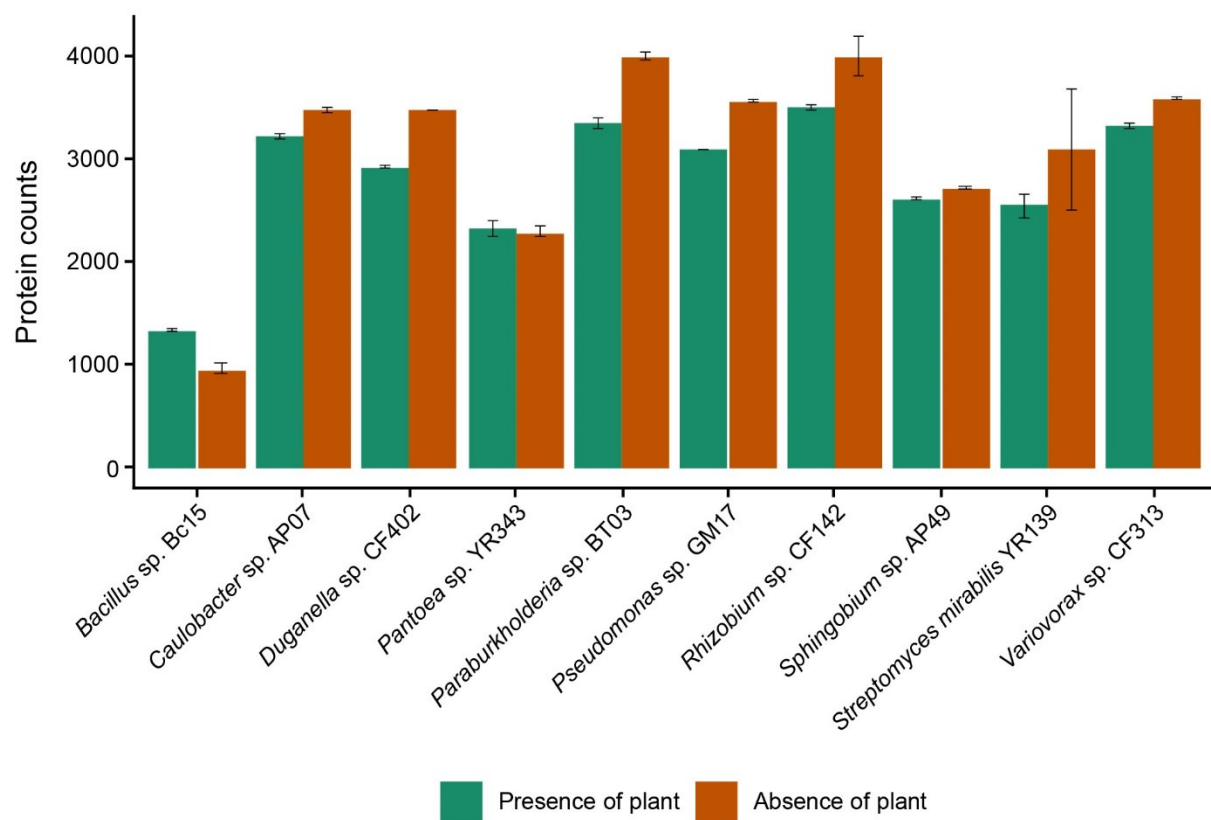


Figure 5.2: Average number of proteins identified (n=3) in proteomic analysis of ten bacterial strains grown individually in the absence or presence of the plant. Error bars represent standard deviation of mean.

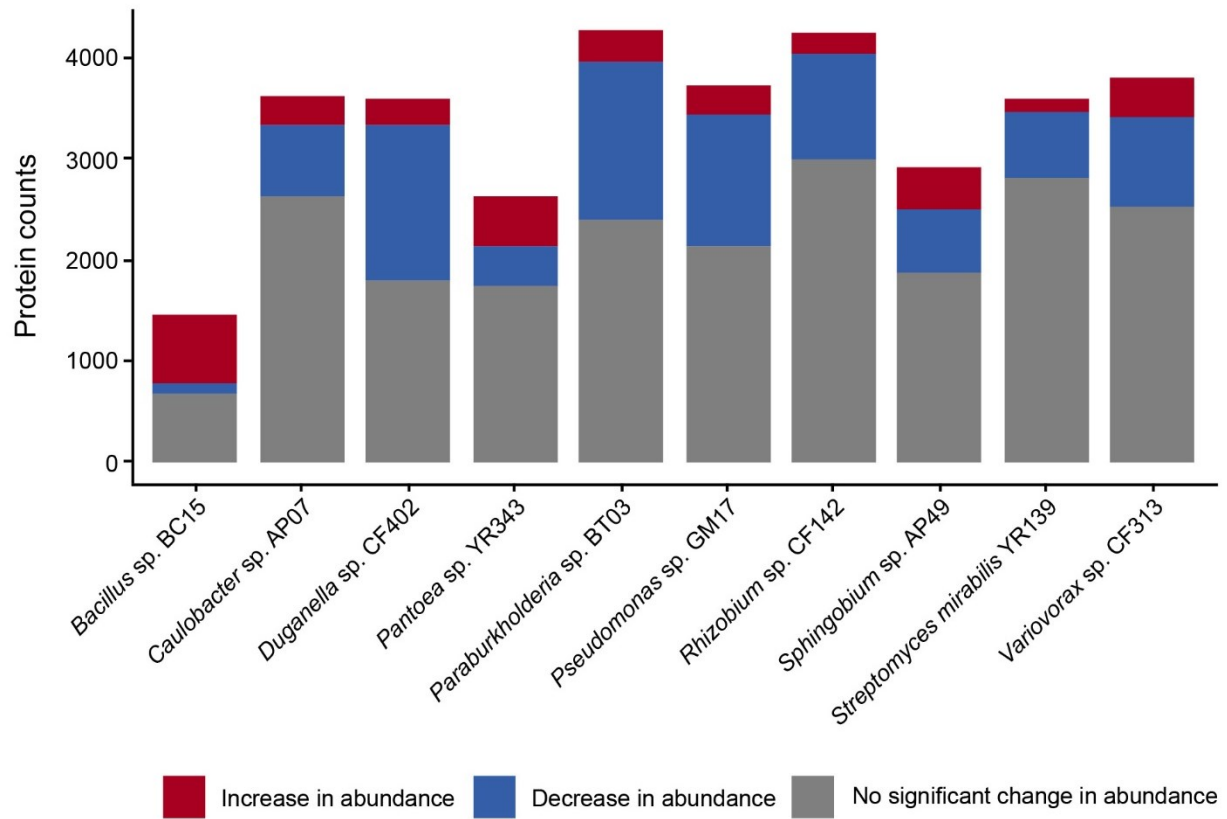


Figure 5.3: Stacked bar charts showing the number of differentially abundant proteins for ten bacterial strains grown individually in the presence of a plant compared to when grown in the absence of a plant.

decrease in abundance in *Duganella* sp. CF402 when grown in the presence of a plant. In *Pantoea* sp. YR343, although only 33% of its measured proteins displayed significant changes in the presence of a plant compared to the control, more proteins (56%) showed an increased abundance, similar to *Bacillus* sp. BC15. All other bacterial strains displayed a similar trend as *Duganella* sp. CF402, where the majority of the significantly changing proteins exhibited decreases in abundance in the presence of a plant. The shifts in protein abundances across these bacterial strains highlight strain specific patterns in response to the plant root environment under these experimental conditions.

The plant-associated chemical environment impacts the microbiome population structure

We then investigated the impact of the plant-associated chemical environment on proteomic changes of these ten bacterial strains when grown together in a mixed community. Based on previous studies of the 10-member community in plant root colonization²⁵⁶, we prepared a mixture of all ten strains in equal ratios that was then inoculated onto nylon filters both in the presence and absence (control) of a plant root. After 48 hours of incubation, the bacterial cells were collected from filters for metaproteomic analyses. Taxon-specific proteins were identified for the community samples from control experiments and when grown in the presence of plant roots (**Figure 5.4A, 5.4B** and Supplemental Table 3, refer online version). The depth of the measurements decreased somewhat for each strain compared to when grown individually yet remained comparable between the presence or absence of plants. Due to variations in measurement depths for each strain when grown individually and in a community setting, we deemed it was more informative to compare functional categories rather than protein inventories to discern plant-driven and community-driven responses in these bacterial strains.

First, to identify the impact of *Populus* root chemical environment on the microbiome population structure after 48 hours, the relative abundances of each bacterial strain in the community were assessed and compared between control and with plant samples using summed protein abundances (i.e., summed total relative protein abundance per organism divided by the summed total protein abundance of community) (**Figure 5.5**). To assess how the plant affects the microbial community, PERMANOVA was performed using the metaproteomics data. Overall,

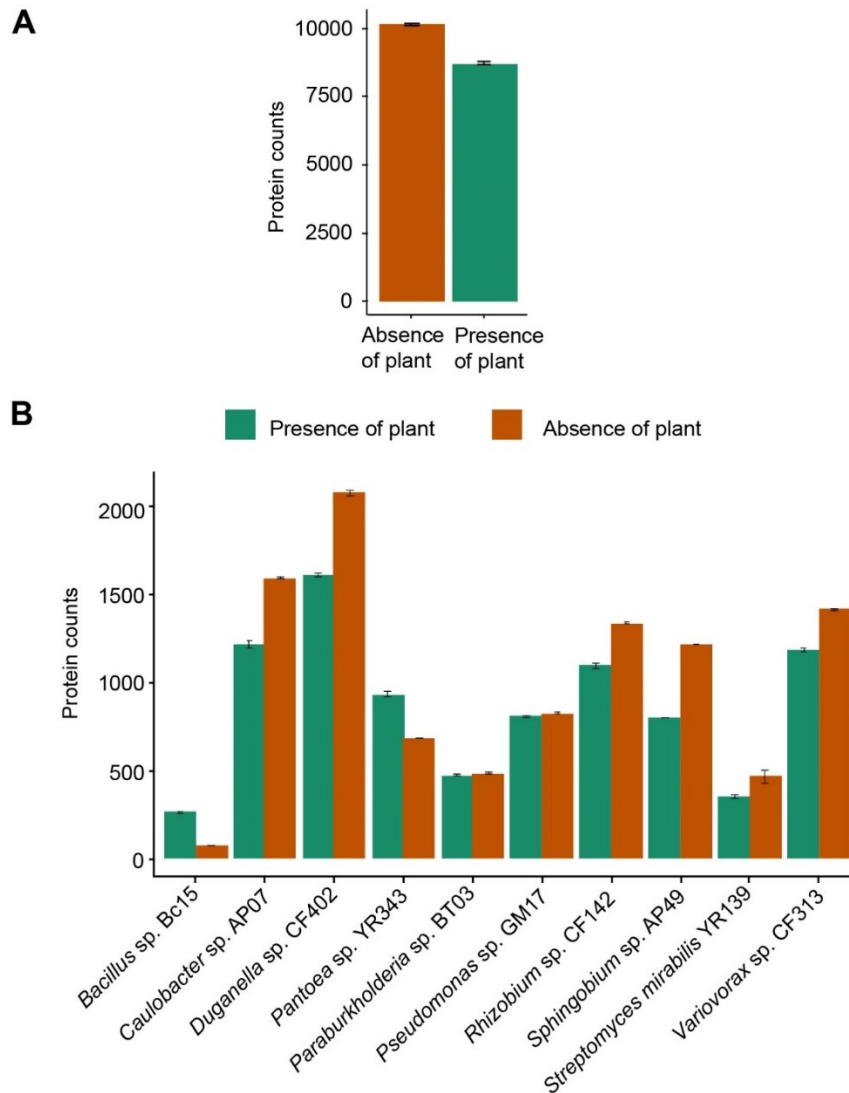


Figure 5.4: Overview of number of proteins identified in metaproteomics analysis for ten bacterial strains grown in a mixed community in the absence or presence of a plant. (A) Average number of proteins identified for the 10-member community grown in the absence or presence of a plant (n=3). Error bars represent standard deviation of mean. (B) Average number of proteins identified for each bacterial strain in the 10-member community grown in the absence or presence of a plant.

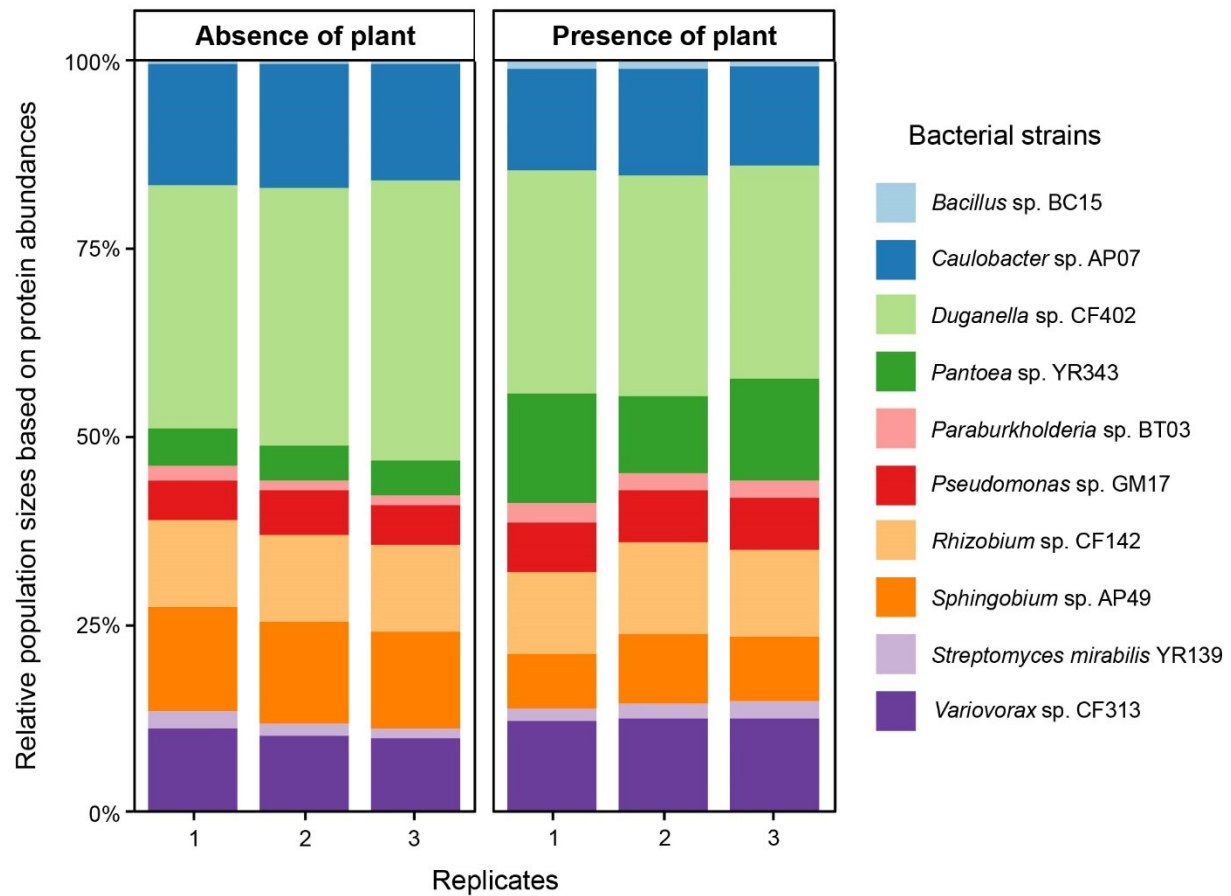


Figure 5.5: Relative population sizes of ten bacterial strains in the mixed community grown in the absence or presence of a plant assessed using summed protein abundances.

the change in the relative protein abundances for the community was not statistically significant ($p=0.1$). However, the individual organism's relative abundances changed between the control and in the presence of the plant. Five members (*Bacillus* sp. BC15, *Pantoea* sp. YR343, *Paraburkholderia* sp. BT03, *Pseudomonas* sp. GM17, and *Variovorax* sp. CF313) of the 10-member community showed statistically significant increases in their relative abundances when grown with the plant compared to the control. Three species (*Caulobacter* sp. AP07, *Duganella* sp. CF402 and *Sphingobium* sp. AP49) showed statistically significant decreases in their relative organism abundance, while changes in relative abundances of *Rhizobium* sp. CF142 and *Streptomyces* sp. YR139 were not statistically significant (Supplemental Table 4, refer online version).

Individual proteome information of each bacterial strain was extracted from the metaproteomics data, and a differential analysis was performed between control samples and those grown in the presence of a plant for each strain to identify individual strain responses to the plant root environment when present in a mixed community. Proteins with significant differences ($p\text{-value} < 0.05$, and absolute $\log_2 \text{FC} > 1$) in abundances were identified for each strain in the community (Supplemental Table 5, refer online version). Varying degrees of proteomic changes were observed across the ten strains in the mixed community in response to the plant root environment (**Figure 5.6**). Interestingly, *Bacillus* sp. BC15 again exhibited significant changes in abundance in a large portion of its proteome ($\sim 88\%$) in response to the presence of a plant root, of which 86% showed increase in abundance. Similarly, $\sim 57\%$ of the proteins in *Pantoea* sp. YR343 showed significant changes in abundance, with $\sim 84\%$ of them increasing in abundance. In all other bacterial strains, the majority of the significantly changing proteins exhibited decreases in abundance in the presence of a plant.

Bacteria shift metabolic function at a taxon-specific level in response to the plant root chemical environment

To gain insights into the metabolic potential of the microbiome in the presence of a plant, we began by first assessing the overall functional changes for each of the ten bacterial strains grown individually in the presence of a plant compared to a control (no plant) condition (**Figure 5.7**).

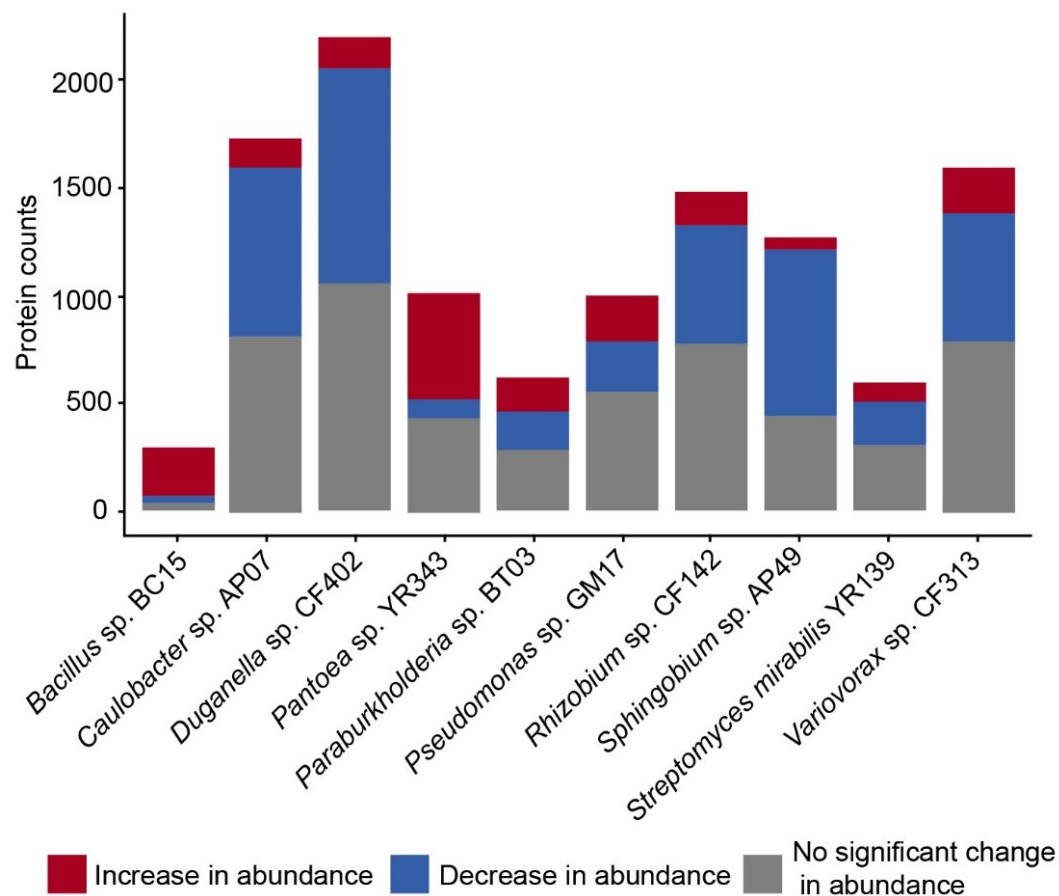


Figure 5.6: Stacked bar charts showing the number of differentially abundant proteins for ten bacterial strains grown in a mixed community in the presence of a plant compared to when grown in the absence of a plant.

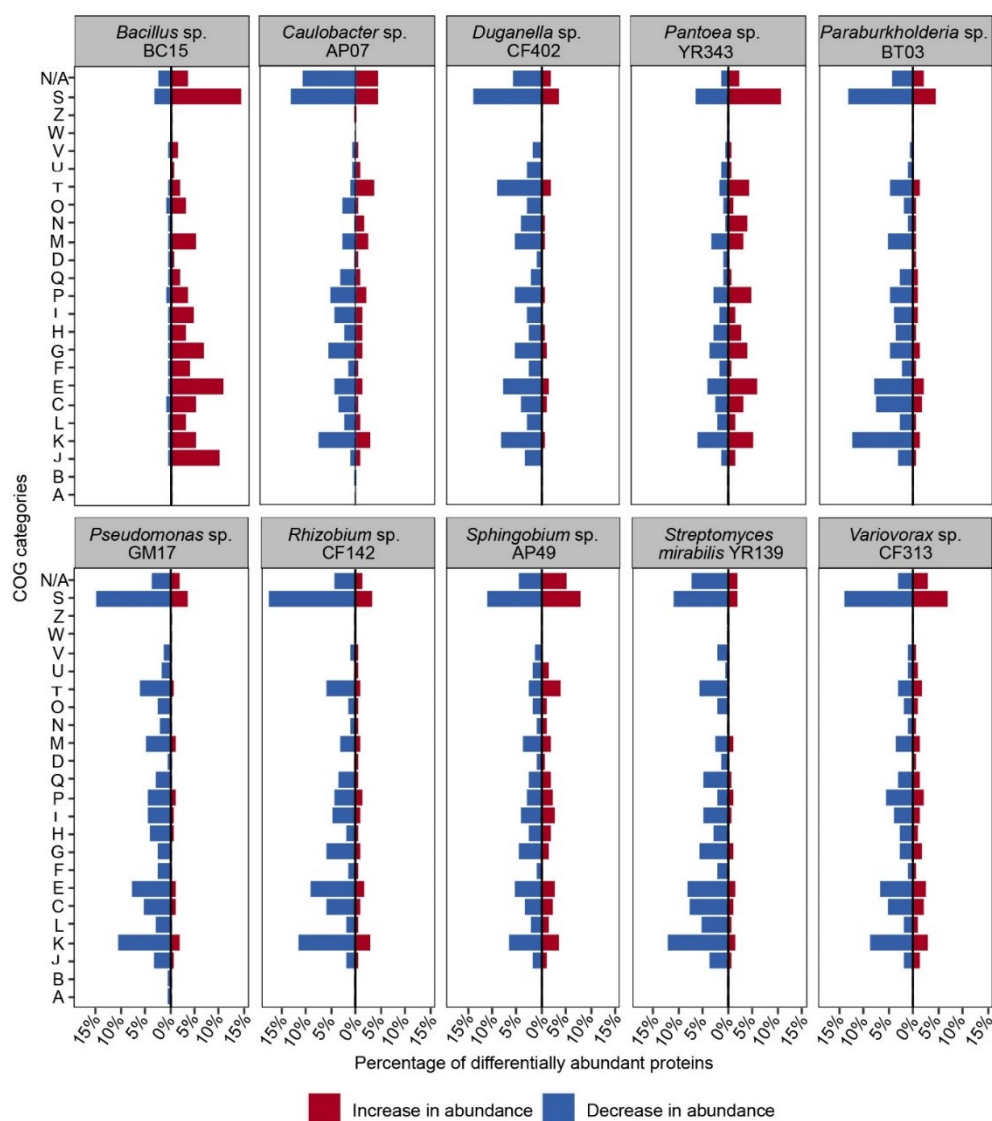


Figure 5.7: Functional analysis of bacterial strains grown individually in the absence or presence of a plant using Clusters of Orthologous Groups (COGs). Stacked bar charts show percentages of differentially abundant proteins in response to the presence of a plant. COG categories A : RNA processing and modification, B : Chromatin structure and dynamics, J : Translation, ribosomal structure and biogenesis, K : Transcription, L : Replication, recombination and repair, C : Energy production and conversion, E : Amino acid transport and metabolism, F : Nucleotide transport and metabolism, G : Carbohydrate transport and metabolism, H : Coenzyme transport and metabolism, I : Lipid transport and metabolism, P : Inorganic ion transport and metabolism, Q : Secondary metabolites biosynthesis, transport and catabolism, D : Cell cycle control, cell division, chromosome partitioning, M : Cell wall/membrane/envelope biogenesis, N : Cell motility, O : Post-translational modification, protein turnover, chaperones, T : Signal transduction mechanisms, U : Intracellular trafficking, secretion, and vesicular transport, V : Defense mechanisms, W : Extracellular structures, Z : Cytoskeleton, S : Function unknown, N/A : Not Assigned.

We classified the differentially abundant proteins into specific functional categories using the Clusters of Orthologous Group (COG), employing eggnog v5.0 in eggnog-mapper v2 (<http://eggog-mapper.embl.de/>)^{260, 261}. Proteins that have several COG categories assigned were counted for each COG category. The analysis was extended beyond the isolated conditions to explore how these metabolic processes change within the context of a constructed community (**Figure 5.8**). Notably, a substantial majority of the differentially abundant proteins identified in all bacterial strains were associated with COG categories related to metabolic processes, specifically C (energy production and conversion), E (amino acid transport and metabolism), F (nucleotide transport and metabolism), G (carbohydrate transport and metabolism), H (coenzyme transport and metabolism), I (lipid transport and metabolism), and P (inorganic ion transport and metabolism). In all isolates, we observed both increases and decreases in the abundances of proteins found in these COG categories, suggesting that the strains were undergoing dynamic metabolic remodeling in response to the chemical environment of the plant. Amino acid transport and metabolism was the most common set of proteins that showed increased abundance within the metabolism category for each of the microbes when grown with the plant. We also found that in the category related to inorganic ion transport and metabolism, *Pantoea* sp. YR343 and *Caulobacter* sp. AP07 showed the largest abundance increase in these proteins. *Sphingobium* sp. AP49, on the other hand, was one of the few isolates to show a significant increase in the abundance of proteins related to lipid metabolism.

While observing changes in metabolism of individuals in the presence of a plant revealed common responses to the chemical environment surrounding the plant, introduction of a microbial community yielded insights into the metabolic flexibility of microbes in an environment where other bacteria may be competing for similar resources. In this study, we found that *Bacillus* sp. BC15 was seemingly unaffected by the addition of the other microbial isolates and exhibited an increased abundance across several categories of metabolic proteins. *Pantoea* sp. YR343 also continued to display increased abundance across several metabolic categories, with a notable increase in proteins related to nucleotide metabolism. Surprisingly, *Sphingobium* sp. AP49 showed a dramatic decrease in the overall number of metabolism-related proteins that showed increased abundance in the presence of a plant. Lastly, we also noticed

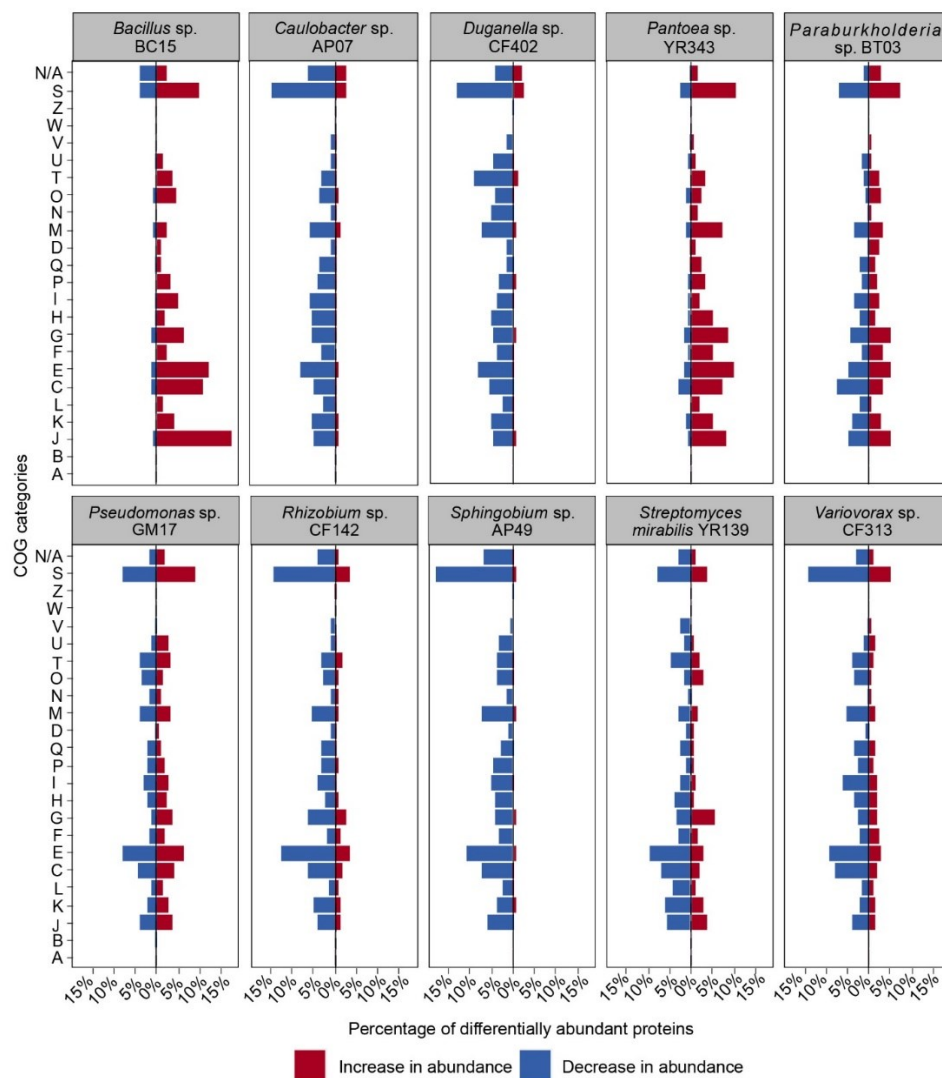


Figure 5.8: Functional analysis of bacterial strains grown in a mixed community in the absence or presence of a plant using Clusters of Orthologous Groups (COGs). Stacked bar charts show percentages of differentially abundant proteins in response to the presence of a plant. COG categories A : RNA processing and modification, B : Chromatin structure and dynamics, J : Translation, ribosomal structure and biogenesis, K : Transcription, L : Replication, recombination and repair, C : Energy production and conversion, E : Amino acid transport and metabolism, F : Nucleotide transport and metabolism, G : Carbohydrate transport and metabolism, H : Coenzyme transport and metabolism, I : Lipid transport and metabolism, P : Inorganic ion transport and metabolism, Q : Secondary metabolites biosynthesis, transport and catabolism, D : Cell cycle control, cell division, chromosome partitioning, M : Cell wall/membrane/envelope biogenesis, N : Cell motility, O : Post-translational modification, protein turnover, chaperones, T : Signal transduction mechanisms, U : Intracellular trafficking, secretion, and vesicular transport, V : Defense mechanisms, W : Extracellular structures, Z : Cytoskeleton, S : Function unknown, N/A : Not Assigned.

increased abundance in proteins involved in carbohydrate metabolism and transport, particularly for *Paraburkholderia* sp. BT03, *Pseudomonas* sp. GM17, *Streptomyces* sp. YR139, and *Variovorax* sp. CF313, in the presence of the community compared to when grown individually with the plant. Because our assessment of COG categories indicated shifts in microbial metabolism in the presence of a plant, we specifically investigated the responses of ATP-binding cassette (ABC) transporter proteins, one of the largest superfamilies of transporters within the ten bacterial strains. The quantitative analysis shown in **Figure 5.9** revealed noteworthy changes in the transporter protein abundances across the ten bacterial strains. The largest number of ABC transporter proteins were identified in the proteomics data from *Rhizobium* sp. CF142 and *Variovorax* sp. CF313 in both the individual isolate and community experiments. When grown individually, all of the isolates showed a larger fraction of ABC transporter proteins that were significantly less abundant in the presence of the plant compared to the fraction that was significantly more abundant, with the exceptions of *Bacillus* sp. BC15 and *Pantoea* sp. YR343. Although fewer ABC transporters were identified in the community dataset overall, the trends seen in the individual isolate data were repeated for the majority of isolates, with the exceptions of *Paraburkholderia* sp. BT03 and *Pseudomonas* sp. GM17, which showed slightly higher or equal fractions of ABC transporters that were significantly more abundant in the presence of the plant compared to the fraction that was significantly less abundant (**Figure 5.9**).

Whether inoculated individually or as part of a mixed community, *Bacillus* and *Pantoea* exhibited a significant increase in the abundance of proteins across a large percentage of their proteomes when grown in the presence of plant. This augmented proteome expression is paralleled by an increase in the relative population sizes of these strains within the community in the presence of a plant compared to the control. Notably, among the ten bacterial strains in the community, *Pantoea* sp. YR343 increased in its relative population size to become the third most abundant member in the community at 48 hours in the presence of a plant. In contrast, *Pantoea* sp. YR343 was among the least abundant members in the absence of the plant (**Figure 5.5**). These findings strongly suggest that the proteomic responses, characterized by an increase in protein abundance, contribute to the success of *Bacillus* and *Pantoea* in the community in the presence of a plant. Note that even though *Pantoea's* favorable response to the presence of the

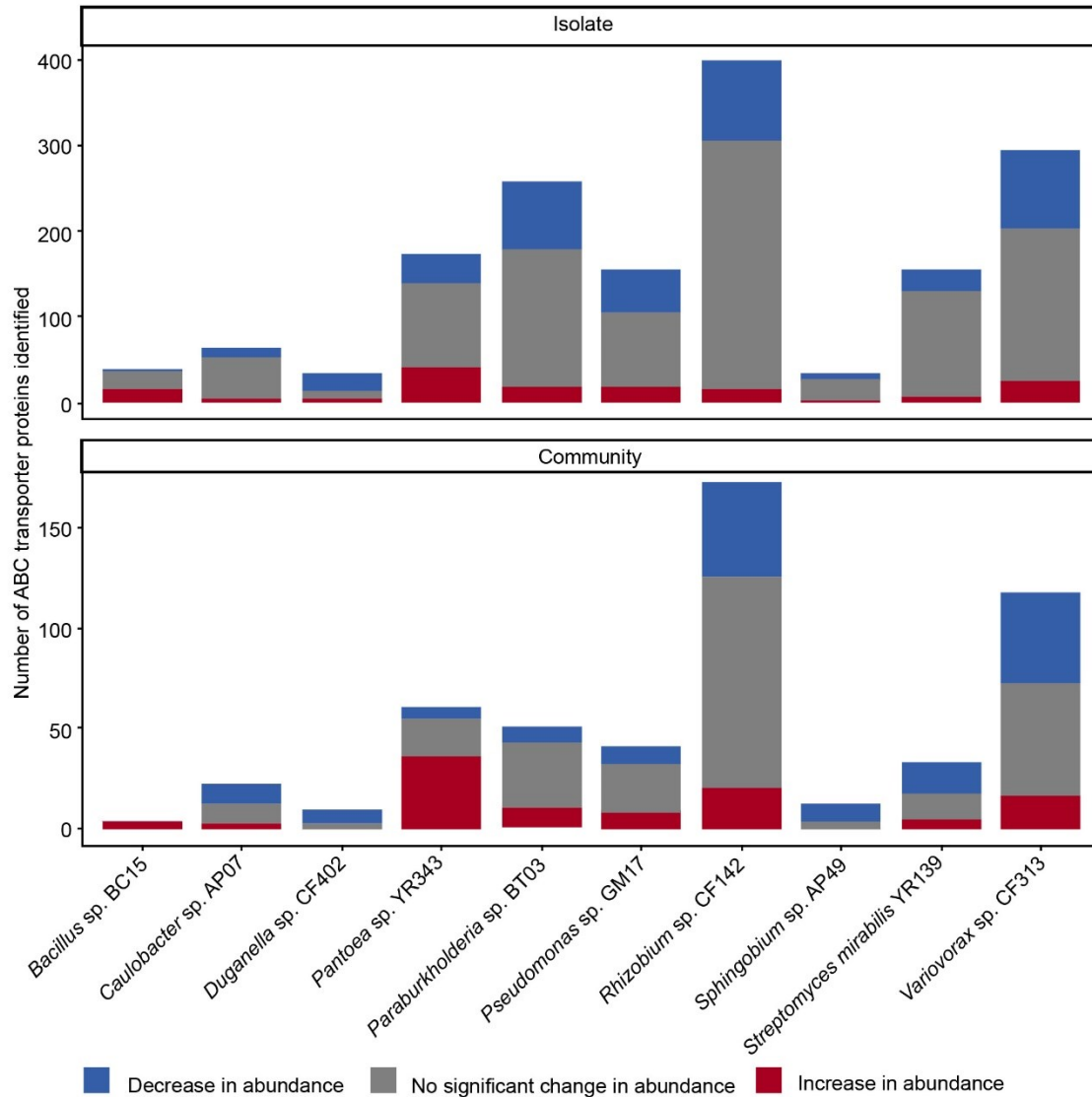


Figure 5.9: Stacked bar charts illustrating the distribution of identified ABC transporters in ten bacterial strains. Each bar represents the total number of ABC transporters identified for a bacterial strain in this study, with segments further divided to depict the counts of ABC transporters that increased in abundance (red), decreased in abundance (blue), or showed no significant change in abundance (grey) in the presence of a plant.

plant is known, the details of the molecular mechanism driving this beneficial association are not known. To delve deeper into the underlying mechanisms and functions driving this shift in relative abundance, we further investigated the proteins that increased in abundance in response to presence of a plant for both *Bacillus* and *Pantoea*.

***Bacillus* sp. BC15 becomes metabolically active in the presence of a plant**

The findings from the differential proteomic analysis for *Bacillus* sp. BC15 grown in the absence or the presence of a plant (as isolates or as individuals within a community), highlighted the dynamic changes in its proteome as a response to plant root. To understand the role of the proteins that increased in abundance in the presence of a plant, a gene ontology (GO) enrichment analysis was performed, utilizing a right-sided hypergeometric test and applying a Benjamini-Hochberg correction to pathways with a p-value less than or equal to 0.05. The focus was specifically on the biological process category, and the analysis was conducted using ClueGO v2.5.6, a Cytoscape plug-in application¹⁹⁹ (**Figure 5.10**). The GO enrichment analysis on proteins that increased in abundance when *Bacillus* sp. BC15 was grown individually revealed several GO terms as significant; these include but are not limited to organonitrogen compound metabolic process, carbohydrate metabolic process, sulfur compound metabolic process, lipid metabolic process, cellular nitrogen compound metabolic process (**Figure 5.10A** and Supplemental Table 6, refer online version). Together, these functions suggest a response of *Bacillus* sp. to the presence of plant roots involving specific metabolic adaptations, consistent with the COG analyses (**Figure 5.7**).

The GO enrichment analysis for *Bacillus* sp. BC15 grown in the community also showed similar results where several GO terms associated with the proteins that increased in abundance in the presence of the plant were found to be significant, these predominantly are involved in several metabolic processes, cellular component biogenesis, generation of precursor metabolites and energy (**Figure 5.10B** and Supplemental Table 6, refer online version). This overlap of significant GO terms highlights similar functional responses by *Bacillus* to the influence of a plant root, whether grown individually or in community.

Protein level analysis revealed an increase in abundance of proteins associated with active metabolism in *Bacillus* in the presence of a plant. Additionally, several proteins involved in

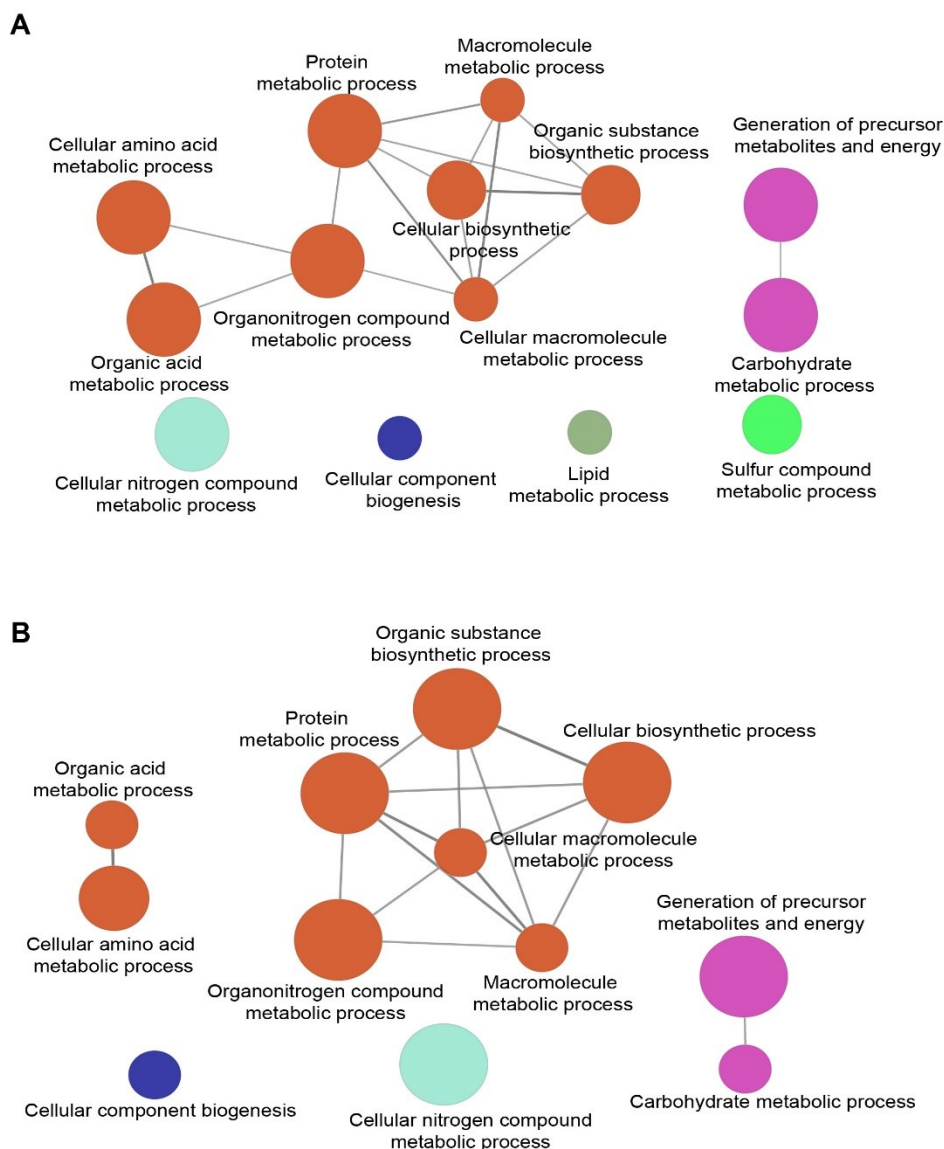


Figure 5.10: Functional changes in *Bacillus* sp. BC15 as a response to the chemical environment of the plant root. Gene ontology (GO) networks representing the significant results from GO enrichment analysis of biological process category (right-sided hypergeometric test, $p\text{-value} \leq 0.05$, Benjamini-Hochberg correction) of proteins that increased in abundance in the presence of a plant. GO enrichment analysis was performed using ClueGo application in Cytoscape (minimum number of genes/GO term = 3, % genes for group merge for GO term grouping = 50) when grown (A) individually or (B) in a mixed community.

various sporulation stages and small, acid-soluble spore proteins were identified in the dataset, some of which are differentially abundant when grown with the plant, suggesting a subset of the population is sporulated (Supplemental Table 7, refer online version). However, the functional analysis of *Bacillus* proteins that increased in abundance in the presence of a plant whether grown individually or in community suggests that these proteins are primarily involved in metabolism and cellular processes. This suggests that the *Bacillus* population becomes metabolically active in response to chemical cues produced by the plant.

***Pantoea* sp. YR343 shows increased abundance of proteins related to chemotaxis, motility, nutrient transport, and metabolism in response to the plant root chemical environment**

To better understand the functions of proteins that showed increased abundance in *Pantoea* sp. YR343 in the presence of the plant, we performed a gene ontology (GO) enrichment analysis as described above. When *Pantoea* was grown individually in the presence of the plant, significant GO terms were found in signal transduction and cell motility (Supplemental Table 8, refer online version). When *Pantoea* sp. YR343 was grown in a community in the presence of a plant, this analysis revealed significant GO terms associated with translation, organonitrogen compound metabolic process, cellular component assembly, cellular amino acid metabolic process, generation of precursor metabolites and energy, and membrane organization (Supplemental Table 8, refer online version).

We next compared the differentially abundant proteins of *Pantoea* sp. YR343 grown individually or in a community to determine whether we could identify a common response to growth in the presence of a plant. Somewhat surprisingly, only a small number of proteins were found in both datasets, with the majority increasing in abundance (**Figure 5.11A** and Supplemental Table 9, refer online version). The functions of these common proteins include cell motility, nutrient transport, and metabolism as common responses to the presence of a plant. Among the proteins that increased in abundance in the presence of plants were several methyl-accepting chemotaxis proteins, a response regulator with a CheY-like receiver domain and flagellar machinery proteins (flagellin, flagellar motor switch protein FliM, flagellar M-ring protein) that are crucial for chemotactic responses and motility of bacteria (**Figure 5.11B**). The methyl-accepting chemotaxis proteins detect the environmental signals and initiate a chemotactic response in bacteria. This

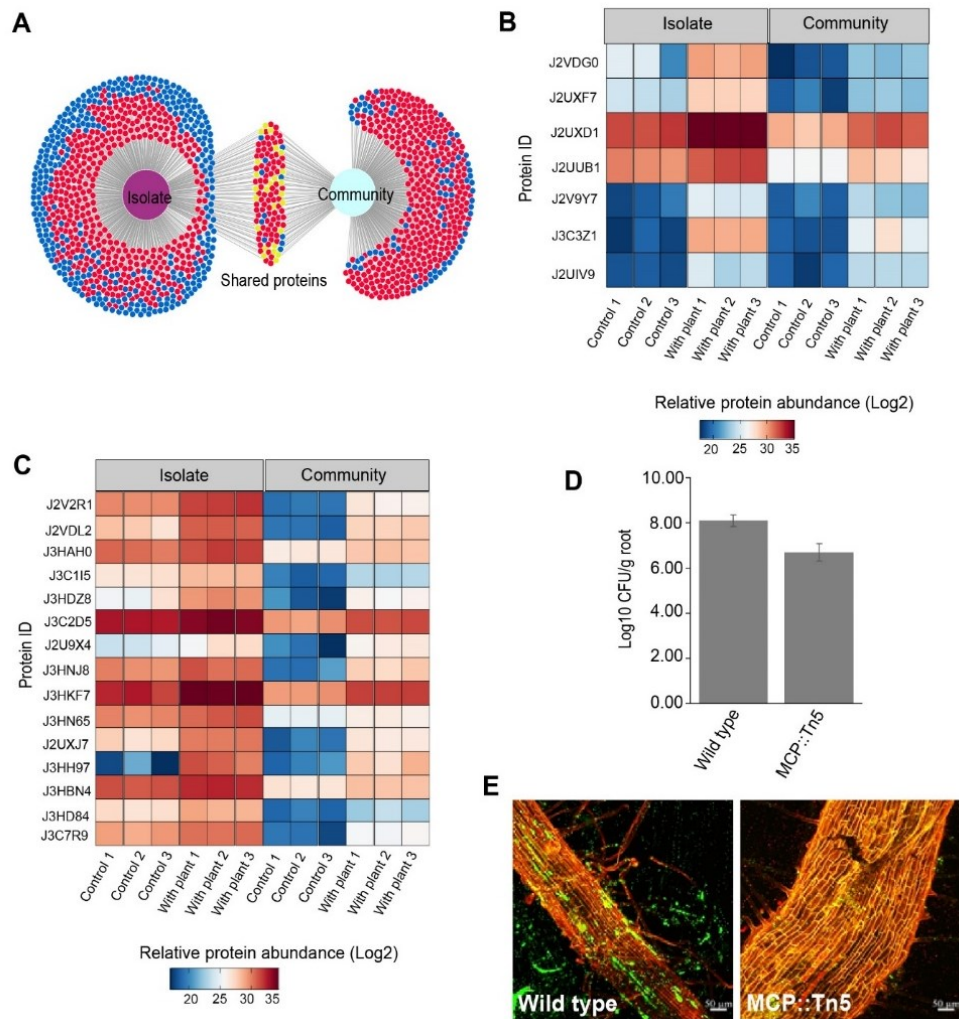


Figure 5.11: Proteomic analysis of *Pantoea* sp. YR343 in response to the chemical environment of a plant root. (A) Overlap of differentially abundant proteins (based on Student's t-test performed between proteins identified in *Pantoea* sp. YR343 grown in the absence or presence of a plant) between individual (left) and mixed community (right) conditions. Each node represents a unique protein accession, and the color indicates whether the relative protein abundance increased (red), decreased (blue), or increased in one condition but decreased in another (yellow) in presence of plant compared to absence of plant. Figure generated using DiVenn 2.0²⁶². (B) Heatmap illustrating the relative protein abundances of chemotaxis and motility related proteins of *Pantoea* sp. YR343 when grown individually and in mixed community. (C) Heatmap illustrating the relative protein abundances of transporter proteins of *Pantoea* sp. YR343 when grown individually and in mixed community. (D) Bar chart showing a transposon insertion in gene (PMI39_02190) encoding a chemoreceptor (Protein ID J3C3Z1) shows reduced colonization compared to the wild type strain based on colony counts (E) Confocal images of *Populus* roots colonized by wild-type *Pantoea* sp. YR343 cells expressing GFP (left) compared to the MCP mutant expressing mCherry (right).

signal is transduced through the chemotaxis machinery and affects the flagellar motor complex allowing the bacteria to move towards favorable conditions such as root exudates. In addition, several transporter proteins essential for uptake of essential nutrients increased in abundance when *Pantoea* sp. YR343 was grown in the presence of a plant, including ABC-type transport system (sugar, ribose, amino acid, thiamine, iron), phosphotransferase system (PTS) proteins involved in uptake of sugars (glucose, Mannose/Fructose/Sorbose family), maltose operon periplasmic protein MalM involved in maltose uptake, and ammonium transporters involved in uptake of ammonium, (**Figure 5.11C**). Increased abundance of these transport proteins suggests that these resources may be more readily available in the presence of plant roots or it may reflect metabolic strategies employed by *Pantoea* sp. YR343 that allow these cells to thrive and compete within the chemical environment surrounding the plant root. Several proteins involved in carbohydrate metabolic processes (phosphofructokinase, a key regulatory enzyme in glycolysis), amino acid biosynthesis (histidine biosynthesis bifunctional protein HisB, involved in biosynthesis of histidine, an essential amino acid) and amino acid metabolism (methionine synthase II, involved in metabolism of methionine, an essential amino acid) also increased in abundance in the presence of a plant. The increase in abundance of proteins involved in transport of nutrients as well as metabolic processes suggest a coordinated regulation of proteins in *Pantoea* in response to of the chemical environment surrounding the plant root.

Finally, we examined the loss of function phenotype of one of the common proteins, a methyl-accepting chemotaxis protein (MCP). *Pantoea* sp. YR343 has 21 predicted methyl-accepting chemotaxis proteins based on a pfam search of proteins possessing a pfam00015 “MCP signal” domain, of which 10 MCPs showed an increase in abundance in the presence of the plant. During previous studies, we isolated a transposon mutant affecting one of these MCPs, encoded by PMI39_02190²⁵². Using fluorescently labeled wild type and mutant strains, we assessed colonization of *P. trichocarpa* BESC819 and found that the MCP mutant strain was significantly impaired in root colonization compared to the wild-type strain based on plating and imaging assays (**Figure 5.11D, 5.11E**). These data suggests that the increased abundance of this protein in the presence of the plant is functionally significant and plays a role in efficient root colonization.

Discussion

The significance of the plant root chemical environment in shaping microbiome assembly is well-established^{24, 263, 264}. In this study, we employed an in-depth proteomics approach to understand the functional adaptations of microbes in response to plant chemical signals. Historically, proteomics analysis of microbes associated with plants has proven difficult because microbial proteins are often swamped by those of the plant, thus requiring novel approaches to study each component separately. Here, we employed a simple nylon filter-based system in which bacteria were inoculated on a filter that was placed on plant roots, allowing chemical communication between the plant and bacteria but not physical contact. While this system can be used to examine responses to plant root exudates, it excludes responses driven by soil factors or physical interactions with the plant root. We validated this approach by using a plant-responsive reporter strain of *Pantoea* sp. YR343 which showed increased GFP fluorescence due to promoter activation in the presence of a plant root compared to control conditions indicating that the cells were able to respond to the chemical environment of the plant through the filter²⁵².

To study responses of individual bacterial strains, an aliquot of culture (as described in methods) was inoculated on filters either with or without a plant. To study community responses, we started with equal ratios of ten bacterial strains to minimize the bias due to unequal starting ratios. This approach of using equal initial ratios allows each species to have an equal opportunity for interactions both with the plant root, as well as other microbes, and ensures the proteomic changes are attributable to experimental conditions i.e, plant root chemical signals rather than initial ratios. In both the microbial isolates as well as the microbial community samples, the proteome measurement depths for individual strains grown in the absence or presence of the plants were comparable. Not surprisingly, the measurement depths for individual strains grown in the community were lower compared to the proteomic measurements of individually grown bacteria. The proteome coverage observed for the metaproteomic studies, however, are similar to other metaproteomic studies using constructed microbial communities¹⁸⁹.

The proteome profiles of individual bacterial strains in a constructed community are influenced by many factors, including interspecies interactions and competition for nutrients and space which ultimately impact the biomass and measurement depth of each strain in the

community sample. To discern whether the changes observed in the proteomes of bacterial strains grown in the absence or the presence of plant are driven by biological effects rather than measurement-depth, we employed a rank-based sorting of proteins based on abundances, which revealed that the decreases in protein abundances within specific functional categories are not a systemic change, supporting the conclusion that the observed proteomic changes are biological.

Within the functional categories, we observed that in most of the bacterial strains the majority of significantly changing proteins decreased in abundance in the presence of the plant compared to the control. Because we only collected one timepoint in this study, however, it is unclear whether this result represents a step in the transition from growth in R2A liquid medium to the plant environment or is indicative of sustained growth in the presence of the plant. It is also possible that the quantity of bacteria applied to the filter was higher than would be sustained by plant exudates in the native rhizosphere, leading to nutrient limitation over time. To distinguish between these possibilities, it would be informative to collect multiple samples over a longer time period to determine how the cells adapt to the plant environment.

Metaproteomic analyses revealed population size shifts in the community in the presence of a plant and the bacterial species that displayed positive responses to the plant environment were pinpointed. The observed relative abundance changes of bacterial strains suggest a dynamic response to the chemical milieu of the rhizosphere (**Figure 5.5**). Notable increases in organism relative abundances exhibited by *Bacillus* sp. BC15 and *Pantoea* sp. YR343 when grown with a plant underscores the intricate interplay between bacterial community structure and the plant root environment. However, it is essential to acknowledge that generalizing these responses might oversimplify the complex dynamics at play. Therefore, gaining a nuanced understanding of the adaptive strategies employed by individual species becomes crucial for understanding the molecular mechanisms governing these responses.

Within this study, we observed that most of the differentially abundant proteins identified were in metabolism-associated COG categories and displayed both increases and decreases in abundance with respect to the presence of a plant. In the community, all the identified ABC transporters in *Bacillus* and about 60% in *Pantoea* increased in abundance in response to the presence of the plant. These observed changes in transporter protein abundances and relative

organism abundances in these strains strongly suggested a strategic reprogramming of nutrient acquisition strategies in the presence of a plant and underscored the significance of not only the sheer quantity but also the dynamic regulation of transporter proteins in bacterial adaptations to the plant milieu. Upregulated nutrient transporters imply an active pursuit of essential resources, potentially indicating symbiotic relationships or competitive interactions within the microbial community. Integrating these transporter results with the increase in abundance of metabolism related proteins, we observe a coordinated interplay between nutrient uptake and metabolic pathways in *Bacillus* and *Pantoea*. While *Bacillus* sp. BC15 and *Pantoea* sp. YR343 both exhibit the highest levels of metabolism-related protein upregulation, there are interesting adaptations within the remaining members of the community when competing with *Bacillus* and *Pantoea*. For example, *Paraburkholderia* sp. BT03 and *Pseudomonas* sp. GM17 show a notable increase in the number and categories of metabolism-related genes that are upregulated in the presence of a community on a plant. We also observed an increase in abundance for proteins that fall into the category of nucleotide metabolism in most of the strains when grown in a community on the plant, with the exception of *Sphingobium* sp. AP49, *Caulobacter* sp. AP07, and *Duganella* sp. CF402.

Our proteomic data also indicated the presence of proteins involved in sporulation in *Bacillus* sp. BC15. *Bacillus* sp. are known to survive harsh environmental conditions by forming endospores that remain quiescent with minimal metabolic activity. Under favorable conditions, there is a morphogenic transition from dormant spores that then germinate into a metabolically active vegetative cell^{265, 266}. Previous studies have shown that *Bacillus* sp. BC15 forms spores in R2A media, however, the dynamics of spore formation in this growth condition are not well understood¹⁹². While inoculating *Bacillus* sp. BC15 in the form of spores in this study was not intentional, our data indicate that *Bacillus* undergoes germination into vegetative cells from dormant spores and becomes metabolically active in the presence of the plant suggesting that there are plant-specific cues capable of signaling germination.

Chemotactic bacteria have been shown to have a competitive advantage for plant colonization²⁶⁷. Consistent with this, we observed that proteins related to chemotaxis were highly represented among the upregulated proteins on the plant both with and without the community. Using a transposon mutant that was identified in a previous study of *Pantoea* sp.

YR343²⁵², we were able to show that loss of function of one of the methyl-accepting chemotaxis proteins (encoded by PMI39_02190) that is more abundant in the presence of the plant resulted in reduced plant colonization compared to the wildtype strain. It will be of particular interest to determine to which plant-related compounds the bacteria are responding and how bacterial colonization patterns correspond to the distribution of those compounds. Combined with the data about metabolism and transport, these data suggest that under these conditions, *Pantoea* sp. YR343 is well-adapted to survival in the presence of a plant. This result is consistent with previous studies using a plant agar-based system in which *Pantoea* sp. YR343 displayed a significant increase in its relative abundance in the presence of a plant²⁵⁸. The identified functions of *Pantoea* as a response to plant chemical environment cover a spectrum of activities that collectively enable it to sense, respond, and adapt to the dynamic conditions presented by the plant root environment. These adaptations are crucial for establishing beneficial interactions, nutrient acquisition, and survival in the rhizosphere. Further studies focusing on these specific protein functions and their roles in interaction with *Populus* roots will provide more detailed insights into the molecular mechanisms underlying *Pantoea*'s adaptation to the *Populus* root environment.

Our comparative proteomic and metaproteomic analysis have provided unique and detailed insights into the responses of various bacterial species to the plant root environment, demonstrating how these species respond and adapt on both an individual level, as well as the community level, in order to effectively navigate nutrient availability and competition in the rhizosphere.

CHAPTER 6: UNTARGETED METABOLOMIC ANALYSIS TO CHARACTERIZE THE DYNAMIC ROOT AND EXUDATE METABOLOMES OF *POPULUS TRICHOCARPA*

Author contributions

Manasa R. Appidi contributed to study design, sample preparation for metabolomics analysis, LC-MS/MS measurements and data analysis. Dr. Paul Abraham provided detailed input into experimental design, MS data acquisition, and assistance with data analysis/ interpretation and figures.

Introduction

Plants release their photosynthesized carbon through their roots into soil. This process called exudation can occur through various passive and active transport mechanisms including diffusion and secretion^{242, 268, 269}. Plant roots exudates contain wide range of compounds including primary metabolites (e.g. sugars, amino acids, organic acids) and secondary metabolites (e.g. phenolic compounds, terpenoids) into its surrounding environment²⁶⁹. The root exudates play several pivotal roles in contributing to plant development, plant-soil-microbe interactions, modify soil properties and carbon sequestration^{240, 268, 270}. The exudates are known for their role in shaping the root associated microbiome composition by providing nutrients to microbes and serve as chemical signals that attract beneficial microbes and defense signals against pathogens, which impacts the plant health and development^{76, 202, 271}. Root exudation is a complex and dynamic process. The composition of the exudates is impacted by several factors including plant species, age of the plant, several external biotic factors such as microbes and abiotic factors such as water and nutrient availability^{22, 272}. Given their integral role in plant health and soil chemistry, studying the root exudates under various stress conditions is crucial for advancing our understanding of how plants adapt to different environmental conditions.

Populus is a short-rotation woody crop, used as a biofuel feedstock^{11, 12}. Modifying plant traits by genetic engineering can improve the conversion of plant biomass to biofuel and bioproducts²⁷³. Lignin is one such plant trait altered in these efforts as it can enhance biomass processing for biofuel production by reducing lignin recalcitrance, and thereby making the

extraction of sugars from cellulose more efficient²⁷⁴. However, altering lignin content can have pleiotropic effects on the plant (impacting other traits, like plant growth and metabolism)²⁷⁵. It is hypothesized that changes in lignin content in *Populus* sp. would alter its root physiology, metabolism, and exudation. These changes are known to subsequently impact associated microbiome composition and function²⁷⁶, which play a major role in plant health. Therefore, it is crucial to study the dynamics in the root and exudate metabolomic profiles of plants with varying lignin contents. The studies focused on understanding the differences in the root and exudate profiles of *Populus* genotypes with varying lignin contents are currently lacking. To address the above-mentioned limited research in *Populus* sp., we aimed to capture the differences between the root intracellular and exudate metabolite profiles of two *Populus trichocarpa* genotypes BESC-24 and BESC-375 varying in lignin contents, BESC-24 has relatively low-lignin content compared to BESC-375^{277, 278}. Additionally, the effect of different nutrient availability conditions, one of the important abiotic factors that impacts the root and exudate metabolome was assessed by collecting roots and exudates.

Root exudation is spatio-temporally dynamic process with varying composition and concentrations of metabolites around different root zones collected at different timepoints that further impacts the surrounding soil chemistry and microbiome composition, making it essential to study these spatio-temporal dynamics²⁷⁹. Studies focused on root exudate analysis typically involve collecting whole exudate samples with less emphasis on spatial sampling of exudates^{63, 276}. Hydroponic culture systems are most commonly used in the root exudate studies but are not suitable for spatial sampling of exudates¹⁵⁷. Quantitative analysis of root exudates across space in soil-based systems is also challenging due to the methodological difficulties and any changes in soil physical characteristics can lead to altered root exudation profiles^{157, 280}. To overcome these limitations with hydroponics and soil-based systems, in this study we used a widely used plant-agar system that enables spatial sampling of root exudates from distinct root regions. Further, to capture the temporal dynamics in the root and exudate metabolic profiles, the samples were collected at two timepoints i.e., 48 hr and 72hr after the plants were transferred to agar plates. These timepoints were selected to capture the temporal differences in metabolomic profiles after allowing the plants to acclimatize to the experimental conditions.

Here, we hypothesized that (1) *Populus trichocarpa* genotypes with differing lignin content will exhibit distinct variations in root metabolites, (2) metabolomic profiles extracted from roots and root exudates of *P. trichocarpa* will differ significantly, (3) distinct metabolite profiles will be identified between different nutrient availability conditions, (4) there will be spatial and temporal differences in metabolite abundances along the plant root axis. To analyze the metabolite profiles of roots and exudates, LC-MS/MS based untargeted metabolomics approach was used. To our knowledge, this is the first comprehensive untargeted metabolomics report of *Populus* root metabolome and spatio-temporally resolved exudate profiles of two *Populus* genotypes with varying lignin content, grown under different nutrient conditions.

Material and methods

***Populus* genotypes and growth conditions**

To assess the effect of variations in lignin contents on root metabolome and exudate profiles, two genotypes of *Populus trichocarpa* namely BESC-24 (low-lignin) and BESC-375 (high-lignin) were used in this study. Axenic stem cuttings of these genotypes were propagated from axenic stock plants. Briefly, axenic sub-cultured shoot tips of each genotype were rooted in 1X Murashige and Skoog (with micro and macronutrients, vitamins and glycine from Caisson Laboratories) with 0.05% MES hydrate (Sigma), 3% sucrose (Sigma), 0.5% activated charcoal (Sigma), 0.01% PPM (Plant Cell Technology) and 0.18% Gelrite (Sigma). The pH of the media was adjusted to 5.7 using KOH before autoclaving. Plants were kept in a growth room for 4-5 weeks until roots developed under a 16-hour/8-hour day/night regime ($120\mu\text{mol m}^{-2}\text{s}^{-1}$) at 22°C. Plants were transferred to experimental conditions after roots were established.

Experimental design

Plant in agar experimental platform is employed in this study, which facilitates in sampling of both the root tissue and spatial sampling of root exudates. Two independent variables were included in this study (A) plant genotype (*Populus trichocarpa* BESC-24 and BESC-375) and (B) nutrient availability (nutrient-rich media containing ¼ strength Hoagland's basal salt in 0.3% agarose and nutrient-limited condition without Hoagland's basal salt in 0.3% agarose) (Hoagland

modified basal salt mixture; PhytoTech Labs) (**Figure 6.1**). At the beginning of the experiment, a total of 24 rooted plant cuttings per genotype were randomly selected and half of them were transferred to petri dishes containing nutrient-rich media and remaining half of the plants were transferred to nutrient-limited media (n=12 for each nutrient condition per genotype), with roots submerged in 25 ml media. To avoid the plant leaves touching the surface of the media, the petri dish lids were replaced with sterile plastic containers. Media controls (n=3 for each nutrient condition) without the plant were also prepared. All plants and media controls were kept in growth chambers with 16-h days at 28°C. Half of the plants for each treatment condition (a combination of genotype and nutrient condition) were collected at each of the two timepoints 48 hr and 72 hr to observe the temporal changes in the metabolite profiles of root and root exudates (n=6 plants for a combination of genotype, nutrient condition and timepoint). At each timepoint, the root tissues from each plant were excised with a sterile scalpel, transferred to a separate Eppendorf tube, frozen immediately in liquid nitrogen and stored at -80°C until further processing. Additionally, equal volumes of agar plugs surrounding distinct root regions (root base, root tips, and distant regions from root base and root tip) were excised to capture the spatial dynamics of root exudation, frozen immediately in liquid nitrogen and stored at -80°C until further processing.

Metabolite extraction from root tissue

The root tissue was dried using a freeze-drier (Labconco Equipment Co., KS, USA) and the dry weights were recorded for each plant. Root tissue was ground into fine powder by cryogrinding in liquid nitrogen (Spex SamplePrep 2010 Geno/Grinder), followed by metabolite extraction using a biphasic extraction method. Briefly, equal volumes of ice-cold hydrated ethyl acetate and LC-MS grade water (total volume = 2mL) were added to powdered root tissue and vortexed for one minute at maximum speed. Samples were incubated on ice for 10 minutes, and subjected to vortexing again, followed by overnight incubation on ice on in the fridge (4°C). Samples were centrifuged for 10 minutes at max speed (21000 g) at 4°C. The ethyl acetate and aqueous supernatant fractions were separated into fresh tubes by aspiration. The ethyl acetate extracts were dried completely in chemical fume hood and resuspended in 50 µL of organic solvent (70% acetonitrile with 0.1% formic acid). The aqueous extracts were filtered using pre-wet 10KDa

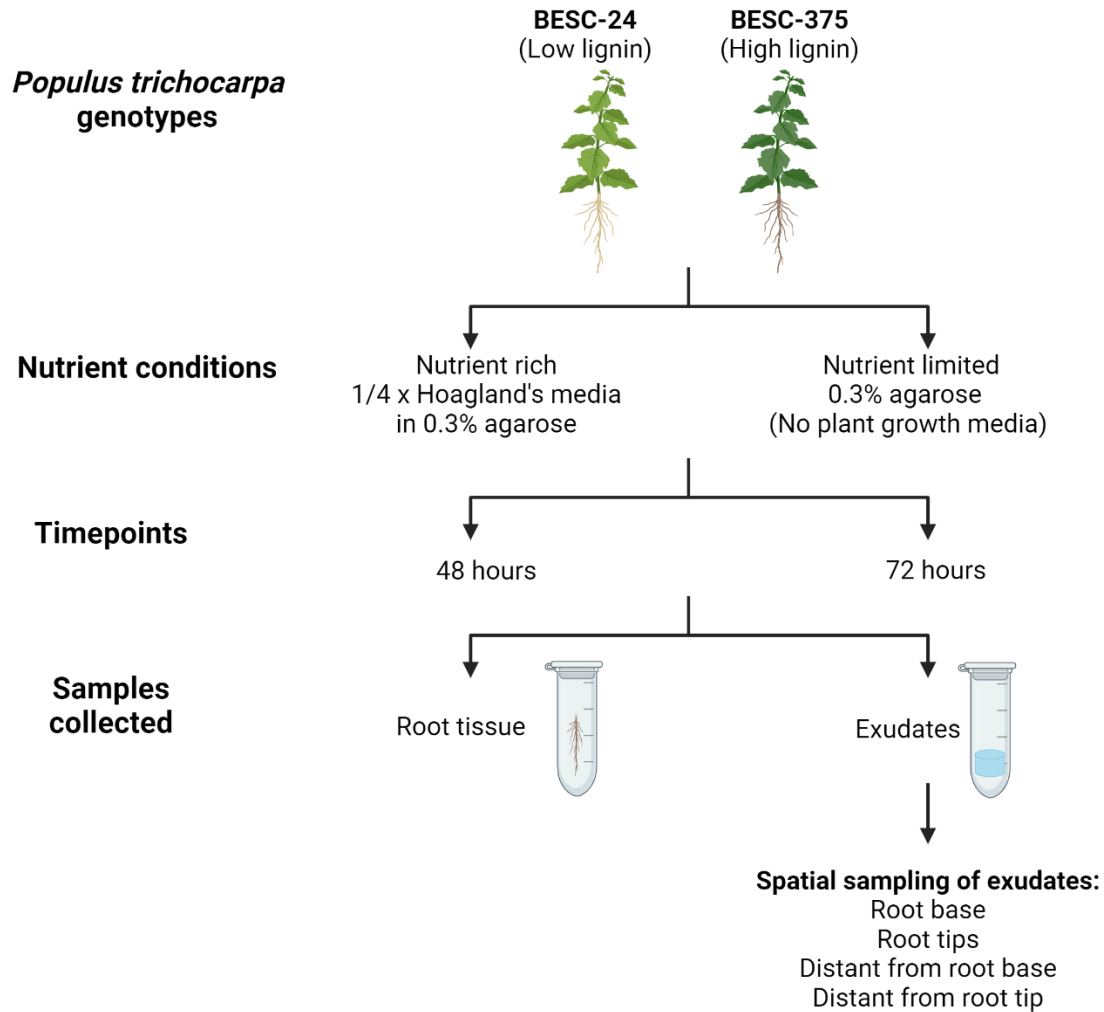


Figure 6.1: Schematic representation of the experimental design. Axenic rooted stem cuttings of *Populus trichocarpa* BESC-24 and BESC-375 genotypes were transferred to nutrient-rich and nutrient-limited media conditions in plant-agar system. Root tissues and spatially resolved exudate samples were collected at 48 hr and 72 hr after transfer. n=6 plants for each timepoint for a combination of genotype and nutrient condition. Figure generated using BioRender.com.

centrifugal filters (Sartorius Vivaspin® 2 Centrifugal Concentrator Polyethersulfone) by centrifugation at 4,500 x g) to remove any residual media or particles. The aqueous extracts were then lyophilized and resuspended in 50 µL of aqueous solvent (98% LC-MS water, 2% acetonitrile with 0.1% formic acid). All samples were stored short-term at 4°C until LC-ESI-MS/MS measurements.

Metabolite extraction from agar plugs for root exudates

The agar plugs collected for root exudate sampling were freeze-dried and metabolites were extracted using a biphasic extraction protocol described above for root tissue. After the extraction and overnight incubation, the samples were centrifuged for 10 minutes at max speed (21000 g) at 4°C. The agar precipitate was discarded after the aqueous and organic supernatant fractions were separated and transferred to new Eppendorf tubes. The aqueous and organic extracts were dried and reconstituted as described in the above section.

Metabolite analysis using Liquid chromatography- tandem mass spectrometry (LC-MS/MS)

All samples were analyzed by liquid chromatography coupled tandem mass spectrometry (LC-MS/MS) using high pressure liquid chromatography coupled with a Q- Exactive Plus mass spectrometer (Thermo Fisher Scientific). For each sample, 10 µL of the resuspended extract was injected and chromatographic separation was performed across an in-house-pulled nanospray emitter (75 µm x 150 mm) packed with 1.7 µm C18 Kinetex RP C18 resin (Phenomenex) by a 30-min linear organic gradient (250 nL/min flow rate) from 5% aqueous solvent (2% acetonitrile and 0.1% formic acid) to 100% organic solvent (70% acetonitrile and 0.1% formic acid). All MS data were acquired with Thermo Xcalibur (version 4.3), using the top N method where N could be up to 5. Data was acquired using positive ionization mode. Target values for the full scan MS spectra were 3×10^6 charges in the 135 to 2000 m/z range with a maximum injection time of 100 ms. Transient times corresponding to a resolution of 70,000 at m/z 200 were chosen and a 2.0 m/z isolation window and isolation offset of 0.5 m/z were used. Fragmentation of precursor ions was performed by stepped higher-energy C-trap dissociation (HCD) with normalized collision energies of 10, 20, and 40 eVs. MS/MS scans were performed at a resolution of 17,500 at m/z

200 with an ion target value of 1.6×10^5 and a 50 ms maximum injection time. Dynamic exclusion was set to 10 s to avoid oversampling measurements of abundant metabolites.

Untargeted metabolomics data processing and analysis

The MS raw data files for aqueous and ethyl acetate fractions were analyzed separately using compound discoverer software, version 3.3.1 (Thermo Fisher Scientific). The Input Files node was used to submit all the raw data associated with each extraction fraction. The Select Spectra node was used with polarity mode Is +. The Detect Compounds node was used with default parameters of 5 ppm mass tolerance, 10,000 minimum peak intensity, use most intense isotope only, chromatographic peak detection S/N threshold 1.5, remove baseline false, and compound detection of $[M+H]^+$, $[M+2H]^+$, $[M+3H]^+$, $[M+H-H_2O]^+$, $[M+K]^+$, $[M+Na]^+$. The Group Compounds node was used with 5 ppm mass tolerance, 0.2 min retention time (RT) tolerance, peak alignment was set to false, preferred ions $[M+H]^+$, $[M+2H]^+$, $[M+3H]^+$ and area integration set to most common ion. The peak rating filter threshold of 4 for a minimum of 3 files. The Search mzVault node was used with precursor mass tolerance of 10 ppm with the HighChem HighRes algorithm to compare MS² spectra from raw files to the NIST_2020_MSMS High Resolution library and GNPS spectral libraries (ALL_GNPS downloaded from <https://external.gnps2.org/gnpslibrary> on 12/08/2023). A scripting node was included in the processing workflow to add InChiKey identifiers for the obtained compound annotations. Compounds that have MS/MS fragment matching score of >0.5 support a Class 2 identification²⁸¹. Metabolite features with poor peak integration were resolved through automated data processing step, where peak areas from indistinguishable features (based on their annotated molecular weight and delta mass) were aggregated using a custom R script (available at <https://github.com/dlcarper/Metabolite-Condensing-Script>). Relative abundances of each compound were assessed by the integrated peak areas across samples. The data was further filtered to remove compound features that had an average abundance value equal to or less than the abundances observed in the media control samples. Permutational multivariate analysis of variance (PERMANOVA) was performed to assess the experimental factors that have a significant influence on the metabolite profiles measured in this study. Additional data post-processing was conducted using MetaboAnalyst v6.0²⁸², default settings were applied for

handling the missing values in the dataset and log10 data transformation was performed to reduce skewness. The data appeared uniformly distributed and consistent, and further normalization steps were deemed unnecessary. Partial Least Squares Discriminant Analysis (PLS-DA) was then performed to evaluate differences between the two genotypes and between root and rhizodeposition samples. Additionally, ANOVA-Simultaneous Component Analysis (ASCA) was applied to assess the interactions between environment (nutrient condition) and time (48 hr and 72 hr) specifically within the BESC-24 genotype. A modest leverage threshold of 0.7 was set to balance sensitivity and specificity, minimizing overfitting while capturing the main effects. The alpha was also set to 0.05 to maintain a standard threshold for identifying statistically significant interactions between environment and sampling time.

Results and discussion

Global results of untargeted metabolomic analysis and data processing to obtain non-redundant metabolite feature lists

The plant metabolic processes and exudates are influenced by many factors in nature such as their genetics, environmental conditions and biotic and abiotic stressors. For a comprehensive understanding of root metabolome and its exudation, it is important to include several factors in the study design. The multifactorial design allows us to study the influence of each factor and their interactions on the metabolite profiles, providing a relatively holistic understanding of the plant exudates compared to studies that focus on effects of single factor on plant root metabolites. In this study, we aimed to obtain a comprehensive view of root and exudate metabolic profiles of *Populus* genotypes with varying lignin content grown under nutrient-rich and nutrient-limited conditions as nutrient availability in the plant's environment is known to be one of the abiotic factors that impacts its metabolism and exudation²⁸. The plant-agar system utilized in this study enabled collection of both roots and exudates for untargeted metabolomic analysis, which is important for comprehensive understanding of metabolites produced and exuded by the plant roots. The root and exudate samples are collected from plants grown for two timepoints 48 hr and 72 hr after transferring to petri dishes to observe the temporal changes in the metabolite profiles for each treatment group. Further, spatially resolving exudates along the

plant root axis is necessary to study the spatial dynamics of the root exudation. In this study, we demonstrated the feasibility of using a cost-effective plant in agar system for spatial sampling of root exudates. At each timepoint, whole root tissue from each plant was collected without spatial sampling due to limited biomass given the age and size of the of the plants. However, the root exudate samples were spatially sampled from distinct spatial locations within the root zone (root base, root tip and regions distant from root) to capture the spatial dynamics of root exudation and its gradient (**Figure 6.1**).

A biphasic extraction method was employed for both root and exudate samples to capture the diverse array of metabolites including both polar and non-polar compounds, untargeted metabolomics analysis was performed to investigate the influence of several experimental factors. Current methods for the root exudate analysis rely mostly on monophasic extractions, which might be limited in its extraction efficiency of the wide range of metabolites including polar (e.g. sugars, amino acids) and non-polar compounds (e.g. fatty acids, lipids) produced and exuded by plant roots. The biphasic extraction method used in this study enhances extraction of metabolites that may be poorly extracted using a monophasic extraction method.

The acquired MS data was processed separately for aqueous and ethyl acetate solvent extractions using compound discoverer (version 3.3.1) and 128787 and 87753 features (mass/retention time pairs) were identified for aqueous and ethyl acetate extractions respectively (**Figure 6.2A**). The resulting data matrices were processed to address redundant peak identifications and poor peak integration by implementing a data processing step that aggregates the peak areas from redundant features that were indistinguishable by annotated molecular weight and annotated delta mass (Da), to obtain non-redundant lists of feature identifications (24096 and 16360 features in polar and non-polar fractions respectively) (**Figure 6.2A**). To ensure the accuracy of the metabolite quantification using our processing step to obtain the non-redundant feature lists, the mean abundance values of sucrose from the refined list was compared with manually curated abundance values using Skyline software and a correlation of 0.69 was observed, demonstrating a reasonable agreement between the in-house implemented processing step and manual integration approaches.

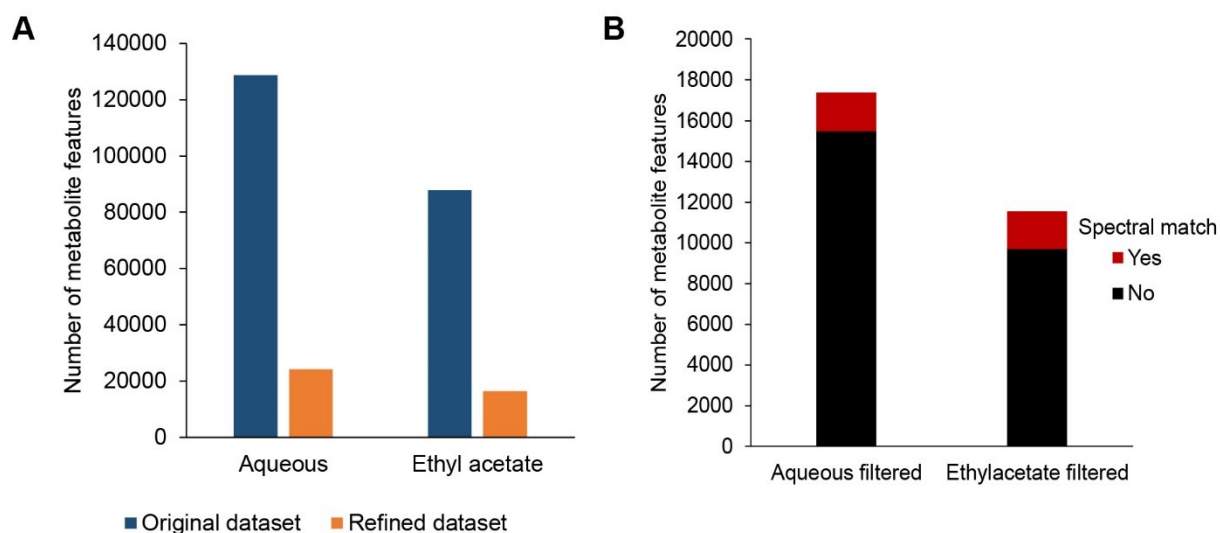


Figure 6.2: Number of metabolite features (A) In original datasets from compound discoverer results and refined non-redundant datasets in both aqueous polar and ethyl acetate organic fractions. (B) Filtered datasets consisting of only the features with mean abundances of experimental treatments greater than in media controls and features with non-missing values in at least 50% of one sample group across the experiment.

To improve the robustness of downstream analysis, the datasets were further filtered to keep only the features with mean abundances of experimental treatments greater than media controls and include only the features with non-missing values in at least 50% of one sample group across the experiment. The filtered lists consist of 17365 features for aqueous extraction and 11547 features in ethyl acetate extraction (**Figure 6.2B**). Of the total metabolite features in these filtered lists, only about 10.9% in aqueous and 15.9% in ethyl acetate extraction have compound annotations based on spectral matching to NIST 2020 and/or GNPS libraries (**Figure 6.2B**). Of the several thousands of metabolite features identified in this study, only about 10-16% has compound annotations and chemical class information. This limited knowledge of metabolite identification represents a significant limitation in understanding the functional roles of the observed metabolites. Although not yet identified, the unannotated metabolite features represent less studied and potentially novel compounds. Therefore, it is crucial to report both annotated and unannotated metabolite features to represent the full complexity of the measured metabolite profiles of roots and exudates. Additionally, reporting unannotated metabolites helps in enriching the metabolomics databases and aid in potential future identification of these compounds which is essential for advancing the field of untargeted metabolomics.

The compounds in the filtered lists were ranked based on their summed abundance values across different experimental conditions in each solvent extraction to identify the most abundant metabolite features and the effect of experimental treatments on their relative abundances. Glutamine, an amino acid commonly found in exudates^{269, 283} was observed to be one of the highly abundant metabolite features in aqueous extraction. Across the aqueous dataset with various sample groups, the relative abundance of glutamine was observed to be dynamic with higher in root tissue samples compared to the exudate samples in both genotypes (**Figure 6.3A**), suggesting its predominant role within the plant. In the ethyl acetate extraction, 9-octadecenamide, a fatty acyl compound known for its role in stress responses in plants was one of the highly abundant metabolites, whose relative abundance might be a result of differences in cell wall chemistries of the two genotypes as its relative abundance is higher in sample groups from BESC-24 genotype compared to BESC-375 (**Figure 6.3B**). Previous studies showed that abundance of 9-Octadecenamide was found to be in higher in exudates of plants in response to

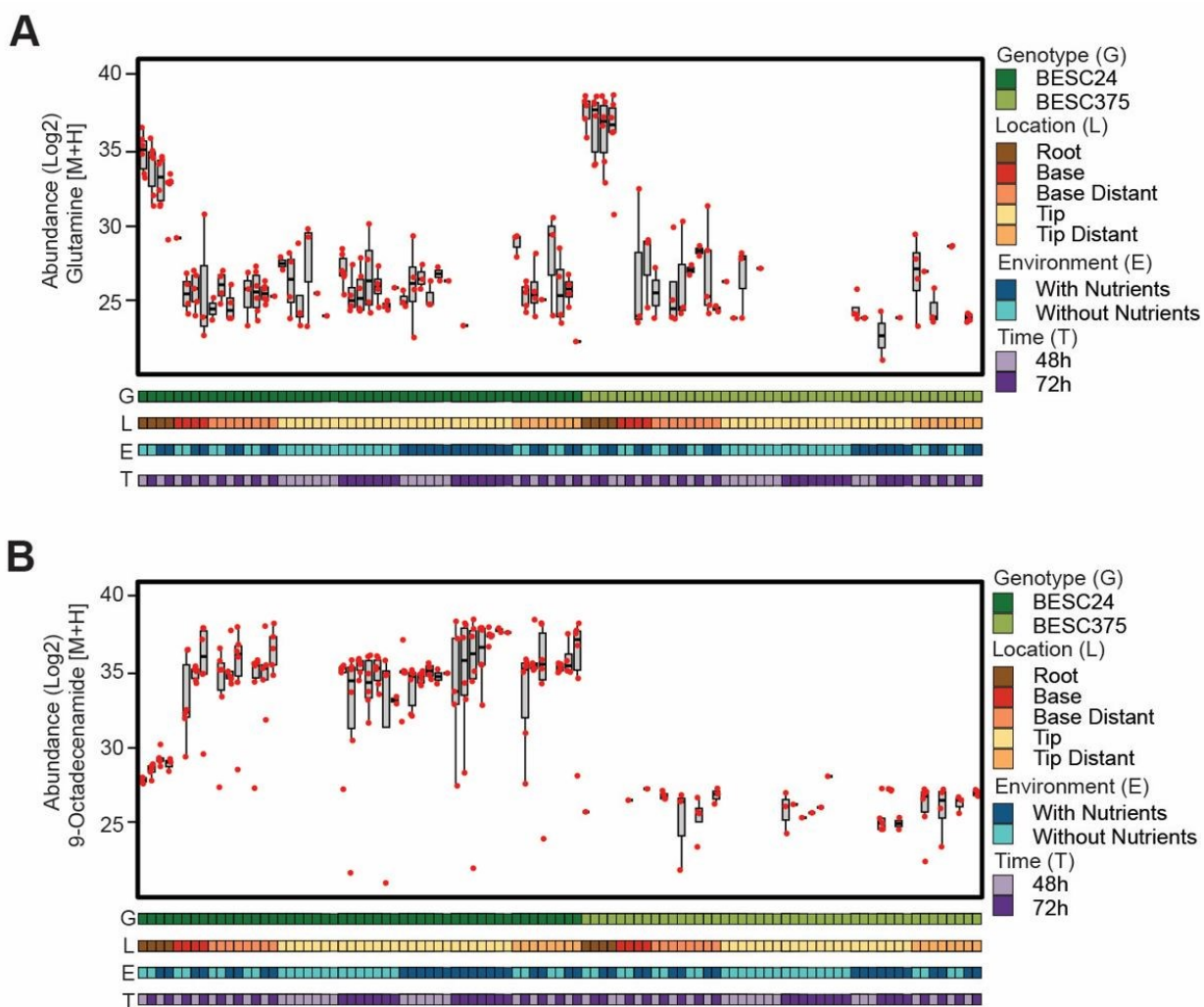


Figure 6.3: Box plots showing Log2 abundance values across experimental conditions for most abundant identified compounds in each extraction solvent (A) Glutamine in aqueous polar fraction (B) 9-Octadecenamide in ethyl acetate organic fraction.

heavy metal stress²⁸⁴. Additionally, 9-Octadecenamide is known for its antimicrobial, anti-inflammatory, and antioxidant activities²⁸⁵.

To understand what chemical classes of metabolites are present and abundant in this untargeted analysis, chemical classification of the annotated compounds (compound that have spectral match to known compound in spectral libraries) was performed using ClassyFire²⁸⁶ and it revealed that metabolite features belonging to superclass organoheterocyclic compounds were most abundant in aqueous extraction contributing to ~54% of total relative abundance of annotated features that have chemical classification information available (**Table 6.1**). In the ethylacetate extraction, lipids and lipid-like molecules were most abundant with ~57% of total relative abundance of annotated features (**Table 6.1**).

Multivariate analysis of metabolite profiles to assess effects of various experimental factors

To assess the influence of various experimental factors including genotype, environment, time, and location (intracellular vs. extracellular) on the overall metabolite profiles measured in this study, permutational multivariate analysis of variance (PERMANOVA) was performed separately on datasets for each extraction solvent (**Table 6.2**). The PERMANOVA results revealed that genotype has a highest significant effect ($p < 0.05$) on the metabolite profiles (27% for aqueous and 30% for ethyl acetate extraction), followed by location (15% for aqueous and 17% for ethyl acetate extraction) in both solvent extractions. Significant effects of interaction between genotype and location (7% for aqueous and 6% for ethyl acetate extraction) was also detected. To visualize the differences between the metabolite profiles of the sample groups (roots and exudates for two genotypes), NMDS plots (Non-Metric Scaling Plot) were used (**Figure 6.4**).

Interestingly, the clustering patterns of NMDS showed a clear separation of four groups comprising of intracellular and extracellular metabolite profiles for each of the two genotypes. Notably, the intracellular and extracellular clusters of BESC-375 aggregated closely compared to BESC-24, suggesting there is less variation between the intracellular and extracellular metabolite profiles of BESC-375. The NMDS plot showed subclusters within the cluster belonging to BESC-24 exudates. As the exudate data consist of several factors including nutrient and time, the subclusters within BESC-24 exudates were examined to identify the underlying factors for this

Table 6.1: Chemical classification results for the annotated metabolite features (A) aqueous polar fraction (B) ethyl acetate organic fraction.

A. Aqueous polar fraction				
Super class	Summed relative abundance	Number of compounds	Relative abundance (%)	Number of compounds (%)
Organoheterocyclic compounds	2.7E+13	409	53.7%	21.8%
Organic acids and derivatives	1.4E+13	263	26.8%	14.0%
Lipids and lipid-like molecules	5.2E+12	418	10.3%	22.3%
Organic oxygen compounds	1.6E+12	192	3.1%	10.3%
Benzenoids	1.3E+12	318	2.7%	17.0%
Phenylpropanoids and polyketides	7.1E+11	171	1.4%	9.1%
Nucleosides, nucleotides, and analogues	4.9E+11	25	1.0%	1.3%
Organic nitrogen compounds	2.9E+11	25	0.6%	1.3%
Organometallic compounds	1.9E+11	2	0.4%	0.1%
Alkaloids and derivatives	2.7E+10	15	0.1%	0.8%
Lignans, neolignans and related compounds	1.2E+10	25	0.0%	1.3%
Organohalogen compounds	3.7E+09	2	0.0%	0.1%
Organic 1,3-dipolar compounds	2.0E+09	4	0.0%	0.2%
Organosulfur compounds	1.1E+09	3	0.0%	0.2%
B. Ethyl acetate organic fraction				
Super class	Summed relative abundance	Number of compounds	Relative abundance (%)	Number of compounds (%)
Lipids and lipid-like molecules	2.0E+13	641	57.1%	35.7%
Benzenoids	5.0E+12	218	14.1%	12.1%
Organoheterocyclic compounds	4.2E+12	289	11.9%	16.1%
Organic oxygen compounds	2.9E+12	160	8.1%	8.9%
Organic acids and derivatives	1.6E+12	141	4.4%	7.9%
Phenylpropanoids and polyketides	1.2E+12	274	3.4%	15.3%
Organic nitrogen compounds	2.4E+11	21	0.7%	1.2%
Organometallic compounds	6.3E+10	3	0.2%	0.2%
Lignans, neolignans and related compounds	2.6E+10	29	0.1%	1.6%
Organohalogen compounds	1.7E+10	1	0.0%	0.1%
Nucleosides, nucleotides, and analogues	1.7E+09	8	0.0%	0.4%
Alkaloids and derivatives	6.1E+08	9	0.0%	0.5%
Organosulfur compounds	2.1E+07	1	0.0%	0.1%

Table 6.2: PERMANOVA results of Bray Curtis distances of metabolomics data (A) aqueous polar fraction (B) ethyl acetate organic fraction.

A. Aqueous polar fraction		
Factor	R²	Pr(>F)
Genotype	0.27	0.001
Location	0.15	0.001
Nutrient	0.08	0.001
Genotype:Location	0.07	0.001
Genotype:Nutrient	0.05	0.001
Timepoint	0.03	0.001
Nutrient:Timepoint	0.02	0.001
Genotype:Timepoint	0.02	0.001
Genotype:Nutrient:Timepoint	0.02	0.001
Genotype:Nutrient:Location	0.01	0.001
Nutrient:Location	0.01	0.001
B. Ethyl acetate organic fraction		
Factor	R²	Pr(>F)
Genotype	0.30	0.001
Location	0.17	0.001
Genotype:Location	0.06	0.001
Nutrient	0.05	0.001
Timepoint	0.03	0.001
Genotype:Timepoint	0.03	0.001
Nutrient:Timepoint	0.03	0.001
Genotype:Nutrient:Timepoint	0.02	0.001
Genotype:Nutrient	0.02	0.001

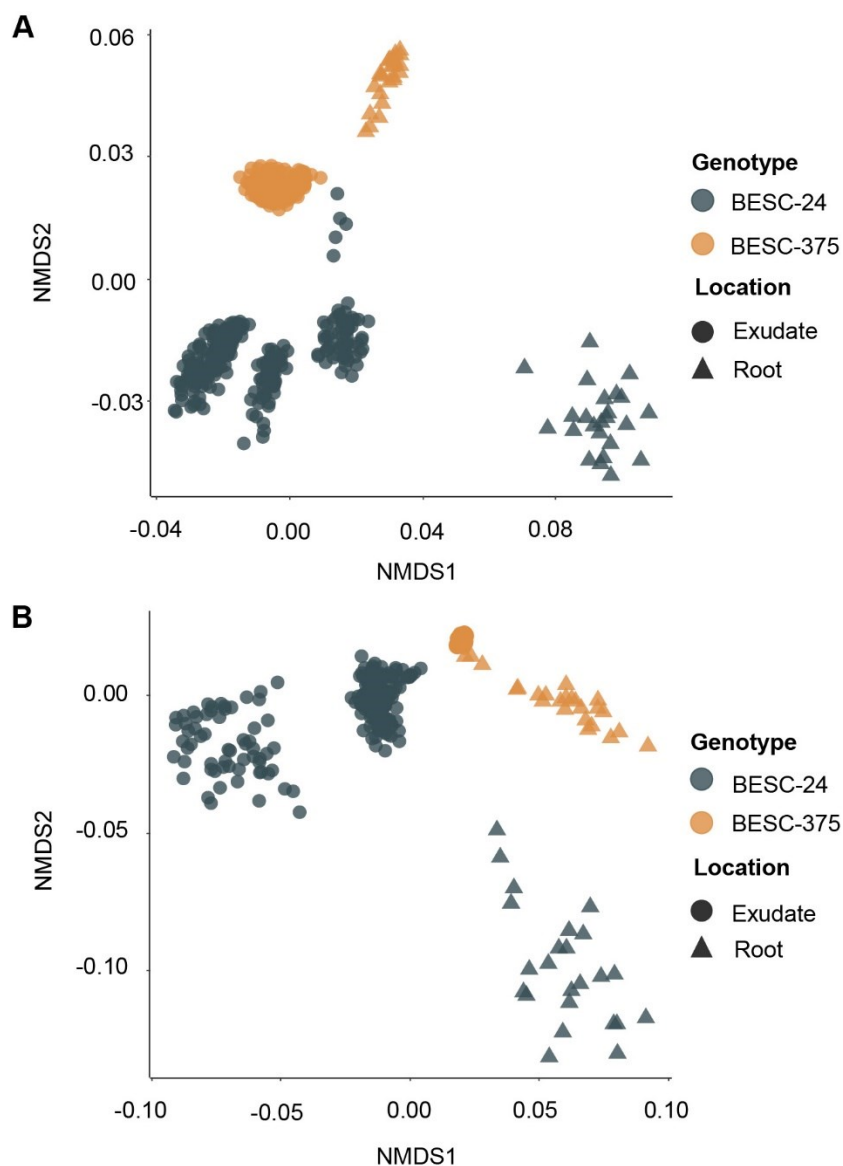


Figure 6.4: Nonmetric multidimensional scaling (NMDS) plots based on Bray Curtis distances of metabolomics data (A) Aqueous polar fraction data showed higher variation is observed in relation to genotype (permutational multivariate analysis of variance [PERMANOVA], $R^2 = 27\%$) followed by the location (15%), (B) Ethyl acetate organic fraction data showed higher variation is observed in relation to genotype (permutational multivariate analysis of variance [PERMANOVA], $R^2 = 30\%$) followed by the location (17%).

separation. The metabolite profiles obtained from two timepoints (48 hr and 72 hr) of nutrient-limited condition clustered together, whereas the metabolite profiles obtained from the two timepoints of nutrient-rich condition clustered separately from each other and nutrient-limited condition. PERMANOVA results showed environment (nutrient condition) also has significant impact on metabolite profiles (8% in aqueous and 5% in ethyl acetate). Overall, these results suggest that there are temporal dynamics driven by the nutrient condition in BESC-24 exudate profiles.

Comparison of metabolite profiles between genotypes

As observed from PERMANOVA analysis, the metabolite profiles of the two genotypes BESC-24 and BESC-375 are significantly different from each other. It is difficult to quantitatively assess the differences between the metabolite profiles of these two genotypes as there are differences in their cell wall chemistries which could influence the metabolite profiles.

Additionally, we observed differences in the root physiology but lack advanced image-based phenotyping data to accurately normalize the data to account for these differences on metabolite profiles. The root system architecture which includes root length, density, surface area, root hair development etc. influences how the roots interact with its environment influencing nutrient and water uptake, exudation, which in turn affects the metabolome composition and their concentrations. Therefore, only qualitative assessment of the metabolite profiles obtained from the two genotypes was conducted. In aqueous extraction consisting of 17365 features, a total of 15349 features were identified for BESC-24 genotype and 10799 features were identified for BESC-375 genotype (**Figure 6.5A**). In the ethyl acetate extraction, a total of 11547 features were identified including 10426 features for BESC-24 and 6822 features for BESC-375 (**Figure 6.5B**). These results indicate that BESC-24 has a more complex metabolic profile compared to BESC-375. Upset plots were generated to identify the common and shared metabolite features between the genotypes and showed large overlaps between the metabolite profiles of these genotypes, 50.6% of total identified features were shared in polar extraction and 49.4% in non-polar extraction. A significantly higher proportion of metabolite features identified were unique to BESC-24 genotype (37.8% and 40.9% in polar and non-polar extractions respectively). There were several metabolite features that were found to be unique to BESC-375 (11.6% and 9.7% in

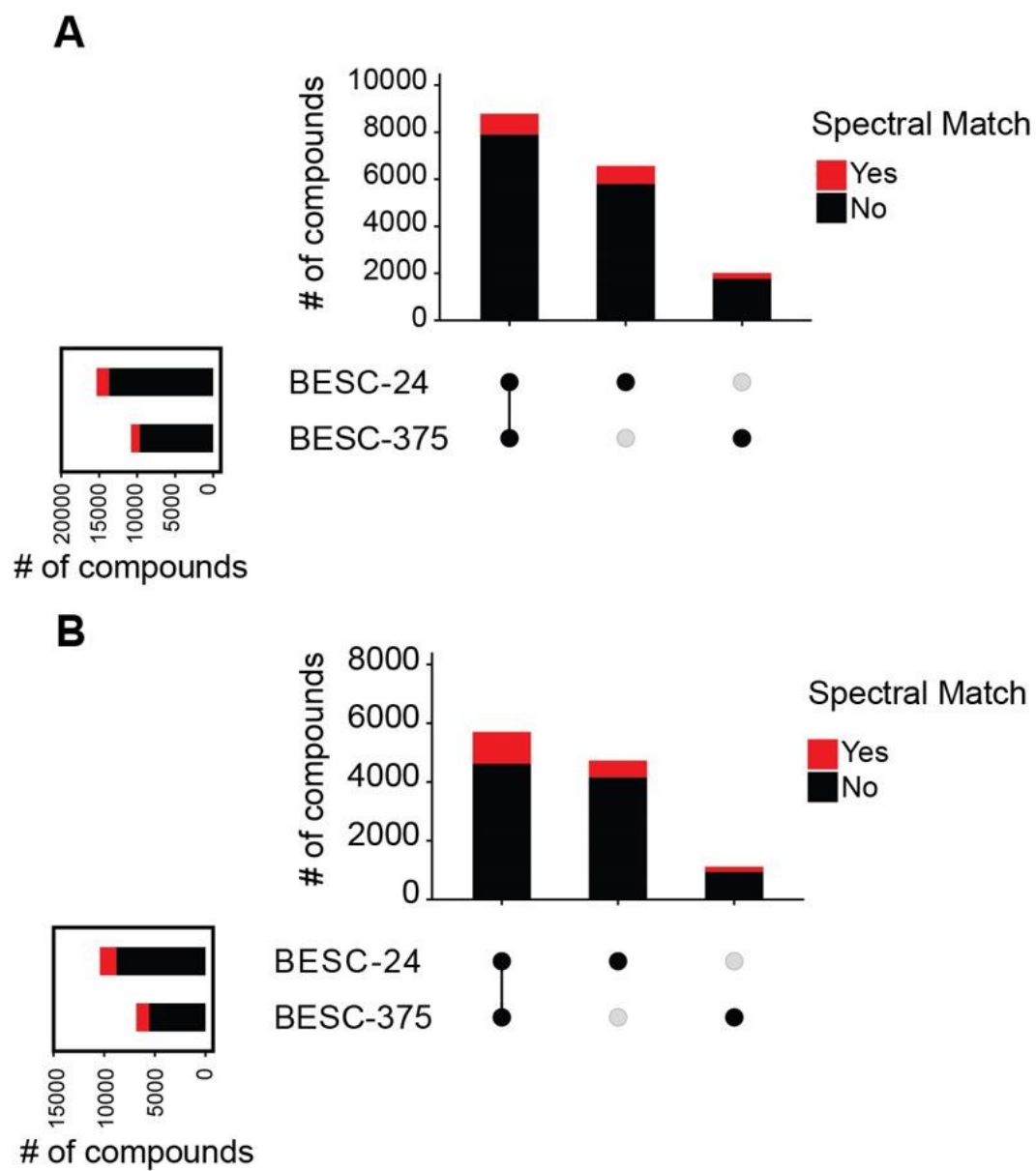


Figure 6.5: Upset plots showing overlap of compounds between the genotypes (A) Aqueous polar fraction, (B) Ethyl acetate organic fraction.

polar and non-polar extractions respectively) (**Figure 6.5A and 6.5B**). The higher number of features identified in BESC-24 correlated with the higher average number of primary roots and dry biomass when compared to BESC-375 (**Figure 6.6**). It is essential to consider root system architecture in the future studies of root and exudate metabolomics.

Previously, using the unsupervised NMDS method for multivariate data analysis, the separation between the two genotypes was observed. To identify which metabolites, contribute most to that differentiation between the genotypes and drive the separation, Partial Least Squares Discriminant Analysis (PLS-DA), a supervised analysis was performed on the data subsets for each extraction separately using MetaboAnalyst (Version 6.0). PLS-DA provides Variable Importance in Projection (VIP) scores to rank the metabolite features by their contribution to the separation. Based on the VIP scores, the key metabolites that differentiated between the genotypes were identified. Of the several metabolites identified, there were only a few metabolites that are annotated and have chemical classification information available. Among these, the metabolite features that have good spectral match with the search databases were further selected to understand the trends in their abundances across the experimental conditions.

For example, loliolide in ethyl acetate extraction was found to be highly abundant in several sample groups of BESC-24 compared to BESC-375 (**Figure 6.7**). Loliolide is a compound belonging to chemical class benzofurans. Loliolide is found in many plant metabolites known for its role in defense response by activating jasmonate-related pathways^{287, 288}. In wheat and rice, loliolide is produced and secreted by roots in response to biotic and abiotic stressors²⁸⁷. In this study, the relative abundance of loliolide did not change significantly in root samples between nutrient-rich and nutrient-limited conditions in BESC-24. However, varying trends were observed across other experimental conditions displaying dynamics due to nutrient condition and sampling timepoints.

Comparison of intracellular and extracellular metabolite profiles

Location was the second highest significant factor that influenced the metabolite profiles based on PERMANOVA. To identify and compare the diversity of metabolites present in root intracellular and exudate samples, a qualitative analysis between the two locations was performed for each genotype separately. The quantitative comparison of the metabolites between

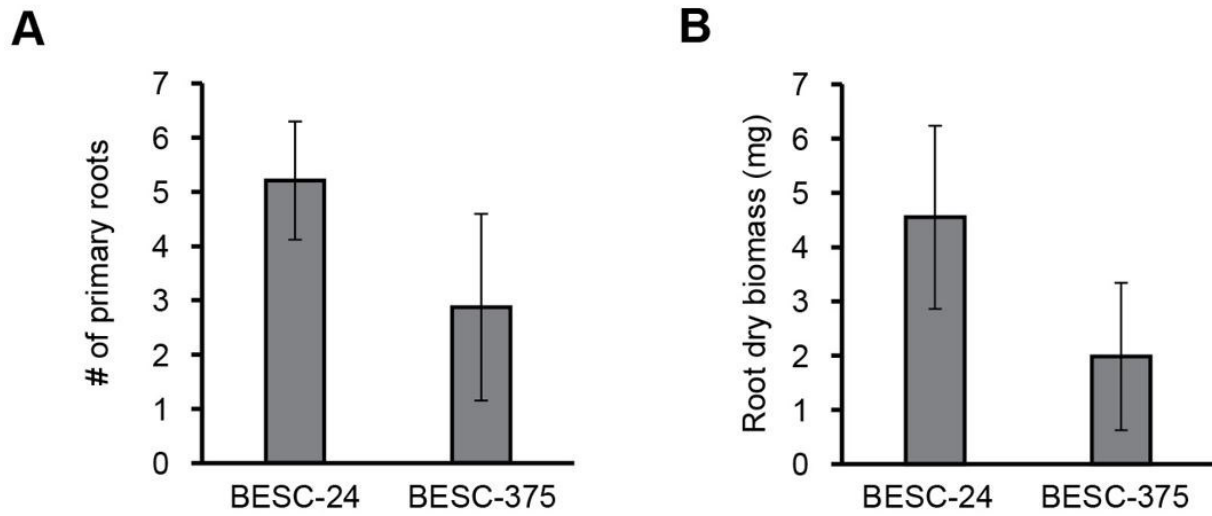


Figure 6.6: Root trait data (A) Average number of primary roots for each genotype (n=24), error bars represent standard deviation (B) Average root dry biomass for each genotype (n=24), error bars represent standard deviation.

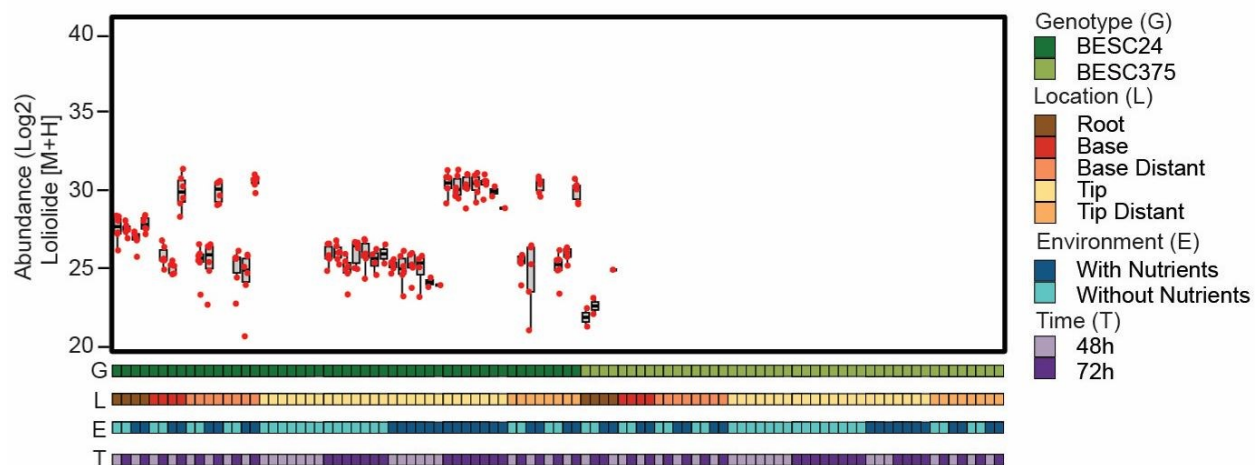


Figure 6.7: Box plots showing Log₂ abundance values across experimental conditions for Loliolide, an important metabolite feature in ethyl acetate organic fraction differentiating between the genotypes based on PLS-DA analysis.

roots and exudates was avoided as the root metabolite profiles were obtained from whole root system per plant, whereas for the exudates, the samples were collected from around distinct root regions that may cause variability in metabolite concentrations due to spatial heterogeneity and metabolite diffusion. It is challenging to account for these variations to conduct quantitative comparison of root intracellular and extracellular metabolite profiles. The data subsets for each genotype were extracted from aqueous and ethyl acetate extraction datasets and compared solely based on presence and absence of metabolites (**Figure 6.8**). For BESC-24, the number of features detected in exudates were higher than in roots for in both aqueous and ethyl acetate extractions. Upset plots showed a higher proportion of detected features (55.2% in aqueous and 53.7% in ethyl acetate) were present in both roots and exudates of BESC-24, indicating that these metabolites might represent a core metabolome that is crucial for BESC-24 physiological functions. For BESC-375 genotype, a similar trend was observed in aqueous extraction with higher number of features in exudates compared to roots, however, in ethyl acetate extraction a higher number of features were detected in roots compared to exudates. Increased measurement depth of metabolite features in exudates compared to root tissue can be due to multiple samples collected for exudates (root base, root tip, distant regions) compared to single sample of root tissue for each plant. Multiple sampling of exudates likely influences the comprehensiveness of the metabolite profile as it increases the chance of measuring certain underrepresented metabolite features. Upset plots show that 31.1% in aqueous extraction and 41.9% of detected features in ethyl acetate extraction were present in both roots and exudates of BESC-375.

As described in above result section focused on genotype comparison, PLS-DA analysis was performed for aqueous and ethyl acetate datasets for each genotype to identify metabolites that differentiate between the intracellular and extracellular metabolite profiles. Based on the VIP scores, the key metabolites that differentiated between the roots and exudates were identified. Of the several metabolites identified, there were only a few metabolites that are annotated and have chemical classification information available. Among these, the metabolite features that have good spectral match with the search databases were further selected to understand the trends in their abundances across the experimental conditions.

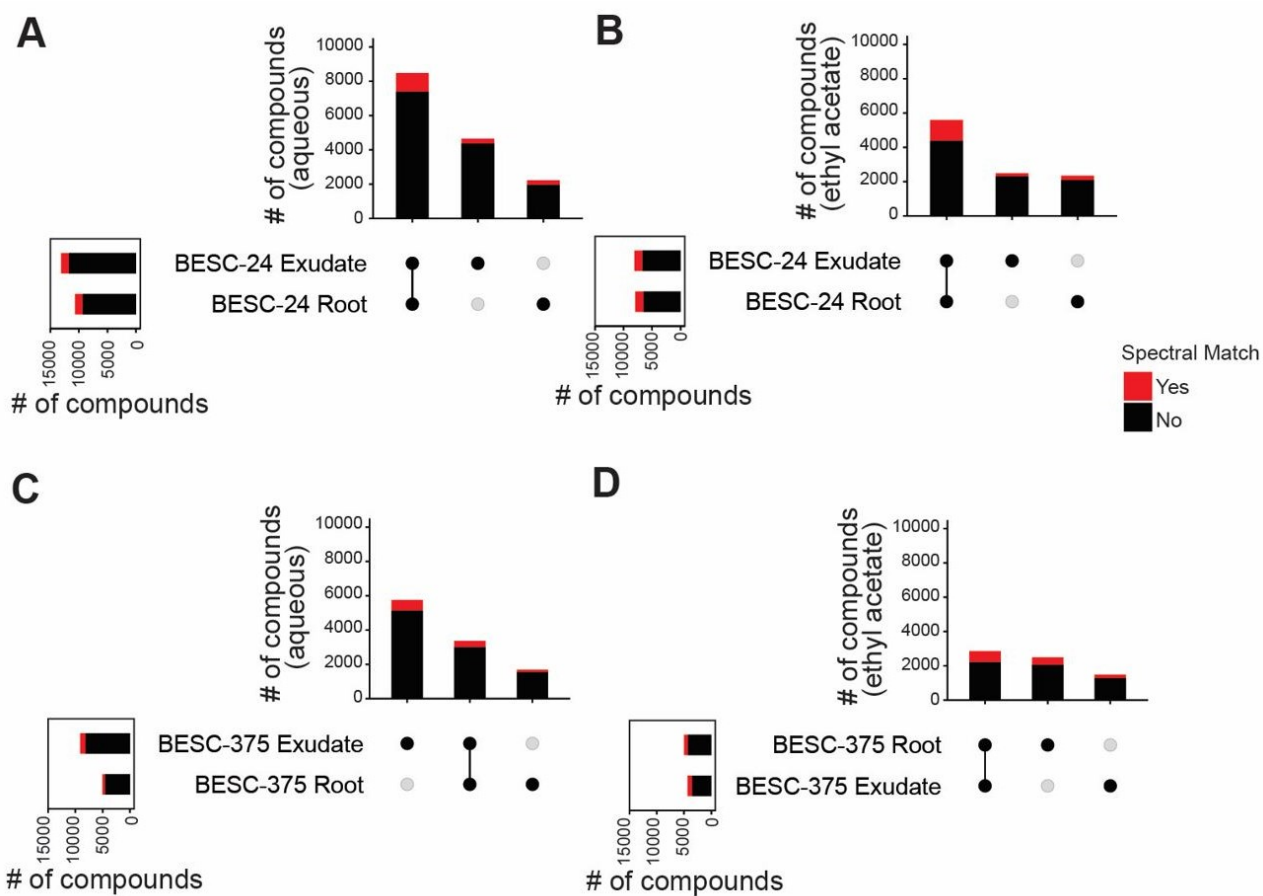


Figure 6.8: Upset plots showing overlap of compounds between the root and exudates samples (A) BESC-24 aqueous polar fraction (B) BESC-24 ethyl acetate organic fraction (C) BESC-375 aqueous polar fraction (D) BESC-375 ethyl acetate organic fraction.

Swainsonine belongs to the chemical class indolizidines, and it was found to be more abundant in exudates compared to the roots of both genotypes (**Figure 6.9A**). Swainsonine is present in a several plant species and known for its toxic roles²⁸⁹. In this study, its presence in exudates indicates its role in defense response mechanisms in both genotypes. Another metabolite, isorhamnetin 3-O-glucoside displayed an opposite trend where it is highly abundant in root tissues compared to exudates, it is a flavonoid, which is known to be widely present in plants and involved in stress response^{290, 291} (**Figure 6.9B**). Other phenolic glycosides salicortin and populin, which are particularly found in willows and known for their roles in stress and defense responses^{292, 293} were identified in this study and have higher abundance in roots compared to exudates in both genotypes. Further investigation is needed to understand the biological relevance of elevated levels of storage of these metabolites within the roots.

Assessing influence of nutrient condition on root and exudate profiles of BESC-24

Based on the clustering patterns observed in the NMDS plots in Figure 6.5, metabolite data belonging to BESC-24 genotype roots and exudates were extracted separately from aqueous and ethyl acetate datasets. The metabolite profiles were compared between the nutrient-rich and nutrient-limited conditions based on presence and absence of metabolites for roots and exudates sample groups to identify nutrient condition specific metabolite features. In all these comparisons, the highest proportion of metabolite features are shared between two nutrient conditions, However, there are about 10% that are unique to either of the two nutrient conditions. These metabolite features that are exclusively identified in samples grown under a particular nutrient condition, suggests the pivotal role played by nutrient availability in altering the metabolite profiles. This comparative analysis highlights the metabolite specificity to nutrient availability or limitation in the plant environment and their potential biological significance. Further, metabolite level analysis revealed the temporal dynamics across the nutrient conditions.

In the root samples of BESC-24, the relative abundance of a dipeptide Asp-Trp identified in aqueous extraction was observed to be higher when grown in nutrient-limited condition compared to nutrient-rich condition at 48 hr (**Figure 6.10A**). Interestingly, its relative abundance decreased at 72 hr in nutrient-limited condition, but it increased in nutrient-rich condition. Another metabolite example that displayed significant variations in abundance between nutrient

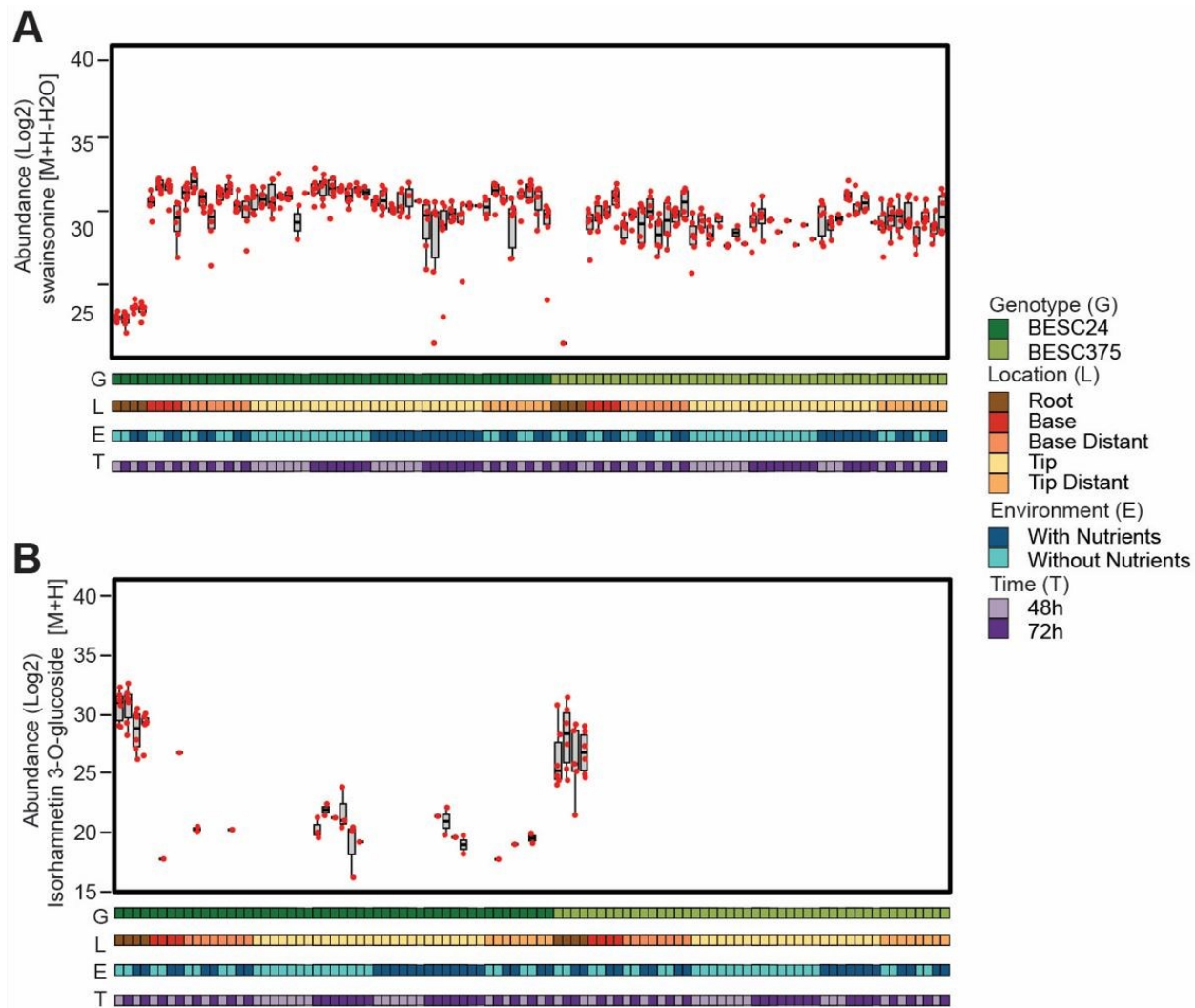


Figure 6.9: Box plots showing Log2 abundance values across experimental conditions for important metabolite features differentiating between root and exudate samples based on PLS-DA analysis (A) Swainsonine in aqueous polar fraction (B) isorhamnetin 3-O-glucoside in ethyl acetate organic fraction.

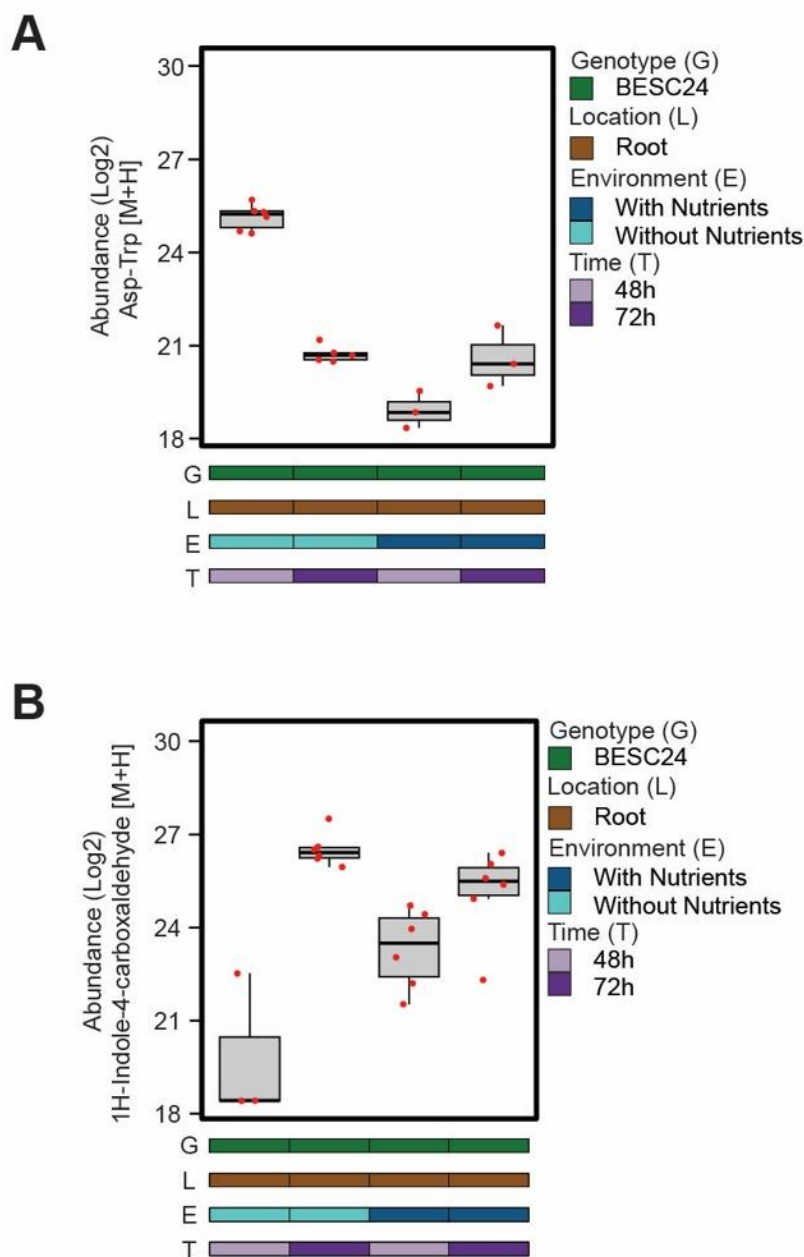


Figure 6.10: Box plots showing Log2 abundance values for metabolites identified in BESC-24 genotype root samples and displayed changes in abundance in response to nutrient condition-timepoint interactions (A) Asp-Trp in aqueous polar fraction (B) 1H-indole-4-carboxaldehyde in ethyl acetate organic fraction.

conditions is 1H-indole-4-carboxaldehyde, where the relative abundance was observed to be higher when grown in nutrient-rich condition compared to nutrient-limited condition at 48 hr (**Figure 6.10B**). For this metabolite, the relative abundance was found to be increased at 72 hr compared to 48 hr timepoint in both nutrient-limited and nutrient-rich conditions.

In BESC-24 exudate samples, S-(5'-Adenosyl)-L-homocysteine, a nucleoside compound which is a biosynthetic precursor to homocysteine was identified in aqueous extraction and displayed changes in abundances between nutrient conditions and sampling timepoints, indicating the dynamics in metabolic pathways influenced by nutrients and time (**Figure 6.11A**). Another metabolite in BESC-24 exudates, azelaic acid which is a fatty acyl compound displayed abundances changes across nutrient conditions and timepoints, with higher abundance in nutrient-rich condition compared to nutrient-limited condition. Additionally, azelaic acid displayed higher abundance values at 72 hr compared to 48 hr timepoint (**Figure 6.11B**). In plants, azelaic acid plays a role in defense responses and increased accumulation of azelaic acid occurs during pathogen infection²⁹⁴ and it is known to be present in root exudates of tomato and a key factor in process of systemically induced root exudation of metabolites (SIREM)²⁹⁵. In this study, the increased levels of azelaic acid in nutrient-rich condition with time could be indicating its role in signaling or defense mechanisms.

Conclusions

In this study, a plant in agar experimental platform was used for collecting root metabolite profiles and the spatially resolved root exudates. Using untargeted metabolomics approach, the effects of various experimental factors including differences in plant lignin content, nutrient availability and sampling timepoints on root metabolite profiles of *Populus trichocarpa* were assessed. This study provided a comprehensive understanding of the differences in the root and exudate profiles between the two *Populus* sp. genotypes with varying lignin contents grown under different nutrient conditions. Several metabolites displaying distinct trends influenced by the plant genotype and nutrient condition were identified. The differences in the abundances of specific metabolites across these conditions underscores the complexity of the plant root metabolic processes and exudation patterns. The findings of this study suggest that genotype-

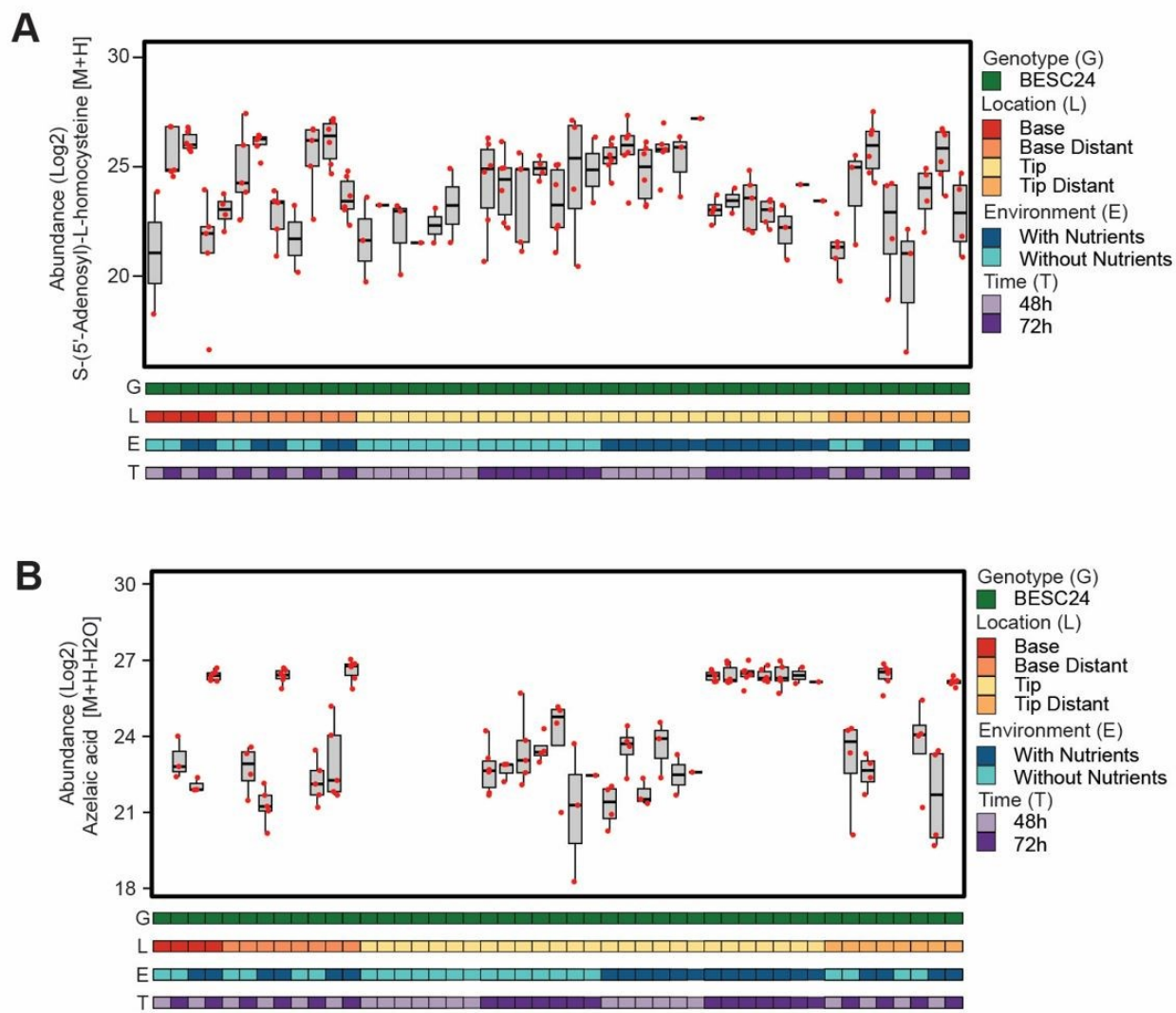


Figure 6.11: Box plots showing Log2 abundance values for metabolites identified in BESC-24 genotype exudate samples and displayed changes in abundance in response to nutrient condition-timepoint interactions (A) S-(5'-Adenosyl)-L-homocysteine in aqueous polar fraction (B) azelaic acid in ethyl acetate organic fraction.

specific conditions, such as cell wall composition, impacts the root and exudate metabolic profiles significantly, while nutrient conditions and time further modulate the abundances of these key metabolites. However, further interrogation of these metabolites is necessary in understanding their functional relevance in *Populus* species and nutrient availability. Further, the temporal dynamics observed in this study highlight the variability in metabolite profiles at various timepoints and the need for temporal studies when studying plant root exudates. This study also highlights the importance of untargeted metabolomics in an unbiased manner in identifying a broad range of metabolites produced by plants to capture the natural complexity of root exudates.

CHAPTER 7: CONCLUSIONS AND FUTURE PERSPECTIVES

Conclusions

Plant-microbe interactions play a pivotal role in plant health and resilience to environmental stresses and therefore impacts crop yields and sustainable agriculture. The complex and dynamic associations between the plant and microbes are influenced by various factors such as host genetics, environmental factors including biotic and abiotic conditions. It is crucial to study the interactions between the plant and its associated microbial communities to gain insights into how these interactions help the plants to adapt and thrive under various environmental conditions. Understanding how the microbial communities are assembled and the factors and interactions that shape them are crucial in plant-microbiome studies but are largely unknown. However, studying these in situ is extremely challenging due to the complexity and spatio-temporal dynamics. Reductionist approaches and synthetic microbial communities (SynComs) are powerful tools in advancing plant microbiome research as they enable dissecting specific interactions and mechanisms under controlled environments. This knowledge can be invaluable in developing sustainable agricultural practices. Further, these studies provide insights into complex relationships between the plants and microbes and contribute to improved understanding of the dynamics in the plant ecosystem, which are vital for stress tolerance, disease resistance and nutrient cycling in changing environmental conditions.

Omics technologies such as genomics, transcriptomics, metaproteomics, and metabolomics are used in studying the plant associated microbiome and each provide unique insights about their composition, functions and interactions. These studies allow us to gain comprehensive understanding of the interactions between plants and their associated microbiome by providing detailed insights into how plants and microbes respond to each other and the influence of microbes on plant growth and stress tolerance mechanisms at a molecular level. Among several omics techniques, metaproteomics enables us to characterize the proteins expressed in the microbial communities, which provides information about the microbial functions, interactions and metabolic pathways, which can be used to determine their responses to available nutrients and/or stressors in their environment. On the other hand, metabolomics allows us to identify the

metabolites produced by plant and associated microbes, which provides insights to the chemical signals produced and released into the environment that are crucial for establishing the interactions between them.

The primary focus of this dissertation was to advance our understanding of the plant-microbiome interactions specifically of *Populus* spp. and its associated rhizosphere microbiome by using a reductionist approach and application of metaproteomics and metabolomics methodologies. In this dissertation research, simplified experimental approaches using SynCom grown under laboratory-controlled conditions was initially employed to gain fundamental understanding of formation of stable microbial assemblies under different nutrient conditions and the inter-species interactions in these communities in liquid media. The complexity of the experimental system was gradually increased by adding the plant host to study the microbial community assemblies and microbial functions in response to presence of plant under controlled laboratory settings. Further, the plant root metabolic profiles were studied to gain insights into the resources produced and released by plant into its surrounding environment.

By using a 10-member SynCom consisting of bacterial species that taxonomically and functionally represent *Populus* rhizosphere and performing metaproteomic analysis, distinct microbial community structures and functions of individual bacterial strains that facilitated their presence and abundance in the community were identified as a result of inter-species interactions under varying nutrient conditions. Although this study was performed under simplified laboratory conditions without presence of plant, it has provided fundamental insights into the microbiome assembly of *Populus* associated bacterial species. This study has led to a follow-up metaproteomic study to understand the effect of initial inoculum ratios of the members of the constructed microbial community on the final stable community structure which is described in chapter 3 of this dissertation. Three bacterial species *Pseudomonas* sp. GM17, *Pantoea* sp. YR343 and *Sphingobium* sp. AP49 were selected based on their presence in the final community of the 10-member study and communities with varying starting ratios of the three bacterial strains were constructed. The metaproteomic analysis at various timepoints showed initial inoculum ratios did not impact which bacterial species dominated the community, however the temporal functional dynamics of the individual bacterial strains led to the final community composition. Overall, this research presented in chapter 3 was crucial in obtaining foundational

knowledge of microbial community assemblies and revealed key microbial players, which needs to be tested in more complex environments in future to confirm their relevance in the rhizosphere microbial communities.

Building up on this research, next we wanted to study the microbiome composition and functions in presence of plant root environment using plant-agar system. Additionally, we aimed to obtain spatially resolved metaproteomes as the spatial dynamics in rhizosphere can impact microbiome composition but recognized the lack of methodologies that facilitate these studies. In chapter 4 of this dissertation, a novel peptide extraction method from agar using a plant-agar microbiome system is described. The study revealed distinct community compositions at distinct root regions likely influenced by the plant root chemical environment. This experimental platform and method enabled spatial sampling of proteomes of plant-root associated bacteria. This methodological advancement can be easily applied to other plant-microbiome research studies and scaled up to larger petri dishes or other plant growing containers such as magenta boxes for larger plants. The method enabling spatial metaproteomics in agar has potential to be applied beyond rhizosphere microbiome research as spatial metaproteomics is of high relevance in other research areas such as understanding spatial organization of microbial communities in biofilms.

Fundamental understanding of the microbial functions to plant root chemicals is highly relevant for advancing our knowledge of plant-microbe interactions. However, it is challenging to dissect the specific effects of plant root chemical environment on the microbiome composition and microbial functions due to the presence of direct interactions between the plant root and microbes. Therefore, methods that enable the study of microbial communities and their responses to root specific chemical signals hold significant potential for advancing rhizosphere microbiome research. The experimental approach using a filter-based system described in chapter 5 of this dissertation allowed us to isolate the microbial responses due to chemical signals specifically from the responses due to direct physical interactions with plant roots. Additionally, studying responses of bacterial strains grown in isolation and in mixed community allowed us to identify responses that are plant root environment specific and not driven by community dynamics. The study revealed the functions employed by specific bacterial species to adapt and thrive in the plant root environment.

Root exudates serve as primary communication source between the plant and surrounding microbiome and hence it is a crucial part of plant-microbiome research. The plant metabolic processes and exudation patterns are influenced by plant genetic and environmental stressors. Chapter 6 of this dissertation provides an in-depth investigation of the dynamics of the root metabolome and root exudate profiles of two *Populus* genotypes with varying cell wall chemistries grown under varying nutrient availability conditions. The untargeted metabolomics results provided deeper understanding of the complex chemical diversity of *Populus* root exudates and their dynamics. To my knowledge, this is the first comprehensive study to characterize *Populus* root metabolome and exudates at a spatial resolution of root base and root tip. Future studies with gradual increments in complexity by adding microbial species will enable the identification of specific exudates that attract or repel certain microbes and how community dynamics change with varying root exudate profiles of *Populus* species. The experimental approach and metabolite extraction method described in chapter 6 are applicable for future studies in rhizosphere research to study effects of various environmental factors on plant root exudates.

Overall, this dissertation research has contributed to significant improvement of our understanding of various processes occurring in the *Populus* rhizosphere including the microbiome community assemblies and microbial functions with and without plant chemical environment and comprehensive understanding of dynamics in root intracellular and extracellular metabolic profiles of *Populus* sp. Additionally technical advancements demonstrated in this dissertation are valuable resources for other plant microbiome studies. This dissertation research laid a foundation to future studies that can build up on these findings by increasing complexity of the experimental system towards a more realistic plant environment in nature.

Remaining challenges, limitations and future perspectives

The complexity and multifaceted interactions between plant and its associated microbiome is increasingly recognized, with ongoing research focusing on unraveling these intricate dynamics. Despite the advances in our understanding, there are still, several challenges and limitations in

studying the plant-microbe interactions. One of the major challenges in studying plant-microbiome is the complexity and therefore simplified experiments with defined microbial consortia are performed to gain fundamental understanding of the microbial functions under defined conditions. Currently, majority of the plant microbiome studies include only bacterial species, with a very few studies also including fungi. Besides bacteria, fungi are also found abundantly in plant microbiome and known for their crucial roles in plant microbiome and their interactions with plants. More studies that include fungal species in the microbiome community experiments are necessary as these studies will provide better understanding of the inter-kingdom interactions that occur in the plant microbiome. As we try to gain holistic understanding of functions of plant microbial communities, there is a need to study higher complexity systems including wide range of microbial taxa and multiple experimental factors, which is essential for understanding the broad spectrum of interactions and functions within the microbiome. Metaproteomic analysis to study the microbial functions can be challenging for the higher complexity samples, due to highly variable abundances of microbial species in a microbiome containing both highly abundant species and extremely low abundant species. The proteins produced by these microbial species can also vary by several orders of magnitude in abundances, this leads to detection of highly abundant proteins and underrepresentation of low abundant but probably biologically important proteins. Certain microorganisms, even when present in low abundance, can have significant effects on the community structure and function.

With the current advancements in the mass spectrometry instrumentations such as the Orbitrap Astral, it is an exciting time to study higher complexity samples. The Orbitrap Astral mass spectrometer contains three mass analyzers, a quadrupole for high selectivity and ion transmission, an orbitrap for high resolution and high dynamic range measurements, and the Astral for high sensitivity and scan rate. With these improved figures of merit, deep proteome coverages within samples with high dynamic range can be achieved with fast and high throughput measurements. Using such advanced instrumentation, chances of detecting low-abundant microbial species that might play crucial roles in microbiome can be increased. Although higher complexity samples and advanced instrumentation offer numerous advantages, they also result in some challenges. These studies will result in highly complex datasets and analyzing them requires advanced computational tools and bioinformatics expertise to process

and interpret these large datasets and therefore there is need to improve the bioinformatics workflows associated with metaproteomic analysis.

Another major limitation in plant-microbiome studies is related to the untargeted metabolomics studies. Mass spectrometry based untargeted metabolomics identifies thousands of compounds, but only a small portion of them are previously known and have chemical information associated with them, which makes only a small portion of the measured data useful to draw any biological interpretations. While the large portion of measured compounds are unannotated, they may consist of novel and biologically relevant compounds. However, most studies do not report them due to the lack of annotations. This is a major limitation in the field of untargeted metabolomics and needs significant improvements in expanding the spectral libraries by characterizing more compounds. Another possible approach is for the scientific community to do a better job of data sharing by creating databases of these compounds in GNPS like platforms and make the data available for collaborations that can enhance identification of novel metabolites and make advancements in metabolomics research.

Given the complexity of the plant-microbiome system, integrating multiple omics studies is a powerful approach in studying the complex interactions between the plant and microbes as they can provide a holistic understanding by combining data from different molecular layers such as DNA, transcripts, proteins and metabolites. For example, integrating the root exudate metabolomics results with metaproteomics results of associated microbial community can provide a systems-level understanding of how root exudates influence the microbial communities. This comprehensive analysis helps in linking specific exudates to microbial functions and plant-microbe interactions. However, integrating multiple omics approaches has several challenges including data qualities can be variable between different omics approaches, some datasets may have higher coverage than others and this can lead to difficulties in integrating them for a meaningful biological interpretation. Additionally, biological interpretation of data from integrating multi-omics approaches requires extensive biological knowledge to understand how changes at genomic, transcriptomic, proteomic, or metabolomic levels are related to each other in the context of plant-microbe interactions.

Another challenge in plant-microbiome research currently is although it is known that the microbiome composition and functions are spatially and temporally dynamic, it is extremely difficult to study these dynamics in soil due to the complexity of the microbiomes, soil properties and methodological limitations. Future efforts in plant microbiome research should focus on obtaining spatially resolved omics measurements from soil analog based systems that closely mimic soil for a more realistic understanding of interactions in rhizosphere. Additionally, there is a need to analyze root system architecture including root length, branching patterns, root hair density in root exudate studies as the root system architecture is critical in determining how and where exudates are released into its environment. Plants often alter root growth and exudation patterns in response to environmental conditions including biotic and abiotic stressors. Combined root exudate analysis with advanced root phenotyping data enables us to correlate specific root morphologies with adaptive exudation strategies but are currently lacking in the field. Integrating the plant root phenotype data with metaproteomics and metabolomics data of root exudates can significantly enhance our understanding of how the structure and organization of the plant root system can influence the microbiome composition by shaping the root exudate profiles. This helps identify how different root traits support microbial interactions. Root growth is a dynamic process, with different phenotypes appearing at various growth stages. By analyzing temporal dynamics in exudate profiles alongside temporal changes in root architecture, we can track how exudate release changes over time in relation to the root development. This dynamic approach reveals the critical time points at which certain exudates are released and may be crucial in microbial recruitment or nutrient mobilization. Such integrated studies bridge the gap between plant morphology and functional ecology and offers a holistic approach to enhance plant resilience and productivity.

Over the past few years, major advancements in the field of plant-microbiome research have been made using omics technologies, that provided unprecedented insights into the complex interplay between the plant host and its associated microbiome. These advancements in plant-microbiome research deepened our fundamental understanding of plant rhizosphere system and holds promise for developing strategies to improve plant health, productivity and resilience for sustainable agriculture.

REFERENCES

- (1) Schirawski, J.; Perlin, M. H. Plant(-)Microbe Interaction 2017-The Good, the Bad and the Diverse. *Int J Mol Sci* 2018, *19* (5). DOI: 10.3390/ijms19051374 From NLM Medline.
- (2) van der Heijden, M. G.; de Bruin, S.; Luckerhoff, L.; van Logtestijn, R. S.; Schlaeppi, K. A widespread plant-fungal-bacterial symbiosis promotes plant biodiversity, plant nutrition and seedling recruitment. *ISME J* 2016, *10* (2), 389-399. DOI: 10.1038/ismej.2015.120 From NLM Medline.
- (3) Rodriguez, R. J.; Henson, J.; Van Volkenburgh, E.; Hoy, M.; Wright, L.; Beckwith, F.; Kim, Y. O.; Redman, R. S. Stress tolerance in plants via habitat-adapted symbiosis. *Isme Journal* 2008, *2* (4), 404-416. DOI: 10.1038/ismej.2007.106.
- (4) Enebe, M. C.; Babalola, O. O. The impact of microbes in the orchestration of plants' resistance to biotic stress: a disease management approach. *Appl Microbiol Biotechnol* 2019, *103* (1), 9-25. DOI: 10.1007/s00253-018-9433-3.
- (5) Vandenkoornhuyse, P.; Quaiser, A.; Duhamel, M.; Le Van, A.; Dufresne, A. The importance of the microbiome of the plant holobiont. *New Phytol* 2015, *206* (4), 1196-1206. DOI: 10.1111/nph.13312 From NLM Medline.
- (6) Sanchez-Cañizares, C.; Jorrín, B.; Poole, P. S.; Tkacz, A. Understanding the holobiont: the interdependence of plants and their microbiome. *Current opinion in microbiology* 2017, *38*, 188-196. DOI: 10.1016/j.mib.2017.07.001.
- (7) Vishwakarma, K.; Kumar, N.; Shandilya, C.; Mohapatra, S.; Bhayana, S.; Varma, A. Revisiting Plant-Microbe Interactions and Microbial Consortia Application for Enhancing Sustainable Agriculture: A Review. *Front Microbiol* 2020, *11*, 560406. DOI: 10.3389/fmicb.2020.560406 From NLM PubMed-not-MEDLINE.
- (8) Phour, M.; Sehwat, A.; Sindhu, S. S.; Glick, B. R. Interkingdom signaling in plant-rhizomicrobiome interactions for sustainable agriculture. *Microbiol Res* 2020, *241*, 126589. DOI: 10.1016/j.micres.2020.126589 From NLM Medline.
- (9) Harris, J. M.; Balint-Kurti, P.; Bede, J. C.; Day, B.; Gold, S.; Goss, E. M.; Grenville-Briggs, L. J.; Jones, K. M.; Wang, A. M.; Wang, Y. C.; et al. What are the Top 10 Unanswered Questions in Molecular Plant-Microbe Interactions? *Mol Plant Microbe In* 2020, *33* (12), 1354-1365. DOI: 10.1094/Mpmi-08-20-0229-Cr.
- (10) Matyssek, R.; Lüttge, U.; zu Castell, W. Evolution of Holobiont-Like Systems: From Individual to Composed Ecological and Global Units. In *Progress in Botany Vol. 83*, Lüttge, U., Cánovas, F. M., Risueño, M.-C., Leuschner, C., Pretzsch, H. Eds.; Springer International Publishing, 2023; pp 203-248.

- (11) Bryant, N. D.; Pu, Y.; Tschaplinski, T. J.; Tuskan, G. A.; Muchero, W.; Kalluri, U. C.; Yoo, C. G.; Ragauskas, A. J. Transgenic Poplar Designed for Biofuels. *Trends Plant Sci* 2020, 25 (9), 881-896. DOI: 10.1016/j.tplants.2020.03.008 From NLM Medline.
- (12) Hinchee, M.; Rottmann, W.; Mullinax, L.; Zhang, C.; Chang, S.; Cunningham, M.; Pearson, L.; Nehra, N. Short-rotation woody crops for bioenergy and biofuels applications. *In Vitro Cell Dev Biol Plant* 2009, 45 (6), 619-629. DOI: 10.1007/s11627-009-9235-5 From NLM PubMed-not-MEDLINE.
- (13) Jansson, S.; Douglas, C. J. Populus: a model system for plant biology. *Annu Rev Plant Biol* 2007, 58, 435-458. DOI: 10.1146/annurev.arplant.58.032806.103956 From NLM Medline.
- (14) Tuskan, G. A.; Difazio, S.; Jansson, S.; Bohlmann, J.; Grigoriev, I.; Hellsten, U.; Putnam, N.; Ralph, S.; Rombauts, S.; Salamov, A.; et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 2006, 313 (5793), 1596-1604. DOI: 10.1126/science.1128691 From NLM Medline.
- (15) Gottel, N. R.; Castro, H. F.; Kerley, M.; Yang, Z. M.; Pelletier, D. A.; Podar, M.; Karpinets, T.; Uberbacher, E.; Tuskan, G. A.; Vilgalys, R.; et al. Distinct Microbial Communities within the Endosphere and Rhizosphere of
Roots across Contrasting Soil Types. *Appl Environ Microb* 2011, 77 (17), 5934-5944. DOI: 10.1128/Aem.05255-11.
- (16) 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.; et al. Cultivating the Bacterial Microbiota of Populus Roots. *mSystems* 2021, 6 (3), e0130620. DOI: 10.1128/mSystems.01306-20 From NLM PubMed-not-MEDLINE.
- (17) Trivedi, P.; Leach, J. E.; Tringe, S. G.; Sa, T.; Singh, B. K. Plant-microbiome interactions: from community assembly to plant health. *Nat Rev Microbiol* 2020, 18 (11), 607-621. DOI: 10.1038/s41579-020-0412-1 From NLM Medline.
- (18) Panke-Buisse, K.; Poole, A. C.; Goodrich, J. K.; Ley, R. E.; Kao-Kniffin, J. Selection on soil microbiomes reveals reproducible impacts on plant function. *ISME J* 2015, 9 (4), 980-989. DOI: 10.1038/ismej.2014.196 From NLM Medline.
- (19) Bandyopadhyay, P.; Bhuyan, S. K.; Yadava, P. K.; Varma, A.; Tuteja, N. Emergence of plant and rhizospheric microbiota as stable interactomes. *Protoplasma* 2017, 254 (2), 617-626. DOI: 10.1007/s00709-016-1003-x.
- (20) Bakker, P. A.; Berendsen, R. L.; Doornbos, R. F.; Wittermans, P. C.; Pieterse, C. M. The rhizosphere revisited: root microbiomics. *Front Plant Sci* 2013, 4, 165. DOI: 10.3389/fpls.2013.00165 From NLM PubMed-not-MEDLINE.

- (21) Ma, W.; Tang, S.; Dengzeng, Z.; Zhang, D.; Zhang, T.; Ma, X. Root exudates contribute to belowground ecosystem hotspots: A review. *Front Microbiol* 2022, *13*, 937940. DOI: 10.3389/fmicb.2022.937940 From NLM PubMed-not-MEDLINE.
- (22) Badri, D. V.; Vivanco, J. M. Regulation and function of root exudates. *Plant Cell Environ* 2009, *32* (6), 666-681. DOI: 10.1111/j.1365-3040.2008.01926.x From NLM Medline.
- (23) Alahmad, A.; Harir, M.; Fochesato, S.; Tulumello, J.; Walker, A.; Barakat, M.; Ndour, P. M. S.; Schmitt-Kopplin, P.; Cournac, L.; Laplaze, L.; et al. Unraveling the interplay between root exudates, microbiota, and rhizosheath formation in pearl millet. *Microbiome* 2024, *12* (1). DOI: ARTN 1
10.1186/s40168-023-01727-3.
- (24) Pantigoso, H. A.; Newberger, D.; Vivanco, J. M. The rhizosphere microbiome: Plant-microbial interactions for resource acquisition. *J Appl Microbiol* 2022, *133* (5), 2864-2876. DOI: 10.1111/jam.15686 From NLM Medline.
- (25) Sasse, J.; Martinoia, E.; Northen, T. Feed Your Friends: Do Plant Exudates Shape the Root Microbiome? *Trends Plant Sci* 2018, *23* (1), 25-41. DOI: 10.1016/j.tplants.2017.09.003 From NLM Medline.
- (26) Saeed, Q.; Xiukang, W.; Haider, F. U.; Kucerik, J.; Mumtaz, M. Z.; Holatko, J.; Naseem, M.; Kintl, A.; Ejaz, M.; Naveed, M.; et al. Rhizosphere Bacteria in Plant Growth Promotion, Biocontrol, and Bioremediation of Contaminated Sites: A Comprehensive Review of Effects and Mechanisms. *Int J Mol Sci* 2021, *22* (19). DOI: 10.3390/ijms221910529 From NLM Medline.
- (27) Thepbandit, W.; Athinuwat, D. Rhizosphere Microorganisms Supply Availability of Soil Nutrients and Induce Plant Defense. *Microorganisms* 2024, *12* (3). DOI: 10.3390/microorganisms12030558 From NLM PubMed-not-MEDLINE.
- (28) Alori, E. T.; Glick, B. R.; Babalola, O. O. Microbial Phosphorus Solubilization and Its Potential for Use in Sustainable Agriculture. *Front Microbiol* 2017, *8*, 971. DOI: 10.3389/fmicb.2017.00971 From NLM PubMed-not-MEDLINE.
- (29) Kudoyarova, G.; Arkhipova, T.; Korshunova, T.; Bakaeva, M.; Loginov, O.; Dodd, I. C. Phytohormone Mediation of Interactions Between Plants and Non-Symbiotic Growth Promoting Bacteria Under Edaphic Stresses. *Front Plant Sci* 2019, *10*. DOI: ARTN 1368
10.3389/fpls.2019.01368.
- (30) Spaepen, S.; Vanderleyden, J.; Remans, R. Indole-3-acetic acid in microbial and microorganism-plant signaling. *FEMS Microbiol Rev* 2007, *31* (4), 425-448. DOI: 10.1111/j.1574-6976.2007.00072.x From NLM Medline.

- (31) Raaijmakers, J. M.; Paulitz, T. C.; Steinberg, C.; Alabouvette, C.; Moënne-Loccoz, Y. The rhizosphere: a playground and battlefield for soilborne pathogens and beneficial microorganisms. *Plant and Soil* 2009, *321* (1-2), 341-361. DOI: 10.1007/s11104-008-9568-6.
- (32) Nadarajah, K.; Abdul Rahman, N. S. N. Plant-Microbe Interaction: Aboveground to Belowground, from the Good to the Bad. *Int J Mol Sci* 2021, *22* (19). DOI: 10.3390/ijms221910388 From NLM Medline.
- (33) Berendsen, R. L.; Pieterse, C. M.; Bakker, P. A. The rhizosphere microbiome and plant health. *Trends Plant Sci* 2012, *17* (8), 478-486. DOI: 10.1016/j.tplants.2012.04.001 From NLM Medline.
- (34) Tecon, R.; Mitri, S.; Ciccarese, D.; Or, D.; van der Meer, J. R.; Johnson, D. R. Bridging the Holistic-Reductionist Divide in Microbial Ecology. *mSystems* 2019, *4* (1). DOI: 10.1128/mSystems.00265-18 From NLM.
- (35) Fitzpatrick, C. R.; Salas-Gonzalez, I.; Conway, J. M.; Finkel, O. M.; Gilbert, S.; Russ, D.; Teixeira, P.; Dangel, J. L. The Plant Microbiome: From Ecology to Reductionism and Beyond. *Annu Rev Microbiol* 2020, *74*, 81-100. DOI: 10.1146/annurev-micro-022620-014327 From NLM Medline.
- (36) Liu, Y. X.; Qin, Y.; Bai, Y. Reductionist synthetic community approaches in root microbiome research. *Curr Opin Microbiol* 2019, *49*, 97-102. DOI: 10.1016/j.mib.2019.10.010 From NLM Medline.
- (37) Zengler, K.; Hofmockel, K.; Baliga, N. S.; Behie, S. W.; Bernstein, H. C.; Brown, J. B.; Dinneny, J. R.; Flöge, S. A.; Forry, S. P.; Hess, M.; et al. EcoFABs: advancing microbiome science through standardized fabricated ecosystems. *Nature Methods* 2019, *16* (7), 567-571. DOI: 10.1038/s41592-019-0465-0.
- (38) Oyserman, B. O.; Cordovez, V.; Flores, S. S.; Leite, M. F. A.; Nijveen, H.; Medema, M. H.; Raaijmakers, J. M. Extracting the GEMs: Genotype, Environment, and Microbiome Interactions Shaping Host Phenotypes. *Front Microbiol* 2020, *11*, 574053. DOI: 10.3389/fmicb.2020.574053 From NLM PubMed-not-MEDLINE.
- (39) Marin, O.; Gonzalez, B.; Poupin, M. J. From Microbial Dynamics to Functionality in the Rhizosphere: A Systematic Review of the Opportunities With Synthetic Microbial Communities. *Front Plant Sci* 2021, *12*, 650609. DOI: 10.3389/fpls.2021.650609 From NLM PubMed-not-MEDLINE.
- (40) Sergaki, C.; Lagunas, B.; Lidbury, I.; Gifford, M. L.; Schafer, P. Challenges and Approaches in Microbiome Research: From Fundamental to Applied. *Front Plant Sci* 2018, *9*, 1205. DOI: 10.3389/fpls.2018.01205 From NLM PubMed-not-MEDLINE.
- (41) de Souza, R. S. C.; Armanhi, J. S. L.; Arruda, P. From Microbiome to Traits: Designing Synthetic Microbial Communities for Improved Crop Resiliency. *Front Plant Sci* 2020, *11*, 1179. DOI: 10.3389/fpls.2020.01179 From NLM PubMed-not-MEDLINE.

- (42) Carper, D. L.; Appidi, M. R.; Mudbhari, S.; Shrestha, H. K.; Hettich, R. L.; Abraham, P. E. The Promises, Challenges, and Opportunities of Omics for Studying the Plant Holobiont. *Microorganisms* 2022, *10* (10). DOI: 10.3390/microorganisms10102013 From NLM PubMed-not-MEDLINE.
- (43) Nicholson, J. K.; Lindon, J. C. Systems biology - Metabonomics. *Nature* 2008, *455* (7216), 1054-1056. DOI: 10.1038/4551054a.
- (44) Shade, A. Diversity is the question, not the answer. *Isme Journal* 2017, *11* (1), 1-6. DOI: 10.1038/ismej.2016.118.
- (45) Thompson, L. R.; Sanders, J. G.; McDonald, D.; Amir, A.; Ladau, J.; Locey, K. J.; Prill, R. J.; Tripathi, A.; Gibbons, S. M.; Ackermann, G.; et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 2017, *551* (7681), 457-463. DOI: 10.1038/nature24621 From NLM Medline.
- (46) Nilsson, R. H.; Ryberg, M.; Abarenkov, K.; Sjökvist, E.; Kristiansson, E. The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. *Fems Microbiology Letters* 2009, *296* (1), 97-101. DOI: 10.1111/j.1574-6968.2009.01618.x.
- (47) Shade, A.; Stopnisek, N. Abundance-occupancy distributions to prioritize plant core microbiome membership. *Curr Opin Microbiol* 2019, *49*, 50-58. DOI: 10.1016/j.mib.2019.09.008 From NLM Medline.
- (48) Beilsmith, K.; Thoen, M. P. M.; Brachi, B.; Gloss, A. D.; Khan, M. H.; Bergelson, J. Genome-wide association studies on the phyllosphere microbiome: Embracing complexity in host-microbe interactions. *The Plant journal : for cell and molecular biology* 2019, *97* (1), 164-181. DOI: 10.1111/tpj.14170 From NLM Medline.
- (49) Lambais, M. R.; Barrera, S. E.; Santos, E. C.; Crowley, D. E.; Jumpponen, A. Phyllosphere Metaproteomes of Trees from the Brazilian Atlantic Forest Show High Levels of Functional Redundancy. *Microbial Ecology* 2017, *73* (1), 123-134. DOI: 10.1007/s00248-016-0878-6.
- (50) Abram, F. Systems-based approaches to unravel multi-species microbial community functioning. *Comput Struct Biotechnol J* 2015, *13*, 24-32. DOI: 10.1016/j.csbj.2014.11.009 From NLM PubMed-not-MEDLINE.
- (51) Vayssier-Taussat, M.; Albina, E.; Citti, C.; Cosson, J. F.; Jacques, M. A.; Lebrun, M. H.; Le Loir, Y.; Ogliastro, M.; Petit, M. A.; Roumagnac, P.; et al. Shifting the paradigm from pathogens to pathobiome: new concepts in the light of meta-omics. *Front Cell Infect Microbiol* 2014, *4*, 29. DOI: 10.3389/fcimb.2014.00029 From NLM Medline.
- (52) VerBerkmoes, N. C.; Denev, V. J.; Hettich, R. L.; Banfield, J. F. Systems biology: Functional analysis of natural microbial consortia using community proteomics. *Nat Rev Microbiol* 2009, *7* (3), 196-205. DOI: 10.1038/nrmicro2080 From NLM Medline.

- (53) Hettich, R. L.; Pan, C. L.; Chourey, K.; Giannone, R. J.: Harnessing the Power of High Performance Mass Spectrometry to Identify the Suite of Proteins That Control Metabolic Activities in Microbial Communities. *Anal Chem* 2013, 85 (9), 4203-4214. DOI: 10.1021/ac303053e.
- (54) Renu; Gupta, S. K.; Rai, A. K.; Sarim, K. M.; Sharma, A.; Budhlakoti, N.; Arora, D.; Verma, D. K.; Singh, D. P. Metaproteomic data of maize rhizosphere for deciphering functional diversity. *Data Brief* 2019, 27. DOI: ARTN 104574
10.1016/j.dib.2019.104574.
- (55) Knief, C.; Delmotte, N.; Chaffron, S.; Stark, M.; Innerebner, G.; Wassmann, R.; von Mering, C.; Vorholt, J. A. Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *ISME J* 2012, 6 (7), 1378-1390. DOI: 10.1038/ismej.2011.192 From NLM Medline.
- (56) Bao, Z. H.; Okubo, T.; Kubota, K.; Kasahara, Y.; Tsurumaru, H.; Anda, M.; Ikeda, S.; Minamisawa, K. Metaproteomic Identification of Diazotrophic Methanotrophs and Their Localization in Root Tissues of Field-Grown Rice Plants. *Appl Environ Microb* 2014, 80 (16), 5043-5052. DOI: 10.1128/Aem.00969-14.
- (57) Salvato, F.; Hettich, R. L.; Kleiner, M. Five key aspects of metaproteomics as a tool to understand functional interactions in host-associated microbiomes. *Plos Pathogens* 2021, 17 (2). DOI: ARTN e1009245
10.1371/journal.ppat.1009245.
- (58) Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd. An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* 2001, 73 (23), 5683-5690. DOI: 10.1021/ac010617e From NLM Medline.
- (59) Salvato, F.; Vintila, S.; Finkel, O. M.; Dangl, J. L.; Kleiner, M. Evaluation of Protein Extraction Methods for Metaproteomic Analyses of Root-Associated Microbes. *Mol Plant Microbe Interact* 2022, 35 (11), 977-988. DOI: 10.1094/MPMI-05-22-0116-TA From NLM Medline.
- (60) Bauermeister, A.; Mannocho-Russo, H.; Costa-Lotufo, L. V.; Jarmusch, A. K.; Dorrestein, P. C. Mass spectrometry-based metabolomics in microbiome investigations. *Nat Rev Microbiol* 2022, 20 (3), 143-160. DOI: 10.1038/s41579-021-00621-9 From NLM Medline.
- (61) Gupta, S.; Schillaci, M.; Roessner, U. Metabolomics as an emerging tool to study plant-microbe interactions. *Emerg Top Life Sci* 2022, 6 (2), 175-183. DOI: ArtN EtlS20210262
10.1042/EtlS20210262.

- (62) Hiruma, K. Roles of Plant-Derived Secondary Metabolites during Interactions with Pathogenic and Beneficial Microbes under Conditions of Environmental Stress. *Microorganisms* 2019, 7 (9). DOI: 10.3390/microorganisms7090362 From NLM PubMed-not-MEDLINE.
- (63) McLaughlin, S.; Zhalnina, K.; Kosina, S.; Northen, T. R.; Sasse, J. The core metabolome and root exudation dynamics of three phylogenetically distinct plant species. *Nat Commun* 2023, 14 (1), 1649. DOI: 10.1038/s41467-023-37164-x From NLM Medline.
- (64) Jing, H.; Wang, H.; Wang, G.; Liu, G.; Cheng, Y. The mechanism effects of root exudate on microbial community of rhizosphere soil of tree, shrub, and grass in forest ecosystem under N deposition. *ISME Commun* 2023, 3 (1), 120. DOI: 10.1038/s43705-023-00322-9 From NLM PubMed-not-MEDLINE.
- (65) Oburger, E.; Jones, D. L. Sampling root exudates - Mission impossible? *Rhizosphere-Neth* 2018, 6, 116-133. DOI: 10.1016/j.rhisph.2018.06.004.
- (66) Kawasaki, A.; Okada, S.; Zhang, C.; Delhaize, E.; Mathesius, U.; Richardson, A. E.; Watt, M.; Gilliam, M.; Ryan, P. R. A sterile hydroponic system for characterising root exudates from specific root types and whole-root systems of large crop plants. *Plant Methods* 2018, 14, 114. DOI: 10.1186/s13007-018-0380-x From NLM PubMed-not-MEDLINE.
- (67) Lopez-Guerrero, M. G.; Wang, P.; Phares, F.; Schachtman, D. P.; Alvarez, S.; van Dijk, K. A glass bead semi-hydroponic system for intact maize root exudate analysis and phenotyping. *Plant Methods* 2022, 18 (1), 25. DOI: 10.1186/s13007-022-00856-4 From NLM PubMed-not-MEDLINE.
- (68) Sasse, J.; Kant, J.; Cole, B. J.; Klein, A. P.; Arsova, B.; Schlaepfer, P.; Gao, J.; Lewald, K.; Zhalnina, K.; Kosina, S.; et al. Multilab EcoFAB study shows highly reproducible physiology and depletion of soil metabolites by a model grass. *New Phytologist* 2019, 222 (2), 1149-1160. DOI: 10.1111/nph.15662.
- (69) Ziegler, J.; Schmidt, S.; Chutia, R.; Muller, J.; Bottcher, C.; Strehmel, N.; Scheel, D.; Abel, S. Non-targeted profiling of semi-polar metabolites in Arabidopsis root exudates uncovers a role for coumarin secretion and lignification during the local response to phosphate limitation. *J Exp Bot* 2016, 67 (5), 1421-1432. DOI: 10.1093/jxb/erv539 From NLM Medline.
- (70) Chaparro, J. M.; Badri, D. V.; Bakker, M. G.; Sugiyama, A.; Manter, D. K.; Vivanco, J. M. Root Exudation of Phytochemicals in Arabidopsis Follows Specific Patterns That Are Developmentally Programmed and Correlate with Soil Microbial Functions. *Plos One* 2013, 8 (2). DOI: ARTN e55731
10.1371/journal.pone.0055731.
- (71) Strehmel, N.; Bottcher, C.; Schmidt, S.; Scheel, D. Profiling of secondary metabolites in root exudates of Arabidopsis thaliana. *Phytochemistry* 2014, 108, 35-46. DOI: 10.1016/j.phytochem.2014.10.003 From NLM Medline.

- (72) Suzuki, K.; Okazaki, K.; Tawaraya, K.; Osaki, M.; Shinano, T. Gas chromatography-mass spectrometry associated global analysis of rice root exudates under aseptical conditions. *Soil Sci Plant Nutr* 2009, *55* (4), 505-513. DOI: 10.1111/j.1747-0765.2009.00390.x.
- (73) Tawaraya, K.; Horie, R.; Wagatsuma, T.; Saito, K.; Oikawa, A. Metabolite profiling of shoot extract, root extract, and root exudate of rice under nitrogen and phosphorus deficiency. *Soil Sci Plant Nutr* 2018, *64* (3), 312-322. DOI: 10.1080/00380768.2018.1476828.
- (74) Carvalhais, L. C.; Dennis, P. G.; Fedoseyenko, D.; Hajirezaei, M. R.; Borriss, R.; von Wirén, N. Root exudation of sugars, amino acids, and organic acids by maize as affected by nitrogen, phosphorus, potassium, and iron deficiency. *J Plant Nutr Soil Sc* 2011, *174* (1), 3-11. DOI: 10.1002/jpln.201000085.
- (75) Stringlis, I. A.; Yu, K.; Feussner, K.; de Jonge, R.; Van Bentum, S.; Van Verk, M. C.; Berendsen, R. L.; Bakker, P.; Feussner, I.; Pieterse, C. M. J. MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. *Proc Natl Acad Sci U S A* 2018, *115* (22), E5213-E5222. DOI: 10.1073/pnas.1722335115 From NLM Medline.
- (76) Hu, L.; Robert, C. A. M.; Cadot, S.; Zhang, X.; Ye, M.; Li, B.; Manzo, D.; Chervet, N.; Steinger, T.; van der Heijden, M. G. A.; et al. Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota. *Nat Commun* 2018, *9* (1), 2738. DOI: 10.1038/s41467-018-05122-7 From NLM Medline.
- (77) Phillips, R. P.; Erlitz, Y.; Bier, R.; Bernhardt, E. S. New approach for capturing soluble root exudates in forest soils. *Functional Ecology* 2008, *22* (6), 990-999. DOI: 10.1111/j.1365-2435.2008.01495.x.
- (78) Wilson, R. M.; Tfaily, M. M.; Kolton, M.; Johnston, E. R.; Petro, C.; Zalman, C. A.; Hanson, P. J.; Heyman, H. M.; Kyle, J. E.; Hoyt, D. W.; et al. Soil metabolome response to whole-ecosystem warming at the Spruce and Peatland Responses under Changing Environments experiment. *P Natl Acad Sci USA* 2021, *118* (25). DOI: ARTN e2004192118
10.1073/pnas.2004192118.
- (79) Li, Y.; Xu, L.; Letuma, P.; Lin, W. Metabolite profiling of rhizosphere soil of different allelopathic potential rice accessions. *BMC Plant Biol* 2020, *20* (1), 265. DOI: 10.1186/s12870-020-02465-6 From NLM Medline.
- (80) Gao, Y. Z.; Yang, Y.; Ling, W. T.; Kong, H. L.; Zhu, X. Z. Gradient Distribution of Root Exudates and Polycyclic Aromatic Hydrocarbons in Rhizosphere Soil. *Soil Sci Soc Am J* 2011, *75* (5), 1694-1703. DOI: 10.2136/sssaj2010.0244.
- (81) Petriacq, P.; Williams, A.; Cotton, A.; McFarlane, A. E.; Rolfe, S. A.; Ton, J. Metabolite profiling of non-sterile rhizosphere soil. *The Plant journal : for cell and molecular biology* 2017, *92* (1), 147-162. DOI: 10.1111/tpj.13639 From NLM Medline.

- (82) Karpievitch, Y. V.; Polpitiya, A. D.; Anderson, G. A.; Smith, R. D.; Dabney, A. R. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. *Ann Appl Stat* 2010, 4 (4), 1797-1823. DOI: 10.1214/10-AOAS341 From NLM PubMed-not-MEDLINE.
- (83) Zhou, B.; Xiao, J. F.; Tuli, L.; Ransom, H. W. LC-MS-based metabolomics. *Mol Biosyst* 2012, 8 (2), 470-481. DOI: 10.1039/c1mb05350g From NLM Medline.
- (84) Pitt, J. J. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin Biochem Rev* 2009, 30 (1), 19-34. From NLM PubMed-not-MEDLINE.
- (85) Rusli, H.; Putri, R. M.; Alni, A. Recent Developments of Liquid Chromatography Stationary Phases for Compound Separation: From Proteins to Small Organic Compounds. *Molecules* 2022, 27 (3). DOI: ARTN 907
10.3390/molecules27030907.
- (86) Xie, F.; Smith, R. D.; Shen, Y. F. Advanced proteomic liquid chromatography. *J Chromatogr A* 2012, 1261, 78-90. DOI: 10.1016/j.chroma.2012.06.098.
- (87) Criscuolo, A.; Zeller, M.; Cook, K.; Angelidou, G.; Fedorova, M. Rational selection of reverse phase columns for high throughput LC-MS lipidomics. *Chem Phys Lipids* 2019, 221, 120-127. DOI: 10.1016/j.chemphyslip.2019.03.006.
- (88) Alsaleh, M.; Barbera, T. A.; Andrews, R. H.; Sithithaworn, P.; Khuntikeo, N.; Loilome, W.; Yongvanit, P.; Cox, I. J.; Syms, R. R. A.; Holmes, E.; et al. Mass Spectrometry: A Guide for the Clinician. *J Clin Exp Hepatol* 2019, 9 (5), 597-606. DOI: 10.1016/j.jceh.2019.04.053 From NLM PubMed-not-MEDLINE.
- (89) Lenco, J.; Jadeja, S.; Naplekov, D. K.; Krokhin, O.; Khalikova, M. A.; Chocholous, P.; Urban, J.; Broeckhoven, K.; Nováková, L.; Svec, F. Reversed-Phase Liquid Chromatography of Peptides for Bottom-Up Proteomics: A Tutorial. *J Proteome Res* 2022, 21 (12), 2846-2892. DOI: 10.1021/acs.jproteome.2c00407.
- (90) Churchwell, M. I.; Twaddle, N. C.; Meeker, L. R.; Doerge, D. R. Improving LC-MS sensitivity through increases in chromatographic performance: Comparisons of UPLC-ES/MS/MS to HPLC-ES/MS/MS. *J Chromatogr B* 2005, 825 (2), 134-143. DOI: 10.1016/j.jchromb.2005.05.037.
- (91) Wilson, S. R.; Vehus, T.; Berg, H. S.; Lundanes, E. Nano-LC in proteomics: recent advances and approaches. *Bioanalysis* 2015, 7 (14), 1799-1815. DOI: 10.4155/Bio.15.92.
- (92) Shen, Y.; Smith, R. D. Advanced nanoscale separations and mass spectrometry for sensitive high-throughput proteomics. *Expert Rev Proteomics* 2005, 2 (3), 431-447. DOI: 10.1586/14789450.2.3.431 From NLM Medline.

- (93) Zhang, X.; Fang, A.; Riley, C. P.; Wang, M.; Regnier, F. E.; Buck, C. Multi-dimensional liquid chromatography in proteomics--a review. *Anal Chim Acta* 2010, *664* (2), 101-113. DOI: 10.1016/j.aca.2010.02.001 From NLM Medline.
- (94) Stoll, D. R.; Carr, P. W. Two-Dimensional Liquid Chromatography: A State of the Art Tutorial. *Anal Chem* 2017, *89* (1), 519-531. DOI: 10.1021/acs.analchem.6b03506.
- (95) Nagele, E.; Vollmer, M.; Horth, P.; Vad, C. 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. *Expert Rev Proteomics* 2004, *1* (1), 37-46. DOI: 10.1586/14789450.1.1.37 From NLM Medline.
- (96) Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001, *19* (3), 242-247. DOI: 10.1038/85686 From NLM Medline.
- (97) Finehout, E. J.; Lee, K. H. An introduction to mass spectrometry applications in biological research. *Biochem Mol Biol Educ* 2004, *32* (2), 93-100. DOI: 10.1002/bmb.2004.494032020331 From NLM PubMed-not-MEDLINE.
- (98) El-Aneed, A.; Cohen, A.; Banoub, J. Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Appl Spectrosc Rev* 2009, *44* (3), 210-230. DOI: Pii 909281413
10.1080/05704920902717872.
- (99) Awad, H.; Khamis, M. M.; El-Aneed, A. Mass Spectrometry, Review of the Basics: Ionization. *Appl Spectrosc Rev* 2015, *50* (2), 158-175. DOI: 10.1080/05704928.2014.954046.
- (100) Von Helden, G.; Wyttenbach, T.; Bowers, M. T. Inclusion of a Maldi Ion-Source in the Ion Chromatography Technique - Conformational Information on Polymer and Biomolecular Ions. *Int J Mass Spectrom* 1995, *146*, 349-364. DOI: Doi 10.1016/0168-1176(95)04211-3.
- (101) Morris, C. B.; Poland, J. C.; May, J. C.; McLean, J. A. Fundamentals of Ion Mobility-Mass Spectrometry for the Analysis of Biomolecules. *Methods Mol Biol* 2020, *2084*, 1-31. DOI: 10.1007/978-1-0716-0030-6_1 From NLM Medline.
- (102) Smith, R. D.; Loo, J. A.; Loo, R. R. O.; Busman, M.; Udseth, H. R. Principles and Practice of Electrospray Ionization - Mass-Spectrometry for Large Polypeptides and Proteins. *Mass Spectrom Rev* 1991, *10* (5), 359-451. DOI: DOI 10.1002/mas.1280100504.
- (103) Aichler, M.; Walch, A. MALDI Imaging mass spectrometry: current frontiers and perspectives in pathology research and practice. *Lab Invest* 2015, *95* (4), 422-431. DOI: 10.1038/labinvest.2014.156.
- (104) Liigand, P.; Kaupmees, K.; Haav, K.; Liigand, J.; Leito, I.; Girod, M.; Antoine, R.; Kruve, A. Think Negative: Finding the Best Electrospray Ionization/MS Mode for Your Analyte. *Anal*

Chem 2017, 89 (11), 5665-5668. DOI: 10.1021/acs.analchem.7b00096 From NLM PubMed-not-MEDLINE.

(105) Liebler, D. C. *Introduction to proteomics: tools for the new biology*; Springer Science & Business Media, 2001.

(106) Doerr, A. Proteomic. Navigating the negative-mode proteome. *Nat Methods* 2015, 12 (9), 808. DOI: 10.1038/nmeth.3573 From NLM Medline.

(107) Ho, C. S.; Lam, C. W.; Chan, M. H.; Cheung, R. C.; Law, L. K.; Lit, L. C.; Ng, K. F.; Suen, M. W.; Tai, H. L. Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin Biochem Rev* 2003, 24 (1), 3-12. From NLM PubMed-not-MEDLINE.

(108) Li, C.; Chu, S.; Tan, S.; Yin, X.; Jiang, Y.; Dai, X.; Gong, X.; Fang, X.; Tian, D. Towards Higher Sensitivity of Mass Spectrometry: A Perspective From the Mass Analyzers. *Front Chem* 2021, 9, 813359. DOI: 10.3389/fchem.2021.813359 From NLM PubMed-not-MEDLINE.

(109) Niessen, W. M.; Falck, D. Introduction to mass spectrometry, a tutorial. *Analyzing biomolecular interactions by mass spectrometry* 2015, 1-54.

(110) Pirmoradian, M.; Budamgunta, H.; Chingin, K.; Zhang, B.; Astorga-Wells, J.; Zubarev, R. A. Rapid and deep human proteome analysis by single-dimension shotgun proteomics. *Mol Cell Proteomics* 2013, 12 (11), 3330-3338. DOI: 10.1074/mcp.O113.028787 From NLM Medline.

(111) Zubarev, R. A.; Makarov, A. Orbitrap mass spectrometry. ACS Publications: 2013.

(112) Scheltema, R. A.; Hauschild, J.-P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M. The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Molecular & Cellular Proteomics* 2014, 13 (12), 3698-3708.

(113) Neagu, A. N.; Jayathirtha, M.; Baxter, E.; Donnelly, M.; Petre, B. A.; Darie, C. C. Applications of Tandem Mass Spectrometry (MS/MS) in Protein Analysis for Biomedical Research. *Molecules* 2022, 27 (8). DOI: 10.3390/molecules27082411 From NLM Medline.

(114) Jedrychowski, M. P.; Huttlin, E. L.; Haas, W.; Sowa, M. E.; Rad, R.; Gygi, S. P. Evaluation of HCD- and CID-type Fragmentation Within Their Respective Detection Platforms For Murine Phosphoproteomics. *Molecular & Cellular Proteomics* 2011, 10 (12). DOI: ARTN M111.009910 10.1074/mcp.M111.009910.

(115) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* 2007, 4 (9), 709-712. DOI: 10.1038/nmeth1060 From NLM Medline.

(116) Nakayasu, E. S.; Gritsenko, M.; Piehowski, P. D.; Gao, Y.; Orton, D. J.; Schepmoes, A. A.; Fillmore, T. L.; Frohnert, B. I.; Rewers, M.; Krischer, J. P.; et al. Tutorial: best practices and

considerations for mass-spectrometry-based protein biomarker discovery and validation. *Nat Protoc* 2021, 16 (8), 3737-3760. DOI: 10.1038/s41596-021-00566-6 From NLM Medline.

(117) Wilkins, M. R.; Sanchez, J. C.; Gooley, A. A.; Appel, R. D.; HumpherySmith, I.; Hochstrasser, D. F.; Williams, K. L. Progress with proteome projects: Why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng* 1996, 13, 19-50. DOI: Doi 10.1080/02648725.1996.10647923.

(118) Wilmes, P.; Bond, P. L. The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ Microbiol* 2004, 6 (9), 911-920. DOI: 10.1111/j.1462-2920.2004.00687.x From NLM Medline.

(119) Chait, B. T. Mass spectrometry: bottom-up or top-down? *Science* 2006, 314 (5796), 65-66.

(120) Pandeswari, P. B.; Sabareesh, V. Middle-down approach: a choice to sequence and characterize proteins/proteomes by mass spectrometry. *RSC Adv* 2018, 9 (1), 313-344. DOI: 10.1039/c8ra07200k From NLM PubMed-not-MEDLINE.

(121) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top Down proteomics: Facts and perspectives. *Biochem Bioph Res Co* 2014, 445 (4), 683-693. DOI: 10.1016/j.bbrc.2014.02.041.

(122) Dupree, E. J.; Jayathirtha, M.; Yorkey, H.; Mihasan, M.; Petre, B. A.; Dark, C. C. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. *Proteomes* 2020, 8 (3). DOI: ARTN 14

10.3390/proteomes8030014.

(123) Graves, P. R.; Haystead, T. A. Molecular biologist's guide to proteomics. *Microbiology and molecular biology reviews : MMBR* 2002, 66 (1), 39-63; table of contents. DOI: 10.1128/MMBR.66.1.39-63.2002 From NLM Medline.

(124) Dhabaria, A.; Cifani, P.; Reed, C.; Steen, H.; Kentsis, A. A High-Efficiency Cellular Extraction System for Biological Proteomics. *J Proteome Res* 2015, 14 (8), 3403-3408. DOI: 10.1021/acs.jproteome.5b00547 From NLM Medline.

(125) Canas, B.; Pineiro, C.; Calvo, E.; Lopez-Ferrer, D.; Gallardo, J. M. Trends in sample preparation for classical and second generation proteomics. *J Chromatogr A* 2007, 1153 (1-2), 235-258. DOI: 10.1016/j.chroma.2007.01.045 From NLM Medline.

(126) Vertommen, A.; Panis, B.; Swennen, R.; Carpentier, S. C. Evaluation of chloroform/methanol extraction to facilitate the study of membrane proteins of non-model plants. *Planta* 2010, 231 (5), 1113-1125. DOI: 10.1007/s00425-010-1121-1 From NLM Medline.

(127) Niu, L.; Zhang, H.; Wu, Z.; Wang, Y.; Liu, H.; Wu, X.; Wang, W. Modified TCA/acetone precipitation of plant proteins for proteomic analysis. *PLoS One* 2018, 13 (12), e0202238. DOI: 10.1371/journal.pone.0202238 From NLM Medline.

- (128) Batth, T. S.; Tollenaere, M. X.; Ruther, P.; Gonzalez-Franquesa, A.; Prabhakar, B. S.; Bekker-Jensen, S.; Deshmukh, A. S.; Olsen, J. V. Protein Aggregation Capture on Microparticles Enables Multipurpose Proteomics Sample Preparation. *Mol Cell Proteomics* 2019, 18 (5), 1027-1035. DOI: 10.1074/mcp.TIR118.001270 From NLM Medline.
- (129) Chourey, K.; Jansson, J.; VerBerkmoes, N.; Shah, M.; Chavarria, K. L.; Tom, L. M.; Brodie, E. L.; Hettich, R. L. Direct cellular lysis/protein extraction protocol for soil metaproteomics. *J Proteome Res* 2010, 9 (12), 6615-6622. DOI: 10.1021/pr100787q From NLM Medline.
- (130) Qian, C.; Hettich, R. L. Optimized Extraction Method To Remove Humic Acid Interferences from Soil Samples Prior to Microbial Proteome Measurements. *J Proteome Res* 2017, 16 (7), 2537-2546. DOI: 10.1021/acs.jproteome.7b00103 From NLM Medline.
- (131) Suttapitugsakul, S.; Xiao, H.; Smeekens, J.; Wu, R. Evaluation and optimization of reduction and alkylation methods to maximize peptide identification with MS-based proteomics. *Mol Biosyst* 2017, 13 (12), 2574-2582. DOI: 10.1039/c7mb00393e From NLM Medline.
- (132) Dau, T.; Bartolomucci, G.; Rappsilber, J. Proteomics Using Protease Alternatives to Trypsin Benefits from Sequential Digestion with Trypsin. *Anal Chem* 2020, 92 (14), 9523-9527. DOI: 10.1021/acs.analchem.0c00478 From NLM Medline.
- (133) Mansuri, M. S.; Bathla, S.; Lam, T. T.; Nairn, A. C.; Williams, K. R. Optimal conditions for carrying out trypsin digestions on complex proteomes: From bulk samples to single cells. *J Proteomics* 2024, 297, 105109. DOI: 10.1016/j.jprot.2024.105109 From NLM Medline.
- (134) Waas, M.; Pereckas, M.; Jones Lipinski, R. A.; Ashwood, C.; Gundry, R. L. SP2: Rapid and Automatable Contaminant Removal from Peptide Samples for Proteomic Analyses. *J Proteome Res* 2019, 18 (4), 1644-1656. DOI: 10.1021/acs.jproteome.8b00916 From NLM Medline.
- (135) Di Palma, S.; Hennrich, M. L.; Heck, A. J.; Mohammed, S. Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *J Proteomics* 2012, 75 (13), 3791-3813. DOI: 10.1016/j.jprot.2012.04.033 From NLM Medline.
- (136) Josic, D.; Kovac, S. Reversed-phase High Performance Liquid Chromatography of proteins. *Curr Protoc Protein Sci* 2010, Chapter 8, 8 7 1-8 7 22. DOI: 10.1002/0471140864.ps0807s61 From NLM Medline.
- (137) Michalski, A.; Neuhauser, N.; Cox, J.; Mann, M. A Systematic Investigation into the Nature of Tryptic HCD Spectra. *J Proteome Res* 2012, 11 (11), 5479-5491. DOI: 10.1021/pr3007045.
- (138) Cleveland, J. P.; Rose, J. R. Identification of b-/y-ions in MS/MS spectra using a two stage neural network. *Proteome Sci* 2013, 11 (Suppl 1), S4. DOI: 10.1186/1477-5956-11-S1-S4 From NLM PubMed-not-MEDLINE.

- (139) Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G. A.; Ma, B. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics* 2012, *11* (4), M111 010587. DOI: 10.1074/mcp.M111.010587 From NLM Medline.
- (140) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 1994, *5* (11), 976-989. DOI: 10.1016/1044-0305(94)80016-2.
- (141) Dorfer, V.; Pichler, P.; Stranzl, T.; Stadlmann, J.; Taus, T.; Winkler, S.; Mechtler, K. MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J Proteome Res* 2014, *13* (8), 3679-3684.
- (142) Brosch, M.; Yu, L.; Hubbard, T.; Choudhary, J. Accurate and sensitive peptide identification with Mascot Percolator. *J Proteome Res* 2009, *8* (6), 3176-3181.
- (143) The, M.; MacCoss, M. J.; Noble, W. S.; Käll, L. Fast and accurate protein false discovery rates on large-scale proteomics data sets with percolator 3.0. *J Am Soc Mass Spectr* 2016, *27*, 1719-1727.
- (144) Huang, T.; Wang, J.; Yu, W.; He, Z. Protein inference: a review. *Briefings in bioinformatics* 2012, *13* (5), 586-614.
- (145) Orsburn, B. C. Proteome discoverer—a community enhanced data processing suite for protein informatics. *Proteomes* 2021, *9* (1), 15.
- (146) Zhu, W.; Smith, J. W.; Huang, C.-M. Mass spectrometry-based label-free quantitative proteomics. *BioMed research international* 2010, *2010* (1), 840518.
- (147) Välikangas, T.; Suomi, T.; Elo, L. L. A systematic evaluation of normalization methods in quantitative label-free proteomics. *Briefings in bioinformatics* 2018, *19* (1), 1-11.
- (148) Jin, L.; Bi, Y.; Hu, C.; Qu, J.; Shen, S.; Wang, X.; Tian, Y. A comparative study of evaluating missing value imputation methods in label-free proteomics. *Sci Rep-Uk* 2021, *11* (1), 1760.
- (149) Polpitiya, A. D.; Qian, W. J.; Jaitly, N.; Petyuk, V. A.; Adkins, J. N.; Camp, D. G., 2nd; Anderson, G. A.; Smith, R. D. DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* 2008, *24* (13), 1556-1558. DOI: 10.1093/bioinformatics/btn217.
- (150) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 2016, *13* (9), 731-740. DOI: 10.1038/nmeth.3901.
- (151) Xiao, J. F.; Zhou, B.; Ransom, H. W. Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *Trends Analyt Chem* 2012, *32*, 1-14. DOI: 10.1016/j.trac.2011.08.009 From NLM PubMed-not-MEDLINE.

- (152) Patti, G. J.; Yanes, O.; Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* 2012, *13* (4), 263-269. DOI: 10.1038/nrm3314 From NLM Medline.
- (153) Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *J Am Soc Mass Spectr* 2016, *27* (12), 1897-1905. DOI: 10.1007/s13361-016-1469-y.
- (154) Bowen, B. P.; Northen, T. R. Dealing with the Unknown: Metabolomics and Metabolite Atlases. *J Am Soc Mass Spectr* 2010, *21* (9), 1471-1476. DOI: 10.1016/j.jasms.2010.04.003.
- (155) Roberts, L. D.; Souza, A. L.; Gerszten, R. E.; Clish, C. B. Targeted metabolomics. *Curr Protoc Mol Biol* 2012, *Chapter 30*, Unit 30 32 31-24. DOI: 10.1002/0471142727.mb3002s98 From NLM Medline.
- (156) Gupta, S.; Schillaci, M.; Roessner, U. Metabolomics as an emerging tool to study plant-microbe interactions. *Emerg Top Life Sci* 2022, *6* (2), 175-183.
- (157) Salem, M. A.; Wang, J. Y.; Al-Babili, S. Metabolomics of plant root exudates: From sample preparation to data analysis. *Front Plant Sci* 2022, *13*, 1062982. DOI: 10.3389/fpls.2022.1062982 From NLM PubMed-not-MEDLINE.
- (158) Shi, H.; Yang, J.; Li, Q.; PinChu, C.; Song, Z.; Yang, H.; Luo, Y.; Liu, C.; Fan, W. Diversity and correlation analysis of different root exudates on the regulation of microbial structure and function in soil planted with *Panax notoginseng*. *Frontiers in microbiology* 2023, *14*, 1282689.
- (159) Salem, M. A.; Perez de Souza, L.; Serag, A.; Fernie, A. R.; Farag, M. A.; Ezzat, S. M.; Alseekh, S. Metabolomics in the context of plant natural products research: From sample preparation to metabolite analysis. *Metabolites* 2020, *10* (1), 37.
- (160) Casas, M. E.; Matamoros, V. Analytical challenges and solutions for performing metabolomic analysis of root exudates. *Trends Environ Anal* 2021, *31*. DOI: ARTN e00130 10.1016/j.teac.2021.e00130.
- (161) Zheng, J.; Johnson, M.; Mandal, R.; Wishart, D. S. A comprehensive targeted metabolomics assay for crop plant sample analysis. *Metabolites* 2021, *11* (5), 303.
- (162) Sawada, Y.; Hirai, M. Y. Integrated LC-MS/MS system for plant metabolomics. *Computational and Structural Biotechnology Journal* 2013, *4* (5), e201301011.
- (163) Floros, D. J.; Xu, K.; Berthiller, F.; Schwartz-Zimmermann, H. Comparison of chromatographic conditions for the targeted tandem mass spectrometric determination of 354 mammalian metabolites. *J Chromatogr A* 2023, *1697*, 463985.

- (164) Lei, Z.; Huhman, D. V.; Sumner, L. W. Mass spectrometry strategies in metabolomics. *Journal of Biological Chemistry* 2011, 286 (29), 25435-25442.
- (165) Rathore, N.; Hanzelková, V.; Dostálek, T.; Semerád, J.; Schnablová, R.; Cajthaml, T.; Münzbergová, Z. Species phylogeny, ecology, and root traits as predictors of root exudate composition. *New Phytologist* 2023, 239 (4), 1212-1224.
- (166) Heuckeroth, S.; Damiani, T.; Smirnov, A.; Mokshyna, O.; Brungs, C.; Korf, A.; Smith, J. D.; Stincone, P.; Dreolin, N.; Nothias, L.-F. Reproducible mass spectrometry data processing and compound annotation in MZmine 3. *Nature protocols* 2024, 1-45.
- (167) Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal Chem* 2012, 84 (11), 5035-5039.
- (168) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature methods* 2015, 12 (6), 523-526.
- (169) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC bioinformatics* 2010, 11, 1-11.
- (170) Pang, Z.; Lu, Y.; Zhou, G.; Hui, F.; Xu, L.; Viau, C.; Spigelman, A. F.; MacDonald, P. E.; Wishart, D. S.; Li, S. MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucleic Acids Research* 2024, gkae253.
- (171) Scheubert, K.; Hufsky, F.; Petras, D.; Wang, M.; Nothias, L.-F.; Dührkop, K.; Bandeira, N.; Dorrestein, P. C.; Böcker, S. Significance estimation for large scale metabolomics annotations by spectral matching. *Nature Communications* 2017, 8 (1), 1494.
- (172) Schiffman, C.; Petrick, L.; Perttula, K.; Yano, Y.; Carlsson, H.; Whitehead, T.; Metayer, C.; Hayes, J.; Rappaport, S.; Dudoit, S. Filtering procedures for untargeted LC-MS metabolomics data. *BMC bioinformatics* 2019, 20, 1-10.
- (173) Wei, R.; Wang, J.; Su, M.; Jia, E.; Chen, S.; Chen, T.; Ni, Y. Missing value imputation approach for mass spectrometry-based metabolomics data. *Sci Rep-Uk* 2018, 8 (1), 663.
- (174) Sun, J.; Xia, Y. Pretreating and normalizing metabolomics data for statistical analysis. *Genes & Diseases* 2024, 11 (3), 100979.
- (175) Worley, B.; Powers, R. Multivariate analysis in metabolomics. *Current metabolomics* 2013, 1 (1), 92-107.
- (176) Chen, T.; Cao, Y.; Zhang, Y.; Liu, J.; Bao, Y.; Wang, C.; Jia, W.; Zhao, A. Random forest in clinical metabolomics for phenotypic discrimination and biomarker selection. *Evidence-Based Complementary and Alternative Medicine* 2013, 2013 (1), 298183.

(177) Pii, Y.; Mimmo, T.; Tomasi, N.; Terzano, R.; Cesco, S.; Crecchio, C. Microbial interactions in the rhizosphere: beneficial influences of plant growth-promoting rhizobacteria on nutrient acquisition process. A review. *Biol. Fertil. Soils* 2015, *51* (4), 403-415. DOI: 10.1007/s00374-015-0996-1.

(178) de Zelicourt, A.; Al-Yousif, M.; Hirt, H. Rhizosphere Microbes as Essential Partners for Plant Stress Tolerance. *Molecular Plant* 2013, *6* (2), 242-245. DOI: 10.1093/mp/sst028.

(179) Berendsen, R. L.; Pieterse, C. M. J.; Bakker, P. A. H. M. The rhizosphere microbiome and plant health. *Trends in Plant Science* 2012, *17* (8), 478-486. DOI: 10.1016/j.tplants.2012.04.001.

(180) Pineda, A.; Soler, R.; Pozo, M. J.; Rasmann, S.; Turlings, T. C. J. Editorial: Above-belowground interactions involving plants, microbes and insects. *Front Plant Sci* 2015, *6*. DOI: ARTN 318

10.3389/fpls.2015.00318.

(181) Kumar, M.; Ji, B.; Zengler, K.; Nielsen, J. Modelling approaches for studying the microbiome. *Nat Microbiol* 2019, *4* (8), 1253-1267. DOI: 10.1038/s41564-019-0491-9 From NLM Medline.

(182) Joseph, T. A.; Shenhav, L.; Xavier, J. B.; Halperin, E.; Peer, I. Compositional Lotka-Volterra describes microbial dynamics in the simplex. *Plos Comput Biol* 2020, *16* (5). DOI: ARTN e1007917

10.1371/journal.pcbi.1007917.

(183) van der Ark, K. C. H.; van Heck, R. G. A.; Dos Santos, V. A. P. M.; Belzer, C.; de Vos, W. M. More than just a gut feeling: constraint-based genome-scale metabolic models for predicting functions of human intestinal microbes. *Microbiome* 2017, *5*. DOI: ARTN 78

10.1186/s40168-017-0299-x.

(184) Goyal, A.; Wang, T.; Dubinkina, V.; Maslov, S. Ecology-guided prediction of cross-feeding interactions in the human gut microbiome. *Nature Communications* 2021, *12* (1). DOI: ARTN 1335

10.1038/s41467-021-21586-6.

(185) Tartaglia, M.; Bastida, F.; Sciarrillo, R.; Guarino, C. Soil Metaproteomics for the Study of the Relationships Between Microorganisms and Plants: A Review of Extraction Protocols and Ecological Insights. *International Journal of Molecular Sciences* 2020, *21* (22). DOI: ARTN 8455

10.3390/ijms21228455.

(186) Poretsky, R.; Rodriguez, R. L.; Luo, C.; Tsementzi, D.; Konstantinidis, K. T. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One* 2014, 9 (4), e93827. DOI: 10.1371/journal.pone.0093827 From NLM Medline.

(187) Stolpovsky, K.; Martinez-Lavanchy, P.; Heipieper, H. J.; Van Cappellen, P.; Thullner, M. Incorporating dormancy in dynamic microbial community models. *Ecol Model* 2011, 222 (17), 3092-3102. DOI: 10.1016/j.ecolmodel.2011.07.006.

(188) Bowsher, A. W.; Kearns, P. J.; Shade, A. 16S rRNA/rRNA Gene Ratios and Cell Activity Staining Reveal Consistent Patterns of Microbial Activity in Plant-Associated Soil. *Msystems* 2019, 4 (2). DOI: ARTN e00003-19

10.1128/mSystems.00003-19.

(189) Kleiner, M.; Thorson, E.; Sharp, C. E.; Dong, X.; Liu, D.; Li, C.; Strous, M. Assessing species biomass contributions in microbial communities via metaproteomics. *Nat Commun* 2017, 8 (1), 1558. DOI: 10.1038/s41467-017-01544-x.

(190) Kleiner, M. Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities. *Msystems* 2019, 4 (3). DOI: ARTN e00115-19

10.1128/mSystems.00115-19.

(191) Zhou, J. Z.; Liu, W. Z.; Deng, Y.; Jiang, Y. H.; Xue, K.; He, Z. L.; Van Nostrand, J. D.; Wu, L. Y.; Yang, Y. F.; Wang, A. J. Stochastic Assembly Leads to Alternative Communities with Distinct Functions in a Bioreactor Microbial Community. *Mbio* 2013, 4 (2). DOI: ARTN e00584-12

10.1128/mBio.00584-12.

(192) Shrestha, H. K.; Appidi, M. R.; Villalobos Solis, M. I.; Wang, J.; Carper, D. L.; Burdick, L.; Pelletier, D. A.; Doktycz, M. J.; Hettich, R. L.; Abraham, P. E. Metaproteomics reveals insights into microbial structure, interactions, and dynamic regulation in defined communities as they respond to environmental disturbance. *BMC Microbiol* 2021, 21 (1), 308. DOI: 10.1186/s12866-021-02370-4.

(193) Wang, J.; Carper, D. L.; Burdick, L. H.; Shrestha, H. K.; Appidi, M. R.; Abraham, P. E.; Timm, C. M.; Hettich, R. L.; Pelletier, D. A.; Doktycz, M. J. Formation, characterization and modeling of emergent synthetic microbial communities. *Comput Struct Biotechnol J* 2021, 19, 1917-1927. DOI: 10.1016/j.csbj.2021.03.034.

(194) Gao, C. H.; Cao, H.; Cai, P.; Sorensen, S. J. The initial inoculation ratio regulates bacterial coculture interactions and metabolic capacity. *ISME J* 2021, 15 (1), 29-40. DOI: 10.1038/s41396-020-00751-7 From NLM Medline.

- (195) Debray, R.; Herbert, R. A.; Jaffe, A. L.; Crits-Christoph, A.; Power, M. E.; Koskella, B. Priority effects in microbiome assembly. *Nat Rev Microbiol* 2022, 20 (2), 109-121. DOI: 10.1038/s41579-021-00604-w From NLM Medline.
- (196) Wang, J.; Appidi, M. R.; Burdick, L. H.; Abraham, P. E.; Hettich, R. L.; Pelletier, D. A.; Doktycz, M. J. Formation of a constructed microbial community in a nutrient-rich environment indicates bacterial interspecific competition. *mSystems* 2024, 9 (4), e0000624. DOI: 10.1128/msystems.00006-24 From NLM Medline.
- (197) Kall, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* 2007, 4 (11), 923-925. DOI: 10.1038/nmeth1113.
- (198) Darzi, Y.; Letunic, I.; Bork, P.; Yamada, T. iPath3.0: interactive pathways explorer v3. *Nucleic Acids Res* 2018, 46 (W1), W510-W513. DOI: 10.1093/nar/gky299 From NLM Medline.
- (199) Bindea, G.; Mlecnik, B.; Hackl, H.; Charoentong, P.; Tosolini, M.; Kirilovsky, A.; Fridman, W. H.; Pages, F.; Trajanoski, Z.; Galon, J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 2009, 25 (8), 1091-1093. DOI: 10.1093/bioinformatics/btp101.
- (200) Bais, H. P.; Broeckling, C. D.; Vivanco, J. M. Root Exudates Modulate Plant—Microbe Interactions in the Rhizosphere. In *Secondary Metabolites in Soil Ecology*, Karlovsky, P. Ed.; Springer Berlin Heidelberg, 2008; pp 241-252.
- (201) Dennis, P. G.; Miller, A. J.; Hirsch, P. R. Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? *Fems Microbiology Ecology* 2010, 72 (3), 313-327. DOI: 10.1111/j.1574-6941.2010.00860.x.
- (202) Doornbos, R. F.; van Loon, L. C.; Bakker, P. A. H. M. Impact of root exudates and plant defense signaling on bacterial communities in the rhizosphere. A review. *Agron Sustain Dev* 2012, 32 (1), 227-243. DOI: 10.1007/s13593-011-0028-y.
- (203) O'Banion, B. S.; O'Neal, L.; Alexandre, G.; Lebeis, S. L. Bridging the Gap Between Single-Strain and Community-Level Plant-Microbe Chemical Interactions. *Mol Plant Microbe In* 2020, 33 (2), 124-134. DOI: 10.1094/Mpmi-04-19-0115-Cr.
- (204) Goldford, J. E.; Lu, N.; Bajic, D.; Estrela, S.; Tikhonov, M.; Sanchez-Gorostiaga, A.; Segre, D.; Mehta, P.; Sanchez, A. Emergent simplicity in microbial community assembly. *Science* 2018, 361 (6401), 469-474. DOI: 10.1126/science.aat1168.
- (205) Mello, B. L.; Alessi, A. M.; McQueen-Mason, S.; Bruce, N. C.; Polikarpov, I. Nutrient availability shapes the microbial community structure in sugarcane bagasse compost-derived consortia. *Sci Rep* 2016, 6, 38781. DOI: 10.1038/srep38781 From NLM Medline.

- (206) Nobori, T.; Wang, Y.; Wu, J.; Stolze, S. C.; Tsuda, Y.; Finkemeier, I.; Nakagami, H.; Tsuda, K. Multidimensional gene regulatory landscape of a bacterial pathogen in plants. *Nat Plants* 2020, 6 (7), 883-896. DOI: 10.1038/s41477-020-0690-7 From NLM Medline.
- (207) Hirsch, P. R.; Mauchline, T. H. Who's who in the plant root microbiome? *Nat Biotechnol* 2012, 30 (10), 961-962. DOI: 10.1038/nbt.2387 From NLM Medline.
- (208) Ruger, L.; Feng, K.; Dumack, K.; Freudenthal, J.; Chen, Y.; Sun, R.; Wilson, M.; Yu, P.; Sun, B.; Deng, Y.; et al. Assembly Patterns of the Rhizosphere Microbiome Along the Longitudinal Root Axis of Maize (*Zea mays* L.). *Front Microbiol* 2021, 12, 614501. DOI: 10.3389/fmicb.2021.614501 From NLM PubMed-not-MEDLINE.
- (209) Bonkowski, M.; Tarkka, M.; Razavi, B. S.; Schmidt, H.; Blagodatskaya, E.; Koller, R.; Yu, P.; Knief, C.; Hochholdinger, F.; Vetterlein, D. Spatiotemporal Dynamics of Maize (*Zea mays* L.) Root Growth and Its Potential Consequences for the Assembly of the Rhizosphere Microbiota. *Front Microbiol* 2021, 12, 619499. DOI: 10.3389/fmicb.2021.619499 From NLM PubMed-not-MEDLINE.
- (210) Iijima, M.; Griffiths, B.; Bengough, A. G. Sloughing of cap cells and carbon exudation from maize seedling roots in compacted sand. *New Phytol* 2000, 145 (3), 477-482. DOI: 10.1046/j.1469-8137.2000.00595.x From NLM PubMed-not-MEDLINE.
- (211) Knee, E. M.; Gong, F. C.; Gao, M. S.; Teplitski, M.; Jones, A. R.; Foxworthy, A.; Mort, A. J.; Bauer, W. D. Root mucilage from pea and its utilization by rhizosphere bacteria as a sole carbon source. *Mol Plant Microbe In* 2001, 14 (6), 775-784. DOI: Doi 10.1094/Mpmi.2001.14.6.775.
- (212) Walker, T. S.; Bais, H. P.; Grotewold, E.; Vivanco, J. M. Root exudation and rhizosphere biology. *Plant Physiol* 2003, 132 (1), 44-51. DOI: 10.1104/pp.102.019661 From NLM Medline.
- (213) Ramirez-Villacis, D. X.; Finkel, O. M.; Salas-Gonzalez, I.; Fitzpatrick, C. R.; Dangl, J. L.; Jones, C. D.; Leon-Reyes, A. Root Microbiome Modulates Plant Growth Promotion Induced by Low Doses of Glyphosate. *mSphere* 2020, 5 (4). DOI: 10.1128/mSphere.00484-20 From NLM Medline.
- (214) Long, S.; Yang, Y.; Shen, C.; Wang, Y.; Deng, A.; Qin, Q.; Qiao, L. Metaproteomics characterizes human gut microbiome function in colorectal cancer. *NPJ Biofilms Microbiomes* 2020, 6 (1), 14. DOI: 10.1038/s41522-020-0123-4.
- (215) Aufrecht, J. A.; Timm, C. M.; Bible, A.; Morrell-Falvey, J. L.; Pelletier, D. A.; Doktycz, M. J.; Retterer, S. T. Quantifying the Spatiotemporal Dynamics of Plant Root Colonization by Beneficial Bacteria in a Microfluidic Habitat. *Adv Biosyst* 2018, 2 (6). DOI: ARTN 1800048 10.1002/adbi.201800048.
- (216) 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.; et al. A Carotenoid-Deficient Mutant in

Pantoea sp YR343, a Bacteria Isolated from the Rhizosphere of Populus deltoides, Is Defective in Root Colonization. *Front Microbiol* 2016, 7. DOI: ARTN 491

10.3389/fmicb.2016.00491.

(217) Blair, P. M.; Land, M. L.; Piatek, M. J.; Jacobson, D. A.; Lu, T. Y. S.; Doktycz, M. J.; Pelletier, D. A. Exploration of the Biosynthetic Potential of the Populus Microbiome. *Msystems* 2018, 3 (5). DOI: ARTN e00045-18

10.1128/mSystems.00045-18.

(218) Brown, S. D.; Utturkar, S. M.; Klingeman, D. M.; Johnson, C. M.; Martin, S. L.; Land, M. L.; Lu, T. Y. S.; Schadt, C. W.; Doktycz, M. J.; Pelletier, D. A. 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* 2012, 194 (21), 5991-5993. DOI: 10.1128/Jb.01243-12.

(219) 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. Elucidating Duramycin's Bacterial Selectivity and Mode of Action on the Bacterial Cell Envelope. *Front Microbiol* 2018, 9. DOI: ARTN 219

10.3389/fmicb.2018.00219.

(220) Levy, A.; Gonzalez, I. S.; Mittelviefhaus, M.; Clingenpeel, S.; Paredes, S. H.; Miao, J. M.; Wang, K. R.; Devescovi, G.; Stillman, K.; Monteiro, F.; et al. Genomic features of bacterial adaptation to plants. *Nat Genet* 2018, 50 (1), 138-+. DOI: 10.1038/s41588-017-0012-9.

(221) Schaefer, A. L.; Lappala, C. R.; Morlen, R. P.; Pelletier, D. A.; Lu, T.-Y. S.; Lankford, P. K.; Harwood, C. S.; Greenberg, E. P. LuxR-and LuxI-type quorum-sensing circuits are prevalent in members of the Populus deltoides microbiome. *Appl. Environ. Microbiol.* 2013, 79 (18), 5745-5752.

(222) Timm, C. M.; Pelletier, D. A.; Jawdy, S. S.; Gunter, L. E.; Henning, J. A.; Engle, N.; Aufrecht, J.; Gee, E.; Nookaew, I.; Yang, Z. M.; et al. Two Poplar-Associated Bacterial Isolates Induce Additive Favorable Responses in a Constructed Plant-Microbiome System. *Front Plant Sci* 2016, 7. DOI: ARTN 497

10.3389/fpls.2016.00497.

(223) Reasoner, D. J.; Geldreich, E. E. A new medium for the enumeration and subculture of bacteria from potable water. *Appl Environ Microbiol* 1985, 49 (1), 1-7. DOI: 10.1128/aem.49.1.1-7.1985.

(224) Masuda, T.; Tomita, M.; Ishihama, Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res* 2008, 7 (2), 731-740. DOI: 10.1021/pr700658q
From NLM Medline.

(225) Orsburn, B. C. Proteome Discoverer-A Community Enhanced Data Processing Suite for Protein Informatics. *Proteomes* 2021, 9 (1). DOI: ARTN 15

10.3390/proteomes9010015.

(226) Oksanen, J. B., F. Guillaume & Friendly, Michael & Kindt, Roeland & Legendre, Pierre & McGlinn, Daniel & Minchin, Peter & O'Hara, RB & Simpson, Gavin & Solymos, Peter & Stevens, MHH & Szöcs, Eduard & Wagner, Helene. vegan community ecology package version 2.5-7. 2020.

(227) Knief, C.; Delmotte, N.; Chaffron, S.; Stark, M.; Innerebner, G.; Wassmann, R.; von Mering, C.; Vorholt, J. A. Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *Isme Journal* 2012, 6 (7), 1378-1390. DOI: 10.1038/ismej.2011.192.

(228) Hara, S.; Morikawa, T.; Wasai, S.; Kasahara, Y.; Koshiba, T.; Yamazaki, K.; Fujiwara, T.; Tokunaga, T.; Minamisawa, K. Identification of Nitrogen-Fixing Bradyrhizobium Associated With Roots of Field-Grown Sorghum by Metagenome and Proteome Analyses. *Front Microbiol* 2019, 10. DOI: ARTN 407

10.3389/fmicb.2019.00407.

(229) Mattarozzi, M.; Di Zinno, J.; Montanini, B.; Manfredi, M.; Marengo, E.; Fornasier, F.; Ferrarini, A.; Careri, M.; Visioli, G. Biostimulants applied to maize seeds modulate the enzymatic activity and metaproteome of the rhizosphere. *Appl Soil Ecol* 2020, 148. DOI: ARTN 103480

10.1016/j.apsoil.2019.103480.

(230) Munoz-Ucros, J.; Zwetsloot, M. J.; Cuellar-Gempeler, C.; Bauerle, T. L. Spatiotemporal patterns of rhizosphere microbiome assembly: From ecological theory to agricultural application. *J Appl Ecol* 2021, 58 (5), 894-904. DOI: 10.1111/1365-2664.13850.

(231) Mattarozzi, M.; Manfredi, M.; Montanini, B.; Gosetti, F.; Sanangelantoni, A. M.; Marengo, E.; Careri, M.; Visioli, G. A metaproteomic approach dissecting major bacterial functions in the rhizosphere of plants living in serpentine soil. *Anal Bioanal Chem* 2017, 409 (9), 2327-2339. DOI: 10.1007/s00216-016-0175-8.

(232) Estenson, K.; Hurst, G. B.; Standaert, R. F.; Bible, A. N.; Garcia, D.; Chourey, K.; Doktycz, M. J.; Morrell-Falvey, J. L. Characterization of Indole-3-acetic Acid Biosynthesis and the Effects of This Phytohormone on the Proteome of the Plant-Associated Microbe *Pantoea* sp YR343. *J Proteome Res* 2018, 17 (4), 1361-1374. DOI: 10.1021/acs.jproteome.7b00708.

(233) Kumar, S. V.; Abraham, P. E.; Hurst, G. B.; Chourey, K.; Bible, A. N.; Hettich, R. L.; Doktycz, M. J.; Morrell-Falvey, J. L. A carotenoid-deficient mutant of the plant-associated microbe *Pantoea* sp. YR343 displays an altered membrane proteome. *Sci Rep-Uk* 2020, 10 (1). DOI: ARTN 14985

10.1038/s41598-020-71672-w.

(234) Hemkemeyer, M.; Pronk, G. J.; Heister, K.; Kogel-Knabner, I.; Martens, R.; Tebbe, C. C. Artificial soil studies reveal domain-specific preferences of microorganisms for the colonisation of different soil minerals and particle size fractions. *FEMS Microbiol Ecol* 2014, 90 (3), 770-782. DOI: 10.1111/1574-6941.12436.

(235) Poole, P. Shining a light on the dark world of plant root-microbe interactions. *P Natl Acad Sci USA* 2017, 114 (17), 4281-4283. DOI: 10.1073/pnas.1703800114.

(236) Yee, M. O.; Kim, P.; Li, Y. F.; Singh, A. K.; Northen, T. R.; Chakraborty, R. Specialized Plant Growth Chamber Designs to Study Complex Rhizosphere Interactions. *Front Microbiol* 2021, 12. DOI: ARTN 625752

10.3389/fmicb.2021.625752.

(237) Abreu, N. A.; Taga, M. E. Decoding molecular interactions in microbial communities. *Fems Microbiol Rev* 2016, 40 (5), 648-663. DOI: 10.1093/femsre/fuw019.

(238) Zhalnina, K.; Louie, K. B.; Hao, Z.; Mansoori, N.; da Rocha, U. N.; Shi, S.; Cho, H.; Karaoz, U.; Loque, D.; Bowen, B. P.; et al. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat Microbiol* 2018, 3 (4), 470-480. DOI: 10.1038/s41564-018-0129-3.

(239) Reinhold-Hurek, B.; B nger, W.; Burbano, C. S.; Sabale, M.; Hurek, T. Roots Shaping Their Microbiome: Global Hotspots for Microbial Activity. *Annu Rev Phytopathol* 2015, 53, 403-+. DOI: 10.1146/annurev-phyto-082712-102342.

(240) Bais, H. P.; Weir, T. L.; Perry, L. G.; Gilroy, S.; Vivanco, J. M. The role of root exudates in rhizosphere interactions with plants and other organisms. *Annu Rev Plant Biol* 2006, 57, 233-266. DOI: 10.1146/annurev.arplant.57.032905.105159.

(241) Liu, S.; Tao, C.; Zhang, L.; Wang, Z.; Xiong, W.; Xiang, D.; Sheng, O.; Wang, J.; Li, R.; Shen, Z.; et al. Plant pathogen resistance is mediated by recruitment of specific rhizosphere fungi. *Isme j* 2023, 17 (6), 931-942. DOI: 10.1038/s41396-023-01406-z From NLM.

(242) Canarini, A.; Kaiser, C.; Merchant, A.; Richter, A.; Wanek, W. Root Exudation of Primary Metabolites: Mechanisms and Their Roles in Plant Responses to Environmental Stimuli (vol 10, 157, 2019). *Frontiers in Plant Science* 2019, 10. DOI: ARTN 420

10.3389/fpls.2019.00420.

(243) Saleem, M.; Law, A. D.; Sahib, M. R.; Pervaiz, Z. H.; Zhang, Q. M. Impact of root system architecture on rhizosphere and root microbiome. *Rhizosphere-Neth* 2018, 6, 47-51. DOI: 10.1016/j.rhisph.2018.02.003.

- (244) Szoboszlay, M.; Lambers, J.; Chappell, J.; Kupper, J. V.; Moe, L. A.; McNear, D. H. Comparison of root system architecture and rhizosphere microbial communities of Balsas teosinte and domesticated corn cultivars. *Soil Biol Biochem* 2015, *80*, 34-44. DOI: 10.1016/j.soilbio.2014.09.001.
- (245) Garcia, J.; Kao-Kniffin, J. Microbial Group Dynamics in Plant Rhizospheres and Their Implications on Nutrient Cycling. *Front Microbiol* 2018, *9*. DOI: ARTN 1516 10.3389/fmicb.2018.01516.
- (246) Lugtenberg, B.; Kamilova, F. Plant-growth-promoting rhizobacteria. *Annu Rev Microbiol* 2009, *63*, 541-556. DOI: 10.1146/annurev.micro.62.081307.162918.
- (247) Patten, C. L.; Glick, B. R. Bacterial biosynthesis of indole-3-acetic acid. *Can J Microbiol* 1996, *42* (3), 207-220.
- (248) Glick, B. R. Plant growth-promoting bacteria: mechanisms and applications. *Scientifica (Cairo)* 2012, *2012*, 963401. DOI: 10.6064/2012/963401 From NLM PubMed-not-MEDLINE.
- (249) Weiland-Brauer, N. Friends or Foes-Microbial Interactions in Nature. *Biology (Basel)* 2021, *10* (6). DOI: 10.3390/biology10060496.
- (250) Nuccio, E. E.; Starr, E.; Karaoz, U.; Brodie, E. L.; Zhou, J. Z.; Tringe, S. G.; Malmstrom, R. R.; Woyke, T.; Banfield, J. F.; Firestone, M. K.; et al. Niche differentiation is spatially and temporally regulated in the rhizosphere. *Isme Journal* 2020, *14* (4), 999-1014. DOI: 10.1038/s41396-019-0582-x.
- (251) Pathan, S. I.; Vetrovsky, T.; Giagnoni, L.; Datta, R.; Baldrian, P.; Nannipieri, P.; Renella, G. Microbial expression profiles in the rhizosphere of two maize lines differing in N use efficiency. *Plant and Soil* 2018, *433* (1-2), 401-413. DOI: 10.1007/s11104-018-3852-x.
- (252) Bible, A. N.; Chang, M.; Morrell-Falvey, J. L. Identification of a diguanylate cyclase expressed in the presence of plants and its application for discovering candidate gene products involved in plant colonization by *Pantoea* sp. YR343. *PLOS ONE* 2021, *16* (7), e0248607. DOI: 10.1371/journal.pone.0248607.
- (253) 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.; et al. A carotenoid-deficient mutant in *Pantoea* sp. YR343, a bacteria isolated from the rhizosphere of *Populus deltoides*, is defective in root colonization. *Front Microbiol* 2016, *7*, 491. DOI: 10.3389/fmicb.2016.00491 From NLM.
- (254) Bai, Y.; Muller, D. B.; Srinivas, G.; Garrido-Oter, R.; Potthoff, E.; Rott, M.; Dombrowski, N.; Munch, P. C.; Spaepen, S.; Remus-Emsermann, M.; et al. Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* 2015, *528* (7582), 364-369. DOI: 10.1038/nature16192.

- (255) Voges, M.; Bai, Y.; Schulze-Lefert, P.; Sattely, E. S. Plant-derived coumarins shape the composition of an Arabidopsis synthetic root microbiome. *Proc Natl Acad Sci U S A* 2019, *116* (25), 12558-12565. DOI: 10.1073/pnas.1820691116.
- (256) 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.; Veach, A. M.; et al. Cultivating the Bacterial Microbiota of *Populus* Roots. *mSystems* 2021, *6* (3), 10.1128/msystems.01306-01320. DOI: doi:10.1128/msystems.01306-20.
- (257) Jayaraman, D.; Forshey, K. L.; Grimsrud, P. A.; Ane, J. M. Leveraging proteomics to understand plant-microbe interactions. *Front Plant Sci* 2012, *3*, 44. DOI: 10.3389/fpls.2012.00044.
- (258) Appidi, M. R.; Bible, A. N.; Carper, D. L.; Jawdy, S. S.; Giannone, R. J.; Hettich, R. L.; Morrell-Falvey, J.; Abraham, P. E. Development of an Experimental Approach to Achieve Spatially Resolved Plant Root-Associated Metaproteomics Using an Agar-Plate System. *Mol Plant Microbe Interact* 2022, *35* (8), 639-649. DOI: 10.1094/MPMI-01-22-0011-TA.
- (259) Young, K. D. d.; Stankeviciute, G.; Klein, E. A. Sugar-Phosphate Metabolism Regulates Stationary-Phase Entry and Stalk Elongation in *Caulobacter crescentus*. *J Bacteriol* 2020, *202* (4), 10.1128/jb.00468-00419. DOI: doi:10.1128/jb.00468-19.
- (260) Cantalapiedra, C. P.; Hernández-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol* 2021, *38* (12), 5825-5829. DOI: 10.1093/molbev/msab293.
- (261) Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernandez-Plaza, A.; Forslund, S. K.; Cook, H.; Mende, D. R.; Letunic, I.; Rattei, T.; Jensen, L. J.; et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 2019, *47* (D1), D309-D314. DOI: 10.1093/nar/gky1085.
- (262) Sun, L.; Dong, S.; Ge, Y.; Fonseca, J. P.; Robinson, Z. T.; Mysore, K. S.; Mehta, P. DiVenn: An Interactive and Integrated Web-Based Visualization Tool for Comparing Gene Lists. *Front Genet* 2019, *10*, 421. DOI: 10.3389/fgene.2019.00421 From NLM PubMed-not-MEDLINE.
- (263) Dastogeer, K. M. G.; Tumpa, F. H.; Sultana, A.; Akter, M. A.; Chakraborty, A. Plant microbiome-an account of the factors that shape community composition and diversity. *Curr Plant Biol* 2020, *23*. DOI: ARTN 100161
10.1016/j.cpb.2020.100161.
- (264) Jacoby, R. P.; Chen, L.; Schwier, M.; Koprivova, A.; Kopriva, S. Recent advances in the role of plant metabolites in shaping the root microbiome. *F1000Res* 2020, *9*. DOI: 10.12688/f1000research.21796.1 From NLM Medline.
- (265) Tan, I. S.; Ramamurthi, K. S. Spore formation in *Bacillus subtilis*. *Environ Microbiol Rep* 2014, *6* (3), 212-225. DOI: 10.1111/1758-2229.12130.

- (266) Paredes-Sabja, D.; Setlow, P.; Sarker, M. R. Germination of spores of Bacillales and Clostridiales species: mechanisms and proteins involved. *Trends Microbiol* 2011, 19 (2), 85-94. DOI: 10.1016/j.tim.2010.10.004.
- (267) Feng, H.; Fu, R.; Hou, X.; Lv, Y.; Zhang, N.; Liu, Y.; Xu, Z.; Miao, Y.; Krell, T.; Shen, Q.; et al. Chemotaxis of Beneficial Rhizobacteria to Root Exudates: The First Step towards Root-Microbe Rhizosphere Interactions. *Int J Mol Sci* 2021, 22 (13). DOI: 10.3390/ijms22136655 From NLM Medline.
- (268) Vives-Peris, V.; de Ollas, C.; Gomez-Cadenas, A.; Perez-Clemente, R. M. Root exudates: from plant to rhizosphere and beyond. *Plant Cell Rep* 2020, 39 (1), 3-17. DOI: 10.1007/s00299-019-02447-5 From NLM Medline.
- (269) Gargallo-Garriga, A.; Preece, C.; Sardans, J.; Oravec, M.; Urban, O.; Penuelas, J. Root exudate metabolomes change under drought and show limited capacity for recovery. *Sci Rep-Uk* 2018, 8. DOI: ARTN 12696
10.1038/s41598-018-30150-0.
- (270) Jones, D. L.; Nguyen, C.; Finlay, R. D. Carbon flow in the rhizosphere: carbon trading at the soil-root interface. *Plant Soil* 2009, 321 (1-2), 5-33. DOI: 10.1007/s11104-009-9925-0.
- (271) Haichar, F. E.; Santaella, C.; Heulin, T.; Achouak, W. Root exudates mediated interactions belowground. *Soil Biol Biochem* 2014, 77, 69-80. DOI: 10.1016/j.soilbio.2014.06.017.
- (272) Jones, D. L.; Hodge, A.; Kuzyakov, Y. Plant and mycorrhizal regulation of rhizodeposition. *New Phytol* 2004, 163 (3), 459-480. DOI: 10.1111/j.1469-8137.2004.01130.x From NLM PubMed-not-MEDLINE.
- (273) Sticklen, M. Plant genetic engineering to improve biomass characteristics for biofuels. *Curr Opin Biotechnol* 2006, 17 (3), 315-319. DOI: 10.1016/j.copbio.2006.05.003 From NLM Medline.
- (274) Saberi Riseh, R.; Fathi, F.; Lagzian, A.; Vatankhah, M.; Kennedy, J. F. Modifying lignin: A promising strategy for plant disease control. *International journal of biological macromolecules* 2024, 271 (Pt 2), 132696. DOI: 10.1016/j.ijbiomac.2024.132696 From NLM Medline.
- (275) Garzon, C. D.; Lequart, M.; Charras, Q.; Fournet, F.; Bellenger, L.; Sellier-Richard, H.; Giauffret, C.; Vermerris, W.; Domon, J. M.; Rayon, C. The maize low-lignin brown midrib3 mutant shows pleiotropic effects on photosynthetic and cell wall metabolisms in response to chilling. *Plant Physiol Bioch* 2022, 184, 75-86. DOI: 10.1016/j.plaphy.2022.05.006.
- (276) Seitz, V. A.; McGivern, B. B.; Daly, R. A.; Chaparro, J. M.; Borton, M. A.; Sheflin, A. M.; Kresovich, S.; Shields, L.; Schipanski, M. E.; Wrighton, K. C.; et al. Variation in Root Exudate Composition Influences Soil Microbiome Membership and Function. *Appl Environ Microbiol* 2022, 88 (11), e0022622. DOI: 10.1128/aem.00226-22 From NLM Medline.

- (277) Harman-Ware, A. E.; Happs, R. M.; Macaya-Sanz, D.; Doeppke, C.; Muchero, W.; DiFazio, S. P. Abundance of Major Cell Wall Components in Natural Variants and Pedigrees of *Populus trichocarpa*. *Front Plant Sci* 2022, 13, 757810. DOI: 10.3389/fpls.2022.757810 From NLM PubMed-not-MEDLINE.
- (278) Muchero, W.; Guo, J.; DiFazio, S. P.; Chen, J. G.; Ranjan, P.; Slavov, G. T.; Gunter, L. E.; Jawdy, S.; Bryan, A. C.; Sykes, R.; et al. High-resolution genetic mapping of allelic variants associated with cell wall chemistry in *Populus*. *BMC Genomics* 2015, 16 (1), 24. DOI: 10.1186/s12864-015-1215-z From NLM Medline.
- (279) Zhao, X. Y.; Guo, M. R.; Zhang, T. S.; Bai, S. W.; Meng, Y. F.; Tian, Y. S.; Yang, J. X.; Ma, F. Spatiotemporal dynamics of root exudates drive microbial adaptation mechanisms under day-night alterations in constructed wetlands. *Chem Eng J* 2023, 477. DOI: ARTN 147311
10.1016/j.cej.2023.147311.
- (280) Sasse, J.; Martinoia, E.; Northen, T. Feed Your Friends: Do Plant Exudates Shape the Root Microbiome? *Trends in Plant Science* 2018, 23 (1), 25-41. DOI: 10.1016/j.tplants.2017.09.003.
- (281) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 2007, 3 (3), 211-221. DOI: 10.1007/s11306-007-0082-2 From NLM PubMed-not-MEDLINE.
- (282) Pang, Z.; Lu, Y.; Zhou, G.; Hui, F.; Xu, L.; Viau, C.; Spigelman, A. F.; MacDonald, P. E.; Wishart, D. S.; Li, S.; et al. MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucleic Acids Res* 2024, 52 (W1), W398-W406. DOI: 10.1093/nar/gkae253 From NLM Medline.
- (283) Mavrodi, O. V.; McWilliams, J. R.; Peter, J. O.; Berim, A.; Hassan, K. A.; Elbourne, L. D. H.; LeTourneau, M. K.; Gang, D. R.; Paulsen, I. T.; Weller, D. M.; et al. Root Exudates Alter the Expression of Diverse Metabolic, Transport, Regulatory, and Stress Response Genes in Rhizosphere *Pseudomonas*. *Front Microbiol* 2021, 12, 651282. DOI: 10.3389/fmicb.2021.651282 From NLM PubMed-not-MEDLINE.
- (284) Zulfiqar, U.; Yasmin, A.; Fariq, A. Metabolites produced by inoculated *Vigna radiata* during bacterial assisted phytoremediation of Pb, Ni and Cr polluted soil. *PLoS One* 2022, 17 (11), e0277101. DOI: 10.1371/journal.pone.0277101 From NLM Medline.
- (285) Nxumalo, C. I.; Ngidi, L. S.; Shandu, J. S. E.; Maliehe, T. S. Isolation of endophytic bacteria from the leaves of *Anredera cordifolia* CIX1 for metabolites and their biological activities. *BMC Complement Med Ther* 2020, 20 (1), 300. DOI: 10.1186/s12906-020-03095-z From NLM Medline.
- (286) Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; et al. ClassyFire: automated chemical

classification with a comprehensive, computable taxonomy. *J Cheminform* 2016, 8, 61. DOI: 10.1186/s13321-016-0174-y From NLM PubMed-not-MEDLINE.

(287) Li, L. L.; Li, Z.; Lou, Y.; Meiners, S. J.; Kong, C. H. (–)-Loliolide is a general signal of plant stress that activates jasmonate-related responses. *New Phytologist* 2023, 238 (5), 2099-2112.

(288) Murata, M.; Nakai, Y.; Kawazu, K.; Ishizaka, M.; Kajiwar, H.; Abe, H.; Takeuchi, K.; Ichinose, Y.; Mitsuhara, I.; Mochizuki, A.; et al. Loliolide, a Carotenoid Metabolite, Is a Potential Endogenous Inducer of Herbivore Resistance. *Plant Physiol* 2019, 179 (4), 1822-1833. DOI: 10.1104/pp.18.00837.

(289) Cook, D.; Donzelli, B. G.; Creamer, R.; Baucom, D. L.; Gardner, D. R.; Pan, J.; Moore, N.; Krasnoff, S. B.; Jaromczyk, J. W.; Schardl, C. L. Swainsonine biosynthesis genes in diverse symbiotic and pathogenic fungi. *G3: Genes, Genomes, Genetics* 2017, 7 (6), 1791-1797.

(290) Du, L. Y.; Zhao, M.; Xu, J.; Qian, D. W.; Jiang, S.; Shang, E. X.; Guo, J. M.; Duan, J. A. Analysis of the metabolites of isorhamnetin 3-O-glucoside produced by human intestinal flora in vitro by applying ultraperformance liquid chromatography/quadrupole time-of-flight mass spectrometry. *J Agric Food Chem* 2014, 62 (12), 2489-2495. DOI: 10.1021/jf405261a From NLM Medline.

(291) Kalai, F. Z.; Boulaaba, M.; Ferdousi, F.; Isoda, H. Effects of Isorhamnetin on Diabetes and Its Associated Complications: A Review of In Vitro and In Vivo Studies and a Post Hoc Transcriptome Analysis of Involved Molecular Pathways. *Int J Mol Sci* 2022, 23 (2). DOI: 10.3390/ijms23020704 From NLM Medline.

(292) Turoczi, B.; Bakonyi, J.; Szabo, K. A.; Balint, J.; Mathe, I.; Lanyi, S.; Balog, A. In Vitro and In Vivo Effect of Poplar Bud Extracts on *Phytophthora infestans*: A New Effective Biological Method in Potato Late Blight Control. *Plants (Basel)* 2020, 9 (2). DOI: 10.3390/plants9020217 From NLM PubMed-not-MEDLINE.

(293) Harbilas, D.; Vallerand, D.; Brault, A.; Saleem, A.; Arnason, J. T.; Musallam, L.; Haddad, P. S. *Populus balsamifera* Extract and Its Active Component Salicortin Reduce Obesity and Attenuate Insulin Resistance in a Diet-Induced Obese Mouse Model. *Evid Based Complement Alternat Med* 2013, 2013, 172537. DOI: 10.1155/2013/172537 From NLM PubMed-not-MEDLINE.

(294) Jung, H. W.; Tschaplinski, T. J.; Wang, L.; Glazebrook, J.; Greenberg, J. T. Priming in systemic plant immunity. *Science* 2009, 324 (5923), 89-91. DOI: 10.1126/science.1170025 From NLM Medline.

(295) Korenblum, E.; Dong, Y. H.; Szymanski, J.; Panda, S.; Jozwiak, A.; Massalha, H.; Meir, S.; Rogachev, I.; Aharoni, A. Rhizosphere microbiome mediates systemic root metabolite exudation by root-to-root signaling. *P Natl Acad Sci USA* 2020, 117 (7), 3874-3883. DOI: 10.1073/pnas.1912130117.

VITA

Manasa Reddy Appidi was born and raised in Hyderabad, India. Manasa graduated with a Bachelor of Science degree with a specialization in Microbiology and Genetics from Nizam college, Hyderabad, India. After completing her undergraduate degree, she pursued a master's degree in biomedical basis of disease at Sheffield Hallam University, United Kingdom. Then she moved to the United States and obtained a master's degree in molecular biology from Long Island University, Brooklyn, New York. Following the completion of her master's degree in 2016, She spent three years working as an analytical scientist in New Jersey. In 2019, Manasa began her Ph.D. studies in the Genome Science and Technology program at University of Tennessee-Knoxville, where she completed her PhD dissertation work in the Bioanalytical Mass Spectrometry group at Oak Ridge National Laboratory under the supervision of Dr. Robert L. Hettich and Dr. Paul Abraham.