Characterization of Diverse Mechanisms of Salicin Degradation in *Populus* Microbiome Isolates

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Sanjeev Dahal

December 2019

Dedication

I would like to dedicate this dissertation, first and foremost to my family. My father, Mr. Ram Sharan Dahal and my mother, Mrs. Bhagawati Simkhada Dahal always dreamed of providing me with the best education, and they worked hard to bring me opportunities and resources to pursue my dream of becoming a scientist. My brother, Sushant Dahal has always been there as a friend to support me throughout these years. This work is a result of constant encouragement from all the family members and in-laws, and I feel grateful for having them in my life.

I also would like to give my deepest gratitude to all the friends here in Knoxville as well as from my alma mater, University of New Orleans. They have given me a second home here in the States and provided me courage and company all the time. Without them, this journey would be very difficult. From University of New Orleans, I would also like to include Dr. Mary J. Clancy who was my first mentor and introduced me to the wonderful world of molecular biology and genetics.

Finally, I dedicate my dissertation and express special thanks to my dear wife Susmita Sharma Bajgain for supporting me and encouraging me to successfully complete my PhD. You are my best friend and my inspiration.

Acknowledgement

Numerous people spring to mind, but the first person I would like to thank is my advisor Dr. Dale A. Pelletier. I will forever be indebted to him for teaching me the value of patience and endurance in this field. His guidance on the numerous topics in the thesis was always welcoming. Even when I was at some of the lowest points of my life, he was always there to provide me with kind words and encourage me to move on. I could always go and knock on his door, and he would listen and provide me with positive criticism and ways to make my work better. I believe I have grown not only as a researcher but also as a person because of him.

There are several people in the lab to whom I would like to express my gratitude. Tse-Yuan Lu and Leah Burdick have helped me perform several experiments and constantly guided me through laboratory requirements. Previous and current post-doctoral researchers – Collin Timm, Patricia Blair and Dana Carper have also helped me with experiments/analysis and to adjust in the lab. Finally, Aditya Barde, who is a recent Master's student from the lab, has been a good friend and helped me during enduring times.

Some of my collaborators also need to be mentioned here. Dr. Gregory B. Hurst, Dr. Robert L. Hettich and Dr. Suresh Poudel all have been helpful in proteome data generation and analysis. Dr. Timothy J. Tschaplinski, Nancy L. Engle and David Reeves were instrumental in metabolomic assays. Finally, I am thankful to David Garcia who has guided me in several experiments including cell-free reactions and extract preparations. Without the help from the collaborators, I would not be able to make my work as polished as it is.

I would also like to express my appreciation to all the committee members – Dr. Sarah L. Lebeis, Dr. Jennifer L. Morrell-Falvey, Dr. Cong T. Trinh and Dr. Daniel A. Jacobson who have been helpful in making sure that my work is as focused as possible. They have been kind and

critical when required and encouraged me to think like a scientist. Without their guidance, this work would not have been possible. I would also like to thank Dr. Albrecht von Arnim who has provided invaluable advices throughout the years.

Finally, I would like to thank everyone in the Genome Science and Technology program especially Teresa Yeatts who has been helpful throughout my stay here. I am also thankful to all the friends and colleagues in the department as well as outside the department who have made my graduate years livelier by bringing me laughs and cheers whenever they could. I am forever grateful.

Abstract

Investigations of metabolic interactions between plant and microbes are vital to designing approaches to promote plant growth and health. *Populus,* a potential biofuel crop, interacts with its microbiome through numerous metabolites, but research on phenolic glycosides (PGs) is still lacking despite the fact that *Populus* allocates significant amount of carbon to PGs. Therefore, this study asked the following questions: what is the prevalence of microbiome isolates that utilize PGs as growth substrates? what is the diversity in the mechanisms of PG utilization?

For this study, salicin was used as the model PG substrate. Four naturally occurring salicin degradation systems designated – Bgl, Sal, Lac and Cbg that have been described in literature were investigated in 407 *Populus* bacterial isolates with available genome sequences. These four systems appear to be present in phylogenetically diverse microbial associates suggesting independent evolution of these mechanisms. Following this, four bacterial strains – *Caulobacter* sp. AP07, *Citrobacter* sp. YR018, *Agrobacterium rhizogenes* OK036 and *Rahnella* sp. OV744 were selected for further study. *Caulobacter* sp. AP07 was demonstrated to possess two different salicin systems – Lac and Sal that were both found to be active during growth on salicin minimal media. *Citrobacter* sp. YR018, on the other hand, utilized salicin most likely by gaining mutations. A few possible loci that could have been mutated are proposed in this study.

Agrobacterium rhizogenes OK036 was investigated for its unique ability to degrade both the glucosyl and salicyl moieties of salicin unlike other organisms in this study that only utilize glucosyl group. Using proteomic, metabolomic and other experimental assays, *Agrobacterium* was demonstrated to utilize salicin by first cleaving glucose off of salicyl alcohol, which is then channeled into gentisate pathway. A few unique features of this novel pathway were explored further using bioinformatic approaches.

Finally, salicin co-metabolism by co-culture of *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16 was elucidated. Using proteomic and metabolomic along with other quantitative assays OV744 and GM16 were demonstrated to be involved in a unidirectional cross-feeding interaction during growth on salicin minimal media. The co-culture study illustrates a possible mechanism *Populus* might utilize to fine-tune its microbial colonization process.

Table of contents

Chapter 1: Introduction *1*

Plant-microbe interactions **2**

Benefits of plant-microbe interactions and their studies **3**

***Populus* as a model for plant-microbe interaction** **5**

Phenolic glycosides (PGs) **6**

Importance of PGs **8**

Biosynthesis of PGs **9**

Degradation mechanisms of PGs in *Populus*-associated microorganisms and their possible role in shaping the *Populus* microbiome **10**

Chapter 2: Phylogenetic and functional diversity in salicin degrading Populus bacterial isolates *13*

Introduction **14**

Methods and materials **17**

Bacterial strains and culture medium 17

HPLC metabolite analysis of culture supernatant 18

Sample preparation for proteome measurements 18

MS data analysis and evaluation 20

Bioinformatic analysis of salicin degradation genes and gene clusters 20

Data analysis and visualization 21

Results **21**

45 strains are predicted to utilize salicin through either Bgl, Sal or Lac systems 21

Experimental screening demonstrates that approximately 45% of the *Populus* isolates can degrade salicin 24

Caulobacter sp. AP07 utilizes salicin and releases salicyl alcohol 26

.....27

Proteome data support the hypothesis that *Caulobacter* sp. AP07 utilizes salicin through Lac and Sal systems 28

Citrobacter sp. YR018 grows on salicin minimal media but only after prolonged incubation 28

Growth of *Citrobacter* on salicin is most likely due to mutational activation 30

Possible mutational activation sites on *Citrobacter* are proposed 31

Discussion 32

Chapter 3: The salicin degradation pathway in Populus-isolate Agrobacterium rhizogenes OK036 39

Introduction 40

Methods and materials 42

Bacterial strains and culture medium 42

High performance liquid chromatography (HPLC) metabolite analysis 42

Sample preparation for proteome measurements 43

MS data analysis and evaluation 44

Preparation of cell lysates for metabolite detection 45

Sample preparation for metabolite analysis 46

Nano-LC MS/MS 46

LC-MS/MS metabolite data analysis 47

Genomic data analysis 48

Results 49

Agrobacterium rhizogenes OK036 is able to completely utilize salicin for growth 49

A unique salicin degradation pathway is hypothesized for *Agrobacterium rhizogenes* OK036 through proteomic measurements 50

Agrobacterium rhizogenes OK036 cannot grow on intermediates of the proposed pathway 54

Metabolomics assay corroborates the presence of the proposed pathway 55

Identification of the prevalence of the proposed OK036 salicin degradation pathway in other *Populus*-associated strains 57

Discussion 57

Chapter 4: Mechanism of unidirectional by-product cross-feeding of salicyl alcohol between Populus microbiome isolates 64

Introduction 65

Unidirectional by-product cross-feeding 65

Bidirectional by-product cross-feeding 65

By-product reciprocity 65

Bidirectional cooperative cross-feeding 65

Unidirectional cooperative cross-feeding 66

Materials and methods 68

Bacterial strains and culture medium 68

Primer design, DNA isolation and microbial quantification 68

Gas chromatography - mass spectrometry (GC-MS) metabolome analysis 69

Sample preparation for proteome measurements 70

Liquid chromatography/tandem mass spectrometry (LC MS/MS) analysis 71

Proteomics data analysis 72

Recombinant plasmid construction for complementation assay 74

Preparation of electrocompetent cells for complementation assay 75

Complementation assay 76

Bioinformatic analysis 76

Results 77

Rahnella sp. OV744 grows on salicin by utilizing the glucosyl moiety, and in the process secretes salicyl alcohol 77

Proteomics and metabolomic investigation of salicin degradation in *Rahnella* 78

Rahnella sp. OV744 phosphoglucosidase can functionally replace the phosphoglucosidase function of *Pantoea* sp. YR343 mutant....80
Pseudomonas sp. GM16 utilizes salicyl alcohol....82
Proteomics data support the presence of the predicted salicyl alcohol degradation pathway....84
Both *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16 grow in co-culture in salicin by unidirectional cross-feeding mechanism....86
Metabolomics and proteomics data elucidate the mechanism of metabolic cross-feeding interaction between *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16....87
Nature of interaction between *Pseudomonas* sp. GM16 and *Rahnella* sp. OV744 in the salicin co-culture is proposed using growth assays and proteomic analysis....89
Chemotaxis assays suggest that salicin might be involved in mediating the interaction between *Populus*, *Rahnella* and *Pseudomonas*....91

Discussion....92

Chapter 5: Conclusions and future directions....98

Bibliography....103

Appendices....119

Appendix A: Supplementary data for chapter 2....120

Appendix B: Supplementary data for chapter 3....132

Appendix C: Supplementary data for chapter 4....136

Appendix D: Other works I have been part of during my PhD....144

Vita....147

List of tables

Table 1.1: Examples of bacterial mechanisms that have been elucidated to confer positive effect on plants. 4

Table 1.2: Bacterial strains known to degrade salicin. 12

Table 2.1: Locus tags of genes used for determination of gene clusters involved in salicin degradation. 22

Table 2.2 Proteins encoded by *salRCB* and *lacABC* operons were significantly expressed in salicin cultures. 29

Table 3.1: Comparison of abundance of proteins between salicin and glucose cultures of *Agrobacterium rhizogenes* OK036 demonstrated involvement of a cluster of eleven genes most likely required for salicin degradation. 51

Table 4.1: Probable proteins involved in salicin degradation for *Rahnella* sp. OV744 were detected in salicin cultures only. 79

Table 4.2: Proteins of probable pathway of salicyl alcohol degradation were significantly upregulated in *Pseudomonas* sp. GM16. 85

Table S2.1: Taxonomic classification of strains predicted to contain either Bgl, Sal or Lac system. 120

Table S2.2: Genes and their neighboring sequences that might have been mutated for *Citrobacter* to have acquired the ability to grow on salicin. 124

Table S2.3: Three strains that could likely possess both Sal and Lac systems other than the ones already identified in the study. 125

Table S3.1: Genes used as queries for blastp. 132

Table S3.2: Genes of the *Agrobacterium rhizogenes* OK036 cluster were some of the top hits when protein sequences of *Streptomyces* salicylate pathway genes were used as queries in blastp. 132

Table S4.1: Probable salicyl alcohol exporters in *Rahnella* sp. OV744. 136

List of figures

Figure 1.1: A two-step selection hypothesis could explain the host colonization process in rhizosphere and endosphere. 3

Figure 1.2: Structures of salicin and other phenolic glycosides (PGs)……..7

Figure 1.3: PG degradation leads to release of toxic chemicals. 10

Figure 1.4: Biosynthesis of PGs in *Populus*. 11

Figure 2.1: Number of strains categorized by predicted salicin degradation systems (Bgl, Sal or Lac). 23

Figure 2.2: Phylogenetic distribution of salicin degrading strains based on the experimental screening assay..... 25

Figure 2.3: *Caulobacter* sp. AP07 was able to utilize salicin and release salicyl alcohol as a by-product. 27

Figure 2.4: *Citrobacter* sp. YR018 grew on salicin, but only after prolonged incubation. 30

Figure 2.5: Growth of *Citrobacter* on salicin appeared to be due to mutational activation. 31

Figure 3.1: *Agrobacterium rhizogenes* OK036 utilized salicin completely. 49

Figure 3.2: Proposed pathway of salicin degradation in *Agrobacterium rhizogenes* OK036. 53

Figure 3.3: *Agrobacterium rhizogenes* OK036 demonstrated no growth on the intermediates of the proposed salicin pathway. 55

Figure 3.4: Metabolomics data support the presence of proposed pathway in *Agrobacterium rhizogenes* OK036. 56

Figure 3.5: Gene cluster identification in other *Populus*-associated strains. 58

Figure 4.1: *Rahnella* sp. OV744 utilizes salicin and secretes salicyl alcohol in the process..... 81

Figure 4.2: *Rahnella* sp. OV744 phosphoglucosidase gene complements the *Pantoea* sp. YR343 ΔGH1 mutant. 82

Figure 4.3: *Pseudomonas* sp. GM16 can utilize salicyl alcohol but not salicin. 84

Figure 4.4: *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16 can both grow on salicin co-culture by cross-feeding salicyl alcohol. 88

Figure 4.5: Proteomics data support the salicyl alcohol cross-feeding hypothesis between *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16 in the salicin co-culture. 90

Figure 4.6: Proposed model of mechanism of microbial recruitment of *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16 employed by *Populus* through salicin/PGs……97

Figure S2.1: Phylogenetic distribution of salicin degradation pathways of 45 strains predicted to contain either Bgl, Sal or Lac system. 126

Figure S2.2: Comparison of prediction and experimental results between the strains expected to comprise either Bgl, Sal or Lac systems. 126

Figure S2.3: *Caulobacter* possesses a complete *lacC* homolog. 127

Figure S3.1: *Agrobacterium* grows on salicin to an OD_{600} close to 1. 133

Figure S4.1: Possible GH1 enzyme encoding genes and their clusters in *Rahnella* sp. OV744. 137

Figure S4.2: Gel electrophoresis of pSRK.km_OVCompYR digested with SacI and KpnI. 138

Figure S4.3: Growth delay of *Pseudomonas* sp. GM16 in salicyl alcohol is not due to emergence of mutants..... 138

Figure S4.4: Growth curve analysis of *Pseudomonas* sp. GM16 on key intermediates of proposed pathway. 139

Figure S4.5: Stable state of the co-culture demonstrates that *Pseudomonas* sp. GM16 and *Rahnella* sp. OV744 grow in the ratio of ~ 2:1..... 139

Figure S4.6: Growth of *Rahnella* sp. OV744 is likely affected by the growth of *Pseudomonas* sp. GM16 in co-culture. 140

Figure S4.7: Salicyl alcohol (SalOH) does not significantly inhibit the growth of *Rahnella* sp. OV744. 140

Figure S4.8: *Pseudomonas* sp. GM16 secretion does not affect the growth of *Rahnella* sp. OV744. 141

Figure S4.9: Chemotaxis assays demonstrate *Pseudomonas* sp. GM16 is attracted towards salicyl alcohol. 141

List of attachments

Attachment S2.1: Taxonomic classification of 407 *Populus* isolates (AttS2.1_407GenomeData.xlsx)

Attachment S2.2: Blastp and cluster prediction results for all the genes encoding the Bgl system (AttS2.2_BglGenes_407Genomes.xlsx)

Attachment S2.3: Blastp and cluster prediction results for all the genes encoding the Sal system (AttS2.3_SalGenes_407Genomes.xlsx)

Attachment S2.4: Blastp and cluster prediction results for all the genes encoding the Lac system (AttS2.4_LacGenes_407Genomes.xlsx)

Attachment S2.5: Blastp results for *cbg1* gene (Cbg system) (AttS2.5_CbgGenes_407Genomes.xlsx)

Attachment S2.6: 45 genomes classified by their predicted salicin degradation clusters (Bgl, Sal or Lac) (AttS2.6_45PredictedSalicinDegraders_Bgl_Sal_Lac.xlsx)

Attachment S2.7: 166 genomes predicted to not contain any *bglB* homolog (AttS2.7_166strains_withNoPredictedBglB.xlsx)

Attachment S2.8: 67 genomes predicted to not contain any of the four systems in totality (AttS2.8_67strains_withNocompleteSalicinSystem.xlsx)

Attachment S2.9: 28 genomes predicted to not contain any GH1 or GH3 enzymes (AttS2.9_28strains_withNoGH1orGH3.xlsx)

Attachment S2.10: Experimental screening growth data of all 407 isolates in minimal media containing salicin as sole carbon source (AttS2.10_experimentalScreening_407Strains.xlsx)

Attachment S2.11: Details extracted from JGI database for the genes of Lac and Sal systems in *Caulobacter* sp. AP07 (AttS2.11_GeneDetails_AP07.xlsx)

Attachment S2.12: Proteome data comparing protein abundance profiles of glucose and salicin monocultures of *Caulobacter* sp. AP07 (AttS2.12_AP07ProteomeComparisons_GluVSal.xlsx)

Attachment S2.13: Details of GH1 and GH3 genes that are most likely involved in salicin cleavage in *Citrobacter* sp. YR018 (AttS2.13_GeneDetails_YR018.xlsx)

Attachment S3.1: Details extracted from JGI database for the probable genes encoding GH1 and GH3 enzymes in *Agrobacterium rhizogenes* OK036 (AttS3.1_GeneDetails_OK036_GH1_GH3.xlsx)

Attachment S3.2: Proteome data comparing protein abundance profiles of glucose and salicin monocultures of *Agrobacterium rhizogenes* OK036 (AttS3.2_OK036ProteomicsFull.xlsx)

Attachment S3.3: Details of genes of the gene cluster identified from the proteome data (AttS3.3_GeneDetails_cluster_OK036.xlsx)

Attachment S3.4: LC-MS data of the selected metabolites that were identified in the study (AttS3.4_LCMSDataForMetaboliteAnalysis.xlsx)

Attachment S3.5: Blastp hit result and protein families (PFAM and TIGRFAM) of the gene hits (AttS3.5_benzoateCoALigase_blastTo407Genomes.xlsx)

Attachment S4.1: Details of genes of proposed salicin degradation pathway in *Rahnella* sp. OV744 (AttS4.1_proposedGenes_OV744.xlsx)

Attachment S4.2: Proteome data comparing protein expression profiles of glucose and salicin monocultures of *Rahnella* sp. OV744 (AttS4.2_OV744Prot_GluVSalcn.xlsx)

Attachment S4.3: Details of genes of proposed salicyl alcohol utilization pathway in *Pseudomonas* sp. GM16 (AttS4.3_proposedgenes_GM16.xlsx)

Attachment S4.4: Proteomic comparison of expression of glucose and salicyl alcohol monocultures of *Pseudomonas* sp. GM16 (AttS4.4_GM16Prot_GluVSalOH.xlsx)

Attachment S4.5: qPCR dataset of co-cultures harvested, and total DNA extracted at different time points during growth (AttS4.5_qPCR_co-culture.xlsx)

Attachment S4.6: Stable State determination of co-culture using qPCR (AttS4.6_stableStateDetermination.xlsx)

Attachment S4.7: qPCR dataset of *Rahnella* salicin monoculture harvested, and total DNA extracted at different time points during growth (AttS4.7_OV744_monoculture_qPCRData.xlsx)

Attachment S4.8: Protein expression profiles compared between glucose monocultures of *Rahnella* and *Pseudomonas* salicin co-culture (AttS4.8_CocultureVGluMonocultures.xlsx)

Attachment S4.9: Comparing proteome of salicin co-culture with that of *Rahnella* salicin monoculture and of *Pseudomonas* salicyl alcohol monoculture (AttS4.9_CocultureVSalMonocultures.xlsx)

List of abbreviations

2D LC-MS/MS - Two-dimensional liquid chromatography/tandem mass spectrometry

6TMS - 6-trimethylsilyl

AA - Acetic acid

ABC - Ammonium bicarbonate

ACC - 1-aminocyclopropane-1-carboxylate

ACN - Acetonitrile

AMP - Adenosine monophosphate

BH - Benjamini-Hochberg

Blast - Basic local alignment search tool

CID – Collision induced dissociation

DNA - Deoxyribonucleic acid

DOE – Department of Energy

DTT - Dithiothreitol

EDTA - Ethylenediaminetetraacetic acid

ESI - Electrospray ionization

FASTA - Fast adaptive shrinkage threshold algorithm

FDR - False discovery rate

GC-MS - Gas chromatography - mass spectrometry

GH1 - Glycosyl hydrolase family 1

GH3 - Glycosyl hydrolase family 3

GMC - Glucose–methanol–choline

GSH - Glutathione

HCC - 1-hydroxy-6-oxo-2-cyclohexene-1- carboxylic acid

HMM – Hidden markov model

HPLC - High performance liquid chromatography

IAA - Indole acetic acid

IMG - Integrated microbial genomes

IPA - Isopropyl alcohol

JGI - Joint Genome Institute

KEGG - Kyoto Encyclopedia of Genes and Genomes

LC - Liquid chromatography

LTQ - Linear trap quadrupole

MOPS - 3-(N-morpholino)propanesulfonic acid

MudPIT - Multidimensional protein identification technology

MWCO - Molecular weight cut-off

NADPH - Reduced nicotinamide adenine dinucleotide phosphate

NCBI - National Center for Biotechnology Information

NEB - New England Biolabs

NIST - National Institute of Standards and Technology

NSAF - Normalized spectral abundance factors

OMF - Outer membrane factor

PAL - Phenylalanine ammonia lyase

PCR – Polymerase chain reaction

PFAM – Protein families

PHYLIP - Phylogeny inference package

PTS – Phosphotransferase system

qPCR – Quantitative polymerase chain reaction

RBH – Reciprocal blast hit

RNA – Ribonucleic acid

SCX - Strong cation exchange

SDC - Sodium deoxycholate

SDS - Sodium dodecyl sulfate

TIC - Total ion current

TIGRFAM - The Institute for Genomic Research Families

UHPLC – Ultra-high performance liquid chromatography

UNIFAM - UniProt-based families

Chapter 1

Introduction

The focus of this dissertation is on *Populus*-microbiome metabolic interactions mediated through phenolic glycosides (PGs), such as salicin. For this, we aimed to understand what diversity of mechanisms are employed by phylogenetically diverse bacteria to utilize salicin and to propose that such compounds could be one of the driving factors in shaping the *Populus* microbiome.

Plant-microbe interactions

Microbes ranging from viruses to fungi occupy a multitude of habitats on plants both above-ground and below-ground with distinct communities in each [8-15]. Because of the association with plants, microbes can interact with plants to provide numerous benefits such as immune response and substrate acquisition to plants [6, 16-18]. One of the more well-studied regions of interactions between a plant and its microbiome is the rhizosphere which is defined as the region around the root under the influence of the plant [19]. These interactions are partially mediated by the process of rhizodeposition through which roots release organic acids, sugars, amino acids and vitamins to form the rhizosphere [6]. Rhizodeposition provides the nutrient sources to stimulate the growth of microorganisms and offers a distinct niche near the root surface for microbes to metabolically interact with plants and with other microbes.

Apart from the rhizosphere, below-ground microbial community also inhabit regions within the root tissues referred to as the endosphere. Endosphere communities are driven by different forces compared to the ones in the rhizosphere. The working model of microbial assemblages in rhizosphere and endosphere is a two-step selection process (figure 1.1). According to the proposed model, microbial communities in the rhizosphere are influenced by the metabolic signals whereas in the endosphere, host-genotypic factors such as innate immunity help fine-tune

the microbial association [6]. This leads to differential colonization of rhizosphere and endosphere which has been observed in plants such as *Populus* and *Arabidopsis* [8, 20].

Benefits of plant-microbe interactions and their studies

Due to the immobile nature of plants, the presence of microbiome has been deemed crucial to the sustenance and fitness of plants as they provide an extension to plant abilities [6, 21]. A list of benefits provided by the plant-microbe interactions is presented in table 1.1.

Because of the increasingly limited availability of productive land for agriculture, investigators have focused research efforts on developing technologies for efficient growth and health of plants. Traditionally, studies were directed towards plant breeding or using chemical

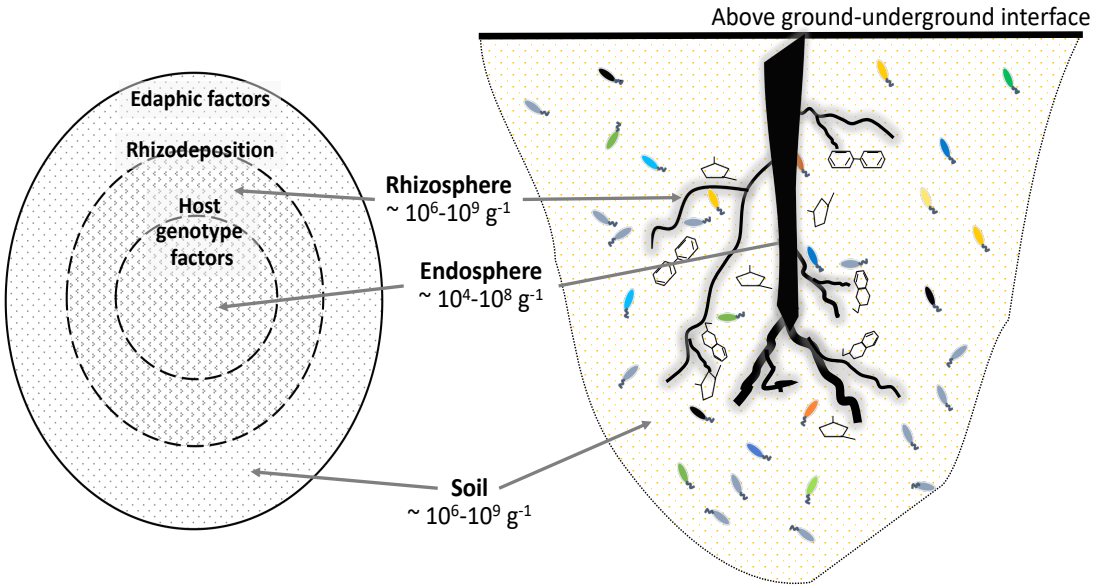


Figure 1.1: A two-step selection hypothesis could explain the host colonization process in rhizosphere and endosphere.
The microbes in soil are influenced by the edaphic (soil-related) factors. First instance of belowground plant-microbe interactions occurs at the region between the roots and the soil which is called the rhizosphere (shaded region in the picture on the right). These interactions are partially facilitated by rhizodeposition. A fraction of the bacterial population in soil and rhizosphere is fine-tuned by host genotypic factors as they migrate into the region within the root tissues called endosphere. The bacterial titers (labeled in the figure) in each region – soil, rhizosphere and endosphere are difficult to calculate but Bulgarelli and colleagues [6] have estimated the numbers based on several studies.

Table 1.1: Examples of bacterial mechanisms that have been elucidated to confer positive effect on plants.

Activity	Strains	Study
Nitrogen fixation (N_2 => NH_3)	▪ *Azoarcus* sp. BH72 – *Kallar* 342 ▪ *Klebsiella pneumoniae* 342 - *Triticum aestivum* L.	[16, 22]
Nitrogen metabolism (NO_3^- => N_2O/NO/NH_3)	▪ *Azospirillum brasilense* Sp245 - *Solanum lycopersicum* Mill.	[23]
Phosphorous acquisition	▪ *Glomus* strains – *Medicago truncatula*	[24]
Iron acquisition	▪ *Pantoea eucalypti* M91 – *Lotus japonicus* Gifu	[25]
Auxin homeostasis	▪ *Azospirillum brasilense* – *Triticum aestivum* ▪ *Pseudomonas putida* – Radish	[26, 27]
1-aminocyclopropane-1-carboxylate (ACC) deaminase activity	▪ *Pseudomonas putida* – Canola seedlings	[28]
Volatile compounds	▪ *Bacillus subtilis* – *Arabidopsis thaliana*	[29]
Antimicrobial compound biosynthesis	▪ *Pseudomonas fluorescens* – *Azospirillum brasilense* – Spring wheat ▪ *Burkholderia pyrrocinia* – *Chromobacterium violaceum* –*Arabidopsis thaliana*	[30, 31]
Induced systemic resistance	▪ *Pseudomonas syringae* – *Lycopersicon esculentum* ▪ *Mesorhizobium loti* – *Lotus japonicus* ▪ *Serratia liquefaciens* – *Pseudomonas putida* – *Lycopersicon esculentum*	[32-34]

fertilizers and pesticides, but such approaches have created unintended ecological and environmental issues. For instance, traditional plant breeding approaches have homogenized genetic variation [35]. Likewise, harmful pesticides and chemical fertilizers are now well-known to contaminate soil and leach into groundwater. Moreover, such approaches have been well-documented to be costly in the long run [19, 36]. Using microbes as an alternative has been touted as a more cost-efficient and an environmentally conscious approach to the sustainable production of crops. For instance, researchers have reported that the use of microbes on fields could cost 1% as much as the cost of using nitrogen fertilizers [36]. Thus, understanding plant-microbe interactions and microbiome functional repertoire could be beneficial for creating eco-friendly and cost-friendly solutions in agriculture.

Populus **as a model for plant-microbe interaction**

Plant-microbe interactions have been studied in numerous plant species such as *Arabidopsis*, *Solanum*, *Oryza*, and *Populus* [10, 15, 20, 37]. Studies of woody plants such as *Populus* provide unique insights because they have longer life cycles and have to adapt to the changing seasons compared to smaller plants with short life cycles such as *Arabidopsis* [38]. *Populus* has been extensively studied for optimization of lignocellulose-based biofuel production because of its relatively small genome size, fast growth, and availability of molecular tools for manipulating this organism [39]. Because of these properties of *Populus* and an overwhelming understanding that optimal microbiome is the best solution to sustainable crop production, focus has been dedicated to investigating the plant-microbe interactions using *Populus* as a model species. As such, a more detailed understanding of molecular mechanisms of microbial selection and colonization of *Populus* is required. In that regards, research to increase our understanding of

host-microbe metabolic interactions is warranted as such communications are believed to be crucial in host-microbe associations [40, 41]. In case of *Populus*, one of its defining features is the amount and diversity of phenolic glycosides (PGs) produced in its tissues (figure 1.2) [42, 43].

Phenolic glycosides (PGs)

PGs refer to metabolites that consist of a phenolic compound conjugated to a sugar (i.e. salicin). In salicin, salicyl alcohol and glucosyl moieties are conjugated by a β(1→4) glycoside bond. Complex PGs such as tremulacin and salicortin are formed by the esterification of one or more hydroxyl groups of the salicyl or glucosyl moieties with various functional groups [7]. *Populus* and other members of *Salicaceae* family produce abundant amounts of PGs with some reports suggesting concentrations of up to 30% of foliar biomass in certain *Populus* clones [44].

The amount and diversity of PGs produced by plants is dependent on various factors such as genotype, developmental stage and environmental factors. The amount of total PGs has been observed to be species-specific [42]. Moreover, studies have demonstrated tremendous clonal variation in strains of *Populus* and *Salix* [45, 46]. Therefore, genotype has been deemed as a major factor in PG production [7]. The developmental stage of the plant can also cause significant differences in the concentration of PGs. Studies have demonstrated that PGs in leaves of juvenile plants of trembling aspen are significantly higher than those in leaves of their adult counterparts [44]. Similarly, Erwin and colleagues [47] determined that concentrations of foliar PGs in *Populus tremuloides* seedlings were significantly higher than those in mature stands. Environmental factors such as nutrient availability too have impact on the PG concentration [48, 49]. Similarly, seasonal variations in PGs have also been reported [50].

PGs are constitutively produced in the members of *Salicaceae*, but some studies have

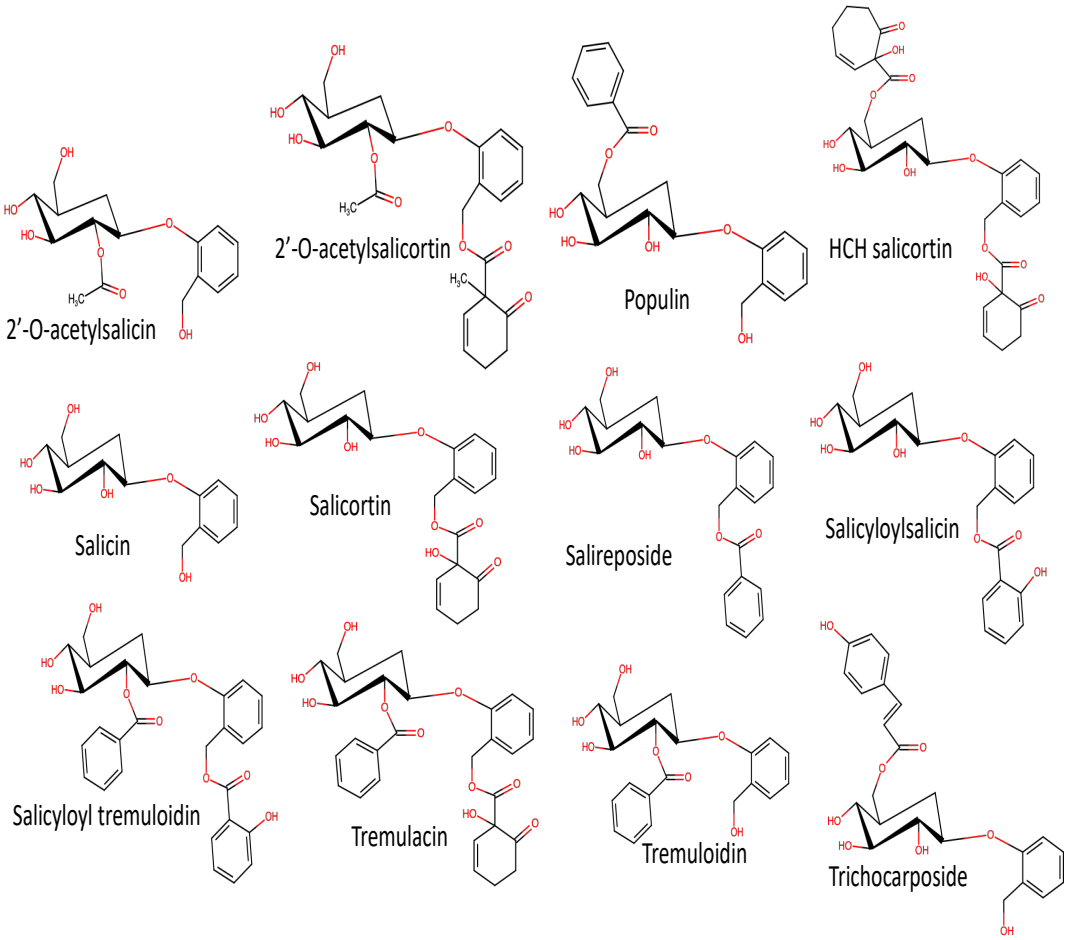


Figure 1.2: **Structures of salicin and other phenolic glycosides (PGs).**
PGs consist of a core structure of salicin which contains glucosyl and salicyl alcohol motifs that are conjugated by β-glycoside bond. Esterification at either glucosyl or salicyl alcohol functional groups by organic or phenolic acids can transform the salicin core structure to more complex PGs such as salicortin, tremulacin, HCH-salicortin and populin. More than 20 such PGs have been described in literature [7].

reported their production is induced upon foliar defoliation. For instance, salicortin and tremulacin levels increased significantly in leaves within 24 hours upon simulated herbivory of *P. tremuloides* [51]. In a study by Fields and Orians [52], it was demonstrated that levels of 2′-cinnamoylsalicortin were higher in the leaves above the damaged leaves (i.e. systemic location). In another study, gypsy moth herbivory was demonstrated to make *P. tremuloides* increase both its local (damaged site) and systemic production of salicortin and tremulacin [53].

Several factors could complicate the measurement of these compounds such as genotype, seasons, herbivore-type and even transport of PGs between different tissues [7]. For instance, salicin has been detected in phloem exudates of *P. deltoides* suggesting that these compounds (or at least salicin) can be translocated between tissues [54]. Therefore, measurement of PG levels is a challenging task.

Importance of PGs

The importance of PGs has been realized in plants as plant-defense and anti-herbivore compounds [55-58]. PGs are demonstrated to alter feeding habit of herbivores such as mammals and insects [46, 51, 58, 59]. This most likely is because their phenolic intermediates can be toxic to herbivores as well as to the plants themselves. Therefore, plants have evolved several mechanisms to limit self-toxicity [60]. These include glycosylation of phenolic compounds to produce PGs and storage of PGs (in vacuoles or cytosol) away from hydrolase enzymes (in chloroplast) until needed. Any form of mechanical damage to plant tissues, however, can cause induction and release of these compounds along with hydrolyzing enzymes. PGs such as salicortin and tremulacin can be hydrolyzed either through the actions of esterases or other enzymes to salicin and 6-hydroxy-2-cyclohexanone (6HCH). 6HCH can be further degraded to catechol. Both

catechol and 6HCH are toxic compounds [7, 51, 61, 62]. In addition to more complex PGs, salicin itself can be degraded by glucosidases to harmful chemical phenolics such salicyl alcohol, salicylaldehyde and salicylic acid (figure 1.3) [60, 63].

In addition to their importance to plants, PGs have found uses in the area of biotechnology and medicine. For instance, salicin and salicylic acid (degradation product of salicin) have been traditionally used as ingredients of analgesic and anti-fever drugs [64]. More recently, researchers have demonstrated that these compounds could be used as antiproliferative and anticancer agents [65-67].

Biosynthesis of PGs

Even though numerous studies have attempted to elucidate the PG synthesis pathways, a complete understanding of the biosynthetic production of PGs has not been achieved to date. Recently, Babst et al. [5] utilized stable-isotope labeling experiments to corroborate earlier studies that phenylalanine is the precursor to PGs. In the first step of the pathway, phenylalanine is converted to cinnamic acid by the action of phenylalanine ammonia lyase enzyme (PAL) before being transformed to salicylaldehyde through a coumaric acid intermediate. Salicylaldehyde is then oxidized to salicyl alcohol which is subsequently glycosylated to form salicin. Salicin can be conjugated to phenolic compounds such as 1-hydroxy-6-oxo-2-cyclohexene-1-carboxylic acid (HCC) to produce higher phenolic glycosides such as salicortin. Alternatively, cinnamic acid can also be converted to benzoyl-CoA which can be transformed to salicyl alcohol and its analogs. Benzoyl-CoA then can be converted to benzyl-benzoate which has been proposed to form salicortin through a salicyl-6-hydroxy-2-cyclohexanone (salicyl-6HCH) intermediate [5]. A concise version of the pathways proposed in the literature has been outlined in figure 1.4.

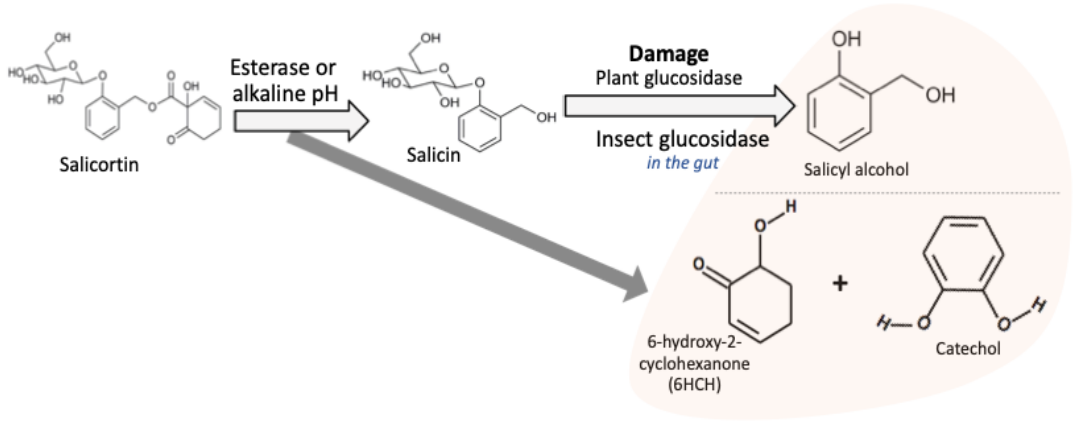


Figure 1.3: **PG degradation leads to release of toxic chemicals.**
PGs are toxic in nature and are stored in different compartments than the glucosidases in plants to limit self-toxicity. PGs such as salicortin can be degraded to salicin by the action of esterase or alkaline pH. 6HCH and catechol are released in the process of salicortin hydrolysis. Furthermore, upon damage, plant releases PGs along with glucosidases which act on salicin to release salicyl alcohol. Insects also possess glucosidases that can degrade salicin. The products salicyl alcohol, 6HCH and catechol are all toxic in nature (highlighted in orange).

Degradation mechanisms of PGs in *Populus*-associated microorganisms and their possible role in shaping the *Populus* microbiome

Recent studies have demonstrated that phenolic compounds are important in *Populus*-microbiome interactions [68, 69]. Since *Populus* produces significant amounts of PGs and these compounds can be potential mediators of *Populus*-microbiome interactions, it is important to understand what proportion of the microbiome population can degrade these compounds and what kind of mechanisms are employed. Therefore, in this work, we aimed to study the diverse mechanisms by which *Populus*-inhabiting bacteria degrade PGs and to demonstrate the importance of PGs to *Populus*-microbiome interactions. For our research purpose, we worked with the commercially available salicin. Salicin degradation involving enzymes of glycosyl hydrolase family 1 or 3 (GH1 or GH3) has been studied in bacteria of different families but not all mechanisms are well-described (Table 1.2).

Individual bacterial mechanisms of salicin degradation have been described in the

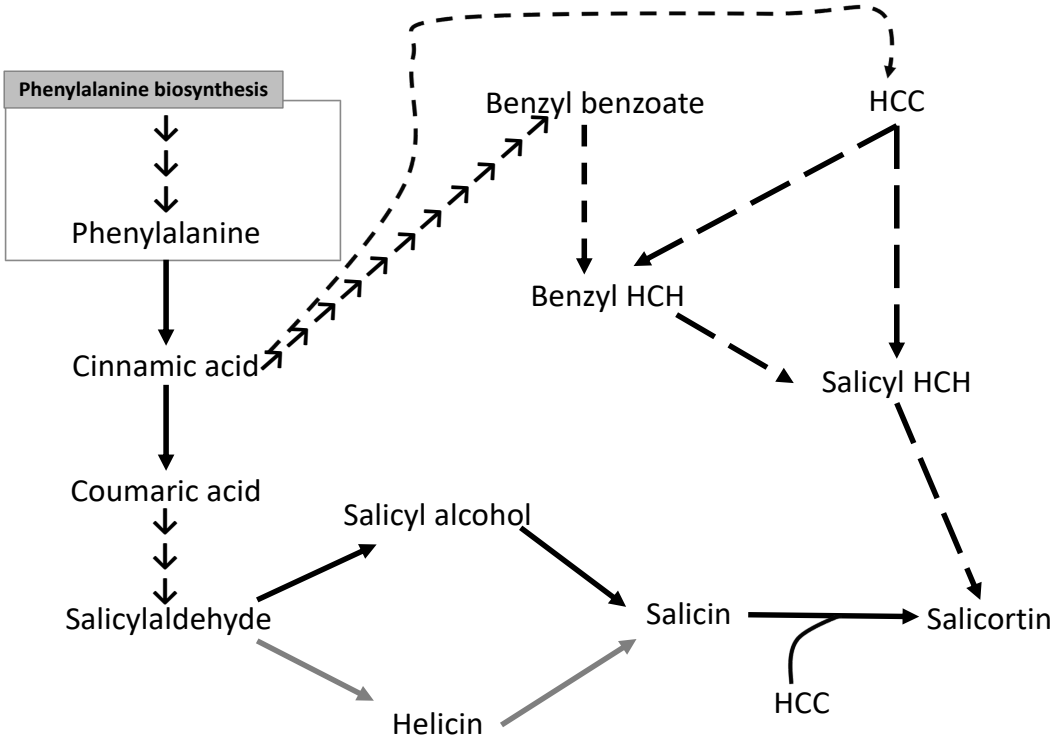


Figure 1.4: Biosynthesis of PGs in *Populus*.
The precursor to PG biosynthesis is phenylalanine produced in the phenylpropanoid pathway. Phenylalanine is converted to cinnamic acid via the action of phenylalanine ammonia lyase (PAL). Subsequently, cinnamic acid and coumaric acid can be transformed to salicylaldehyde which is reduced to salicyl alcohol (or helicin to certain extent). Salicyl alcohol is glucosylated to form salicin. The HCC (6-hydroxy-2-cyclohexanone), which is derived from cinnamic acid, can be conjugated to salicin to form salicortin. Metabolomic data in the study by Babst et al. [5] suggested another pathway for the biosynthesis of salicortin in which cinnamic acid is first converted to benzyl benzoate through a series of steps.

Table 1.2: Bacterial strains known to degrade salicin.
A few bacterial strains have been studied to understand the mechanisms of degradation of complex sugars including salicin. The salicin degrading hydrolases have been identified as either glycosyl hydrolase family 1 (GH1) or 3 (GH3) enzymes except in the case of *C. crescentus* (not detected – ND).

Organism	GH1 or GH3	Study
Escherichia coli	GH1	[70]
Dickeya chrysanthemi	GH1	[71]
Klebsiella aerogenes	GH1	[2]
Azospirillum irakense	GH3	[72]
Caulobacter crescentus	ND	[73]
Dickeya dadantii	GH3	[74]
Agrobacterium tumefaciens	GH3	[75]

literature, but a systematic study of such mechanisms has not been performed, especially in *Populus*-associated strains. We have access to a diverse collection of bacterial strains isolated from roots of wild and field grown *Populus* trees. The genome sequences of these isolates are also available. Here in chapter 2, we investigate the prevalence and diversity of microbial mechanisms of salicin degradation in 407 *Populus*-associated microbial isolates using bioinformatic and experimental methods. The analysis described provides insights into the phylogenetic diversity of salicin degrading bacterial strains and distribution of possible mechanisms in these strains. Furthermore, we characterize two strains *Caulobacter* sp. AP07 and *Citrobacter* sp. YR018 in terms of their salicin degradation mechanisms. In chapter 3, we elucidate a novel salicin degradation pathway in *Agrobacterium rhizogenes* OK036 in which salicin is degraded via salicylate and gentisate intermediates. This organism is able to utilize both glucosyl and salicyl alcohol groups of salicin. Finally, in chapter 4, we investigate the mechanism of co-metabolism of salicin by *Rahnella-Pseudomonas* strains when cultured together on salicin minimal media and go onto develop a cross-feeding model. We also discuss the potential importance of such interactions in the establishment and stability of microbial community in chapter 4. The conclusions and future directions of the projects described in this dissertation are discussed in chapter 5.

Chapter 2

Phylogenetic and functional diversity in salicin degrading *Populus* bacterial isolates

Introduction

Populus is a model perennial woody plant that has been extensively studied for bioproduction of cellulosic biofuel. Woody plants are different from short-lived herbaceous plants because they need to thrive under varying biotic and abiotic conditions remaining in a fixed location for centuries [38]. Therefore, these plants have a long-term and well-established relationship with their microbiome. *Populus* hosts thousands of diverse microorganisms ranging from viruses to fungi around (rhizosphere) and within their roots (endosphere) [76].

The metabolic interaction between the host and microbes have been demonstrated to be an important process in shaping the microbial community [40, 41]. The members of *Salicaceae* family such as *Populus* are known to produce significant amounts of salicylate-containing phenolic glycosides (PGs) such as salicortin and tremulacin. These compounds are important to plants because they can act as anti-herbivore agents [58, 77]. One of the simplest forms of PGs that is commercially available is salicin which consists of salicyl alcohol moiety conjugated to glucosyl group. The bond between these two functional groups can be cleaved either by the action of hydrolytic enzymes or through dilute acid [78].

The role of PGs in shaping microbial communities has only been tangentially examined in a handful of studies. In one such study Sonowal and colleagues [79] proposed that soil bacteria have an evolutionary advantage to retain salicin hydrolytic pathways over their gut counterparts because they are more likely to encounter plant-released compounds such as salicin and arbutin in soil. The researchers demonstrated that soil bacteria consume the glucosyl moiety of salicin and release salicyl alcohol which is toxic to bacterivores such as nematodes and amoeba [79]. In another investigation, it has been demonstrated that gypsy moth inoculated with aspen foliar microbial community grow 64% larger than non-inoculated counterpart when both were fed on

diet with high phenolic glycoside content, leading to the conclusion that the aspen leaf microbiome acquired by the moth helps to detoxify these PGs [56]. In a follow-up study, it was revealed that the gut microbiome of gypsy moth is modified by phenolic glycosides with an increase in *Acinetobacter* (PG degrader) and a decrease in *Ralstonia* populations [56, 80]. In a recent study involving colonization of *P. deltoides* by *Laccaria*, it was demonstrated that the levels of salicin increased by up to four folds after twelve weeks of colonization [68]. Therefore, the role of PGs in plant-microbe interactions and shaping of microbiome has been marginally understood, but a systematic study is required to understand the proportions of bacterial community affected by these interactions and the diverse mechanisms involved in PG utilization.

Only a small number of studies have targeted to understand mechanisms of microbial degradation of salicin. One of these mechanisms was first described in a study by Schaefler and Malamy [81]. In this work, strains of *Enterobacteriaceae* were characterized in terms of phosphoglucosidases and other proteins involved in salicin degradation. The gene encoding the phosphoglucosidase was later described as part of the *bgl* operon (Bgl system) which has been identified in *Erwinia chrysanthemi, Klebsiella aerogenes* and *Escherichia coli*. Other than the phosphoglucosidase (BglB of glycosyl hydrolase family 1 – GH1) gene, the *bgl* operon consists of genes encoding phosphotransferase system (PTS) transporter (BglF) and antiterminator (BglG) proteins. In the Bgl system, the PTS transporter is required to transport and phosphorylate salicin into the cytosol as salicin 6-phosphate which is subsequently acted upon by a glucosidase to yield salicyl alcohol and glucose 6-phosphate [1, 2, 71]. The antiterminator protein has been demonstrated to be required for catabolite repression of the transcription of *bgl* operon in the presence of other sugars such as glucose [82].

Similarly, another operon *lacABC* (Lac system) consisting of genes encoding glucose–

methanol–choline (GMC) oxidoreductase family protein (LacA), cytochrome c family protein (LacB) and gluconate 2-dehydrogenase subunit 3 containing protein (LacC) has been demonstrated to be involved in salicin or disaccharide binding and utilization in *Caulobacter crescentus* and *Agrobacterium tumefaciens* [73, 83, 84]. In addition to these genes, a hypothetical gene is also present within this operon. Deletion of any of the *lac* genes has been demonstrated to either reduce or eliminate the growth of *C. crescentus* strains on salicin [73]. It has been hypothesized that the Lac complex could be involved in oxidizing sugars to keto-sugars which are then acted upon by hydrolase enzymes, but the actual mechanism has not been fully elucidated [85].

A third salicin degradation system has been described in *Azospirillum irakense* KBC1 which consists of *salRCAB* (Sal system) operon comprised of genes encoding LacI/GalR family regulator protein (SalR), outer membrane receptor family protein (SalC) and two functionally redundant glycosyl hydrolase-3 (GH3) proteins (SalA and SalB) [72, 86]. SalC along with TonB has been hypothesized to import salicin into the periplasm where SalA and SalB hydrolyze the substrate to glucose and salicyl alcohol [72, 86, 87]. SalR is a negative regulator which suppresses the *sal* operon [87].

Finally, a coniferin hydrolase – Cbg1 (Cbg system), a member of the GH3 family has been demonstrated to exhibit hydrolase activity with salicin [75]. The enzyme was first isolated from *Agrobacterium tumefaciens* [88]. The protein contains a PA14 domain which has been proposed to be involved in substrate recognition [89]. The hydrolase activity of Cbg1 has been revealed to be linked to an alcohol-dependent transglycosylation process [75, 90].

We began the current project by making genomic predictions of likely mechanisms of salicin degradation in 407 *Populus* bacterial isolates followed by experimental screening on salicin minimal media. The predictions and experimental results matched in ~71% of the strains. To the

best of our knowledge, this is one of the most comprehensive studies on the identification and screening of salicin degraders. Following this, we chose two strains – *Caulobacter* sp. AP07 and *Citrobacter* sp. YR018 for further analysis because they appeared to contain unique mechanisms of salicin degradation that have not been studied so far in *Populus* isolates. Both of these strains fall in the phylum *Proteobacteria* in which *Citrobacter* is in class *Gammaproteobacteria* whereas *Caulobacter* is an *Alphaproteobacteria*. Using both qualitative and quantitative approaches, we propose a unique model of salicin utilization in *Caulobacter* in which it utilizes both Lac and Sal systems for growth on salicin. In the case of *Citrobacter* sp. YR018, it demonstrated the ability to utilize salicin but only after mutational activation of a yet unknown gene/s. Therefore, the result from *Citrobacter* study indicates that *Populus* isolates could evolve to acquire salicin degradation ability.

Methods and materials

Bacterial strains and culture medium

The bacterial strains characterized in this study were previously isolated from *Populus* root tissues. Their genomes have already been sequenced and assembled as described elsewhere [91-93]. The bacterial strains were maintained on R2A (Franklin Lakes, BD Difco) plates.

For growth experiments, individual colonies from R2A plates were grown overnight in R2A liquid medium at 30°C in a shaking incubator at 200 rpm. The overnight cultures were subcultured in 3-(N-morpholino)propanesulfonic acid (MOPS) minimal medium [94] containing the designated carbon substrate (either glucose or salicin) at 3 mM concentration. For growth measurements, optical density at wavelength 600 nm (OD_{600}) was measured using a GENESYS

30 spectrophotometer (Thermo Fisher Scientific) or Synergy H1 microplate reader (BioTek) for 96-well plate format.

HPLC metabolite analysis of culture supernatant

Bacterial strains were grown in triplicates in 10 ml MOPS minimal medium containing 3 mM salicin. Samples were collected during late log phase of growth. Cells were pelleted by centrifugation at speed 10621 x g for 5 minutes in Eppendorf centrifuge 5417R and culture supernatants were collected and stored at -20°C. Frozen samples were thawed and mixed briefly by vortexing. 200 µl of sample was mixed with 500 µl purified water in scintillation vials and processed for metabolite analysis. HPLC metabolite analysis was performed as described elsewhere [95]. Briefly, Agilent 1260 Infinity HPLC was used with C18 column (Thermo Fisher Scientific Hypersil Gold 200/4.6 mm, 5 micron) for chromatography. Mobile phase for this experiment was prepared using 0.1% (v/v) orthophosphoric acid (pH 2.5) and 100% (v/v) acetonitrile. The flow rate was set at 1ml/min and chromatogram was acquired at wavelength 267 nm. Standard solutions of salicin and salicyl alcohol were used for determination of desired peaks.

Sample preparation for proteome measurements

Caulobacter sp. AP07 strain was grown in triplicates on MOPS minimal media containing either glucose (3 mM) or salicin (3 mM). Cell pellets were collected at OD_{600} close to 0.7 by centrifugation at 11952 x g (15 mins, 4°C) and removal of supernatant by decantation. Pellets were flash frozen in liquid N_2 and stored at -80°C.

For proteome measurements, pellets were first thawed, washed and resuspended in lysis buffer (500 µL of 4% sodium deoxycholate (SDC) in 100 mM ammonium bicarbonate (ABC, pH

8.0) and 10mM dithiothreitol (DTT)) for reducing protein. Lysate was transferred to the eppendorf tube containing 1 scoop of zirconium oxide beads for bead beating. The lysate was spun down for 10 minutes at 21000 x g and transferred to a new tube. The crude protein concentration was estimated using nanodrop (Thermo Fisher Scientific). Proteins were alkylated with 30 mM indole acetic acid (IAA) and incubated in the dark for 30 minutes. A total of 200 µg of protein was loaded on top of 10KDa filter (Vivaspin 500) and washed with 500 µL of 100 mM ABC by centrifuging at 10000 x g (~2% SDC concentration). Proteins were digested to peptides with two sequential aliquots of sequencing-grade trypsin (Promega Corp.) at an enzyme-protein ratio of 1:75 (w/w), initially overnight then again for 4 hours at room temperature. The tryptic peptide solution was then spin-filtered through the molecular weight cut-off (MWCO) membrane, adjusted to 1% formic acid to precipitate residual SDC. SDC precipitate was removed from the peptide solution with water-saturated ethyl acetate extraction. Peptide samples were then concentrated via SpeedVac (Thermo Fisher Scientific) and quantified by nanodrop (Thermo Fisher Scientific) using Scopes method [96]. Peptides were adjusted to a final concentration of 0.5 µg/µl and loaded in the autosampler vial for mass spectrometry analysis.

Peptide samples were analyzed by automated 1D LC-MS/MS analysis using a Vanquish UHPLC plumbed directly in-line with a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) outfitted with a back column (10 cm, 5 µm Kinetex C18 RP resin (Phenomenex)) coupled to an in-house pulled nanospray emitter packed with 30 cm, 5 µm Kinetex C18 RP resin (Phenomenex). For each sample, 3 µg of peptides was loaded, desalted, separated and analyzed using 150 minutes organic gradient as previously detailed [97]. Eluting peptides were measured and sequenced by data dependent acquisition on the Q Exactive MS.

MS data analysis and evaluation

All MS/MS spectra collected were processed in Proteome Discoverer (v.2.2, PD) with MS Amanda (v.2.2) [98] and Percolator [99]. Firstly, the spectral recalibration was performed to improve the feature mapping. Spectral data were searched against *Caulobacter* sp. AP07 database to which common laboratory contaminants were appended. The following parameters were set up in MS Amanda to derive fully tryptic peptides: MS1 tolerance = 5 ppm; MS2 tolerance = 0.02 Da; missed cleavages = 2; Carbamidomethyl (C, + 57.021 Da) as static modification; and oxidation (M, + 15.995 Da) and carbamylation (N-terminus, + 43.006 Da) as dynamic modifications. The Percolator FDR threshold was set to 1% at the peptide-to-spectrum matches (PSMs) and peptide level. Both unique and/or razor peptides were used for quantification of a protein. Strict Parsimony Principle parameter was used for protein grouping. The normalization feature of PD was used to normalize the proteins. A student's t-test was performed between glucose versus salicin group to extract the significantly changing proteins (p-value < 0.05).

Bioinformatic analysis of salicin degradation genes and gene clusters

Amino acid sequences of genes previously described in literature to be involved in the metabolism of salicin: *bglGFB*, *lacABC*, *salRCAB* and *cbg1* (see table 2.1) were downloaded from IMG (Integrated Microbial Genomes) [100] database and utilized to query 407 genomes (see attachment S2.1 for details of all 407 genomes) of bacterial isolates from *Populus* roots using standard blastp [101] parameters in Biopython (v. 1.71) [102]. Criteria for calling potential homologs, e-value ≤ 1E-05 was used. Additional criteria of length of alignment ≥ 30% of length of query sequence and percent identity ≥ 25% were used as more stringent cutoffs in some cases. Finally, for cluster determination, a reference gene (*bglB*/*lacA*/*salA*) was selected. Then, up to five

neighboring genes left and right of that particular gene were checked for matches to the individual blastp hits. If yes, then they were assigned as a member of a potential gene cluster. If a complete cluster was identified, then the probable pathway for salicin degradation was predicted to be present in that strain. *cbg1* was not used for prediction of presence or absence of pathway because it appeared to significantly inflate the number of salicin degrading isolates, but it was assigned to strains for record-keeping purpose.

Data analysis and visualization

A distance tree was generated in JGI/IMG (Joint Genome Institute/Integrated Microbial Genome) database using all 407 strains with blast identity threshold set at 30%. The tree was generated using 16S gene alignment which is based on the SILVA database [103] and on programs such as *dnadist* from the Phylogeny Inference Package (PHYLIP) [104]. It was color modified using ggTree (v. 1.14.6) package in R to visualize the distribution of salicin degrading strains [105]. Another tree of only 45 strains was downloaded and color modified for their predicted salicin degradation systems using the method described before. Similarly, d3heatmap (v. 0.6.1.2) package in R was utilized to create heatmaps to compare the experimental and prediction data [106].

Results

45 strains are predicted to utilize salicin through either Bgl, Sal or Lac systems

The 11 protein sequences, representative of four different salicin utilization systems – Bgl, Sal, Lac and Cbg were utilized to search 407 available *Populus*-associated microbial genomes using a custom cluster determination pipeline (see table 2.1 for genes used in this study). First, the

Table 2.1: Locus tags of genes used for determination of gene clusters involved in salicin degradation. Gene sequences (amino acid) from different organisms were used to search for four different salicin degradation systems in 407 *Populus*-associated isolates. Organisms used were: Bgl – *Escherichia coli* K-12 MG1655, Lac – *Caulobacter crescentus* CB15, Sal – *Azospirillum irakense* DSM 11586 and Cbg1 – *Rhizobium radiobacter*.

Genes	Homolog
Ga0063390_02056	BglB (Glycosyl Hydrolase 1)
Ga0063390_02057	BglF (Phosphotransferase system)
Ga0063390_02058	BglG (Antiterminator)
CC1634	Lac A (oxidoreductase, GMC family protein)
CC1632	Lac B (Cytochrome c family protein)
CC1635	Lac C (Hypothetical protein)
P27034	Cbg1 (Beta-glucosidase)
G473DRAFT_2580	SalA (Beta-glucosidase)
G473DRAFT_2579	SalB (Beta-glucosidase)
G473DRAFT_2581	SalC (Outer membrane receptor protein)
G473DRAFT_2582	SalR (Transcriptional regulator)

amino acid sequence of individual genes in the unique operons were searched through blastp followed by determination of the neighboring genes (see methods and materials section). See attachments S2.2-S2.5 for all the blastp hits of individual genes and for data on respective clusters.

Out of the 407 total genomes screened, only 28 did not possess any detectable BglB (GH1), SalA (GH3), SalB (GH3) or Cbg1 (GH3) orthologs. These included members of the following genera: *Methylobacterium*, *Phylobacterium*, *Polaromonas, Acidovorax and Mycobacterium* (see figure 2.1 and attachment S2.9). Cbg1 orthologs were detected in 340 genomes. A distinct phylogenetic pattern for the distribution of Cbg system was not observed most likely because it contains only one gene. Forty-five genomes were predicted to encode orthologs of Bgl, Sal or Lac systems (see attachment S2.6 and table S2.1). A 45-member phylogenetic tree color coded by aforementioned three salicin systems is presented in figure S2.1.

As observed in the figure 2.1, Bgl system was predicted to be present in 22 strains. It is

prevalent in *Bacillus* (in 5/15 total *Bacillus* genomes), *Paenibacillus* (in 6/12 total *Paenibacillus* genomes), *Pantoea* (5/5 total *Pantoea* genomes), *Rahnella* (2/2 total *Rahnella* genomes) and *Microbacterium (*in 2/3 total *Microbacterium* genomes). The *bgl*B (phosphoglucosidase) homologs were predicted to be absent in the genera *Pseudomonas* (59/59 total *Pseudomonas* genomes), *Variovorax* (21/21 total *Variovorax* genomes), *Chryseobacterium* (16/17 total strains) and others (see attachment S2.7).

The Sal cluster was detected in 20 strains and is predominant in the genus *Novosphingobium* (13/17 total *Novosphingobium* genomes). The Lac cluster was predicted to be present in 7 strains primarily belonging to the genus *Caulobacter* (4/5 total *Caulobacter* strains). Surprisingly, four out of the seven strains predicted to encode the Lac cluster also contain a Sal cluster (see table S2.1). Three of those four belong to the genus *Caulobacter*.

Further investigation suggested that 67 genomes contained only partial Bgl, Sal, Lac systems. These include genera *Acidovorax*, *Arthrobacter*, *Burkholderia*, *Ensifer*,

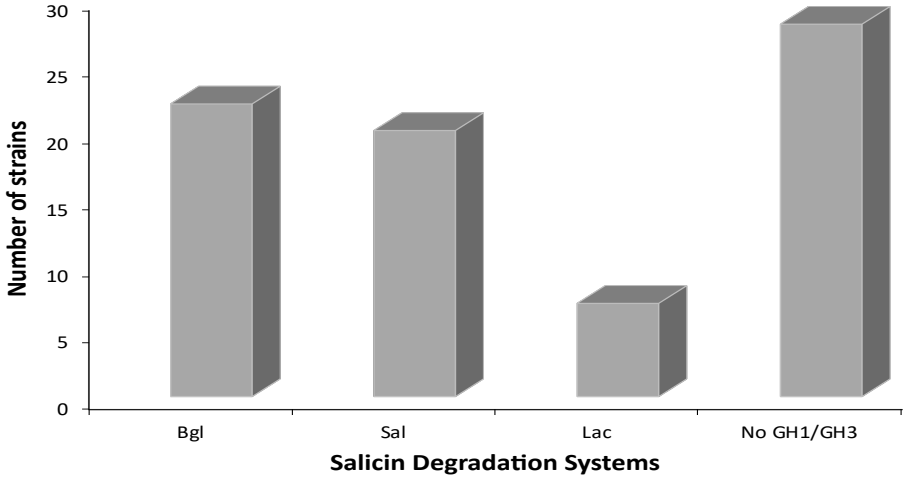


Figure 2.1: Number of strains categorized by predicted salicin degradation systems (Bgl, Sal or Lac). Overall, 45 strains were predicted to possess either Bgl, Sal or Lac system. 22 strains were predicted to contain Bgl system, 20 Sal system and 7 Lac system. Out of seven Lac strains, four also contained Sal system, three of which were members of *Caulobacter* genus. Only 28 strains out of 407 did not contain any detectable glycosyl hydrolase 1 (GH1) or glycosyl hydrolase 3 (GH3) orthologs.

Methylobacterium, *Pseudomonas*, *Tardiphaga* and others from diverse phyla *Proteobacteria*, *Actinobacteria*, *Firmicutes* and *Bacteroidetes* (see attachment S2.8).

Experimental screening demonstrates that approximately 45% of the *Populus* isolates can degrade salicin

We next asked the question who among the 407 genome sequenced isolates could grow with salicin as sole carbon source. To test this, all 407 isolates were grown in minimal media with salicin as sole carbon source in 96-well plates. Of the 407 strains screened, 184 (~45%) were able to grow on salicin (figure 2.2 and attachment S2.10). Strains belonging to genera *Novosphingobium* (12/17 total *Novosphingobium* genomes), *Rhizobium* (19/26 total *Rhizobium* genomes), *Agrobacterium* (5/6 total *Agrobacterium* strains) and *Pseudomonas* (38/59 total *Pseudomonas* genomes) were the noticeable ones that utilized salicin (see attachment S2.10).

Comparing the experimental data and prediction data, it was found that out of 45 strains predicted to contain either Bgl, Sal or Lac system, 32 were experimentally verified to grow on salicin (figure S2.2, attachment S2.10). Seven out of twenty-eight strains predicted to not contain any of the BglB, SalA, SalB or Cbg orthologs were also able to grow on salicin minimal media suggesting presence of yet unknown mechanisms of salicin hydrolysis in these strains.

Two organisms were chosen for further investigation. *Caulobacter* sp. AP07 was selected because it was predicted to contain a *lac* cluster whereas 3 out of 5 *Caulobacter* strains contained both Sal and Lac systems. Upon further examination, this strain seemed to contain an incomplete Sal cluster as well (see figure 2.3c). *Citrobacter* sp. YR018, on the other hand, was selected because during previous growth analysis, it demonstrated growth on salicin minimal medium only after prolonged incubation time (greater than 50 hours).

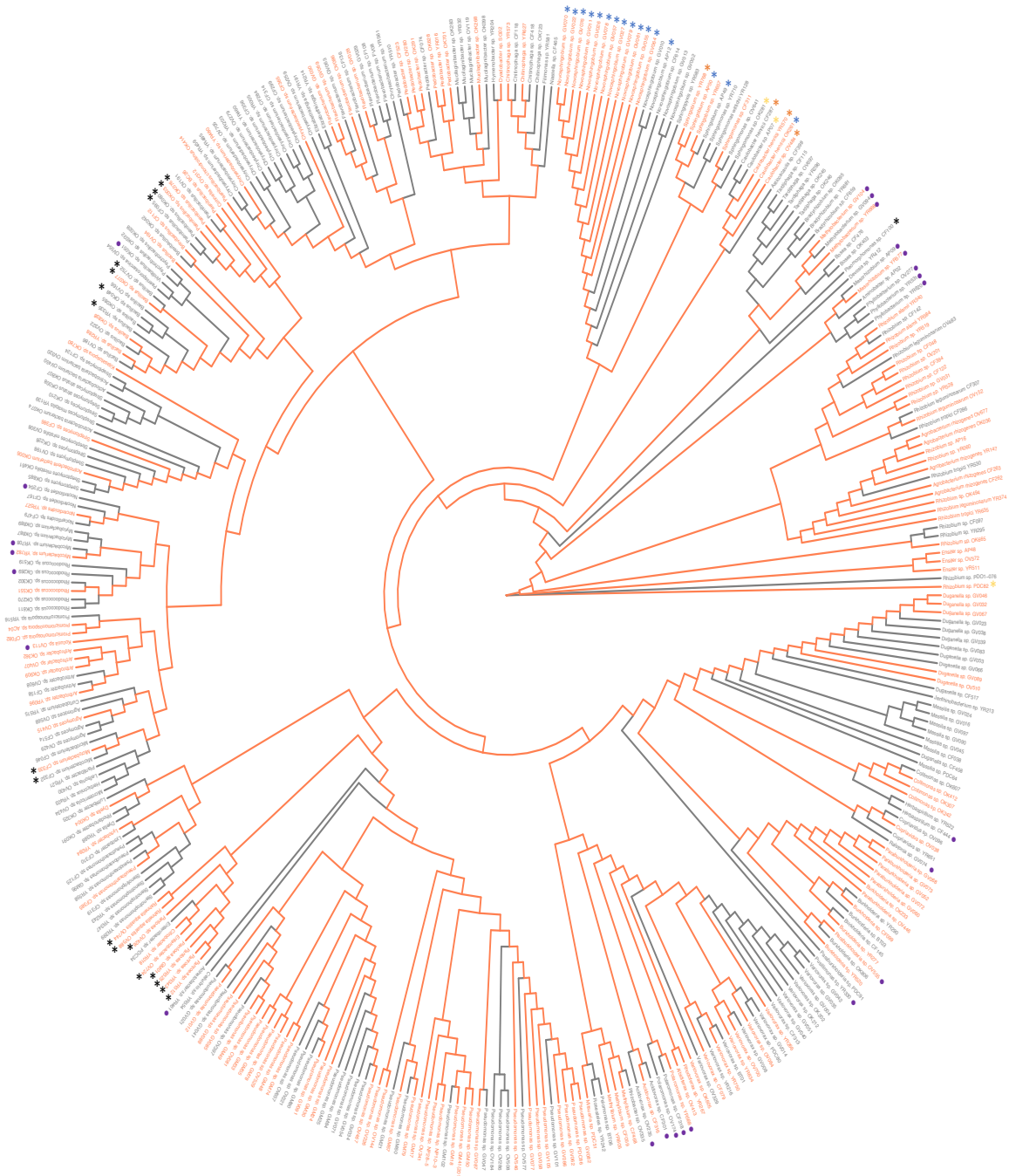


Figure 2.2: Phylogenetic distribution of salicin degrading strains based on the experimental screening assay.
A distance tree was created using JGI/IMG distance tree tool, and then color modified using ggTree package (v. 1.14.6) in R (v. 3.5.1). Salicin degrading strains are highlighted in orange whereas non-growing strains are in black. The prediction results are also added to the figure as stars or circles with color codes: Bgl – Black, Sal – Blue, Lac – Yellow, Sal + Lac – Orange and no GH1/GH3 – Purple.

***Caulobacter* sp. AP07 utilizes salicin and releases salicyl alcohol**

The growth of *Caulobacter* was observed in triplicate on both glucose and salicin minimal media (figure 2.3a). Supernatant samples extracted from the salicin cultures at late log phase were analyzed using HPLC. The metabolite data suggested that salicin was utilized and salicyl alcohol was released (figure 2.3b). Therefore, *Caulobacter* apparently degrades salicin by cleaving glucosyl moiety from salicyl alcohol group. Glucose is then channeled into the central carbon metabolic pathways whereas salicyl alcohol is secreted out into the media. Based on the genome prediction data, *Caulobacter* appears to possess a complete Lac system and an incomplete *sal* operon lacking *salA* homolog (figure 2.3c).

Upon inspection of the AP07 *lac* cluster, PMI01_00694 – PMI01_00697 was determined to resemble that of *C. crescentus* CB15 [73]. However, the length of the potential LacC homolog (PMI01_00694) is only 25 amino acid long and only aligned to the N-terminus of *C. crescentus lacC*. Potential causes of such low alignment length are gene loss, sequencing or assembly errors. So, the only other blastp hit of *lacC* was investigated. The gene PMI01_03448 had 70% blast identity with the amino acid sequence of the query and aligned with C-terminus of *lac*C sequence of *C. crescentus*. Likewise, it contained the required *pfam* domain (Gluconate_2-dh3) found in the *lacC*. The concatenated protein sequence of PMI01_00694 and PMI01_03448 has been provided in supplementary text and is proposed to be the actual *lacC* homolog sequence in AP07 strain (figure S2.3). Possible GH1 homologs (PMI01_04448 and PMI01_02846) were also predicted to be present in *Caulobacter*. All the probable proteins and their detailed information are presented in attachment S2.11.

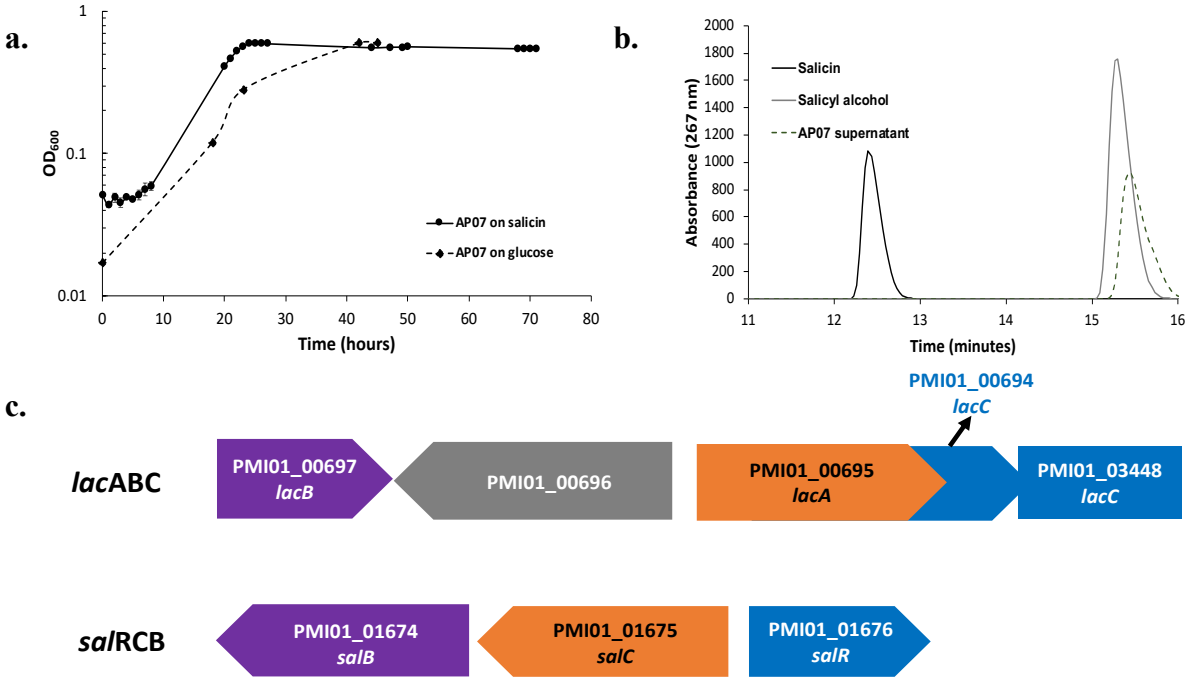


Figure 2.3: *Caulobacter* sp. AP07 was able to utilize salicin and release salicyl alcohol as a by-product. a) *Caulobacter* sp. AP07 was cultured in both glucose and salicin minimal media to determine its growth properties. AP07 was able to grow on both substrates as demonstrated by increase in OD_{600} over time. b) When the supernatant from salicin culture at late-log phase was extracted and analyzed using HPLC, only peak representing salicyl alcohol was observed demonstrating that *Caulobacter* uses only glucosyl group releasing salicyl alcohol into the medium. **c.** Genomic data suggested that *Caulobacter* possessed gene clusters of Lac (*lacABC*) system and an incomplete Sal (*salRCAB*) system. *lacC* ortholog appeared to be missing its C terminal portion because of sequencing/assembly errors but that was easily identified as PMI01_03448 using blastp. Therefore, PMI01_00694-PMI01_03448 is proposed to be a *lacC* homolog. See text for more information.

Proteome data support the hypothesis that *Caulobacter* sp. AP07 utilizes salicin through Lac and Sal systems

To determine the pathway of salicin degradation, proteins extracted from *Caulobacter* sp. AP07 cells grown in salicin and glucose minimal media were subjected to shotgun proteomics using LC-MS/MS. The comparison of the proteomic profiles of glucose and salicin cultures demonstrated higher expression of proteins of both probable Lac and Sal pathways in salicin cultures than those in glucose cultures (see table 2.2). The proteins of the Lac system encoded by PMI01_00694 - PMI01_00697 of AP07 were all significantly upregulated in salicin cultures such that all fold changes were greater than 4 (p-value < 0.05) with the exception of PMI01_00694 (LacC homolog) which was not detected at all most likely because of its small size. The genome-based prediction that PMI01_03448 could encode the C-terminus of LacC was supported by the proteome data as peptides to this portion of the protein were identified (see table 2.2) and determined to be significantly upregulated in salicin cultures (fold change > 5, p-value < 0.05).

Proteins encoded by PMI01_01674 - PMI01_01676, which are probable homologs of *salRCB* were all expressed significantly higher in salicin cultures than in glucose cultures except for the putative transcriptional regulator, SalR. Fold changes were greater than 2 for both SalB and SalC homologs. No probable GH1 proteins (PMI01_04448 and PMI01_02846) were detected in the proteome data. Overall, the proteome data corroborated the genomic prediction that both Lac and Sal systems are functionally active during growth on salicin. See attachment S2.12 for all the proteins detected in this study.

***Citrobacter* sp. YR018 grows on salicin minimal media but only after prolonged incubation**

Initial investigation of the genomic data suggested that *Citrobacter* contained both

Table 2.2 Proteins encoded by *salRCB* and *lacABC* operons were significantly expressed in salicin cultures. Proteomic analysis was performed by comparing the protein abundance between salicin and glucose cultures. Proteins of both Sal and Lac systems were significantly upregulated in salicin. The proteins were also detected in glucose cultures.

Locus tag (Homolog)	Number of peptides identified	Detected in glucose	p-value (BH corrected)	Fold change
PMI01_01675 ("SalC")	35	Yes	0	7.84
PMI01_01676 ("SalB")	23	Yes	0	2.84
PMI01_01674 ("SalR")	1	Yes	0.75	0.38
PMI01_00695 ("LacA")	16	Yes	0	6.19
PMI01_00696 (hypothetical protein)	10	Yes	0	4.72
PMI01_00697 ("LacB")	7	Yes	0	4.47
PMI01_00694 ("LacC")	NA	NA	NA	NA
PMI01_03448 ("LacC")	6	Yes	0.0	5.26

probable BglB (GH1) and SalA/SalB/Cbg1 (GH3) proteins (see attachments S2.13 and table S2.2), any of which could be the hydrolase required for utilizing salicin. Preliminary data indicated that cultures of YR018 had an extremely long lag phase (fifty hours) before detection of growth on salicin minimal media. Therefore, to investigate this observation further, *Citrobacter* was grown on both glucose and salicin minimal media in triplicates. Even though YR018 grew extremely well on glucose achieving an OD_{600} of 0.74 within eight hours, on salicin, this strain again exhibited a fifty-hour lag phase before any increase in OD_{600} was observed. Within the subsequent seventeen hours, the YR018 salicin culture reached an OD_{600} of 0.56 (figure 2.4a). Metabolite data of supernatant from late log phase salicin cultures suggested that *Citrobacter* utilized salicin by cleaving off glucosyl group and channeling to central metabolic pathways (figure 2.4b). Salicyl alcohol is released as a by-product of salicin metabolism.

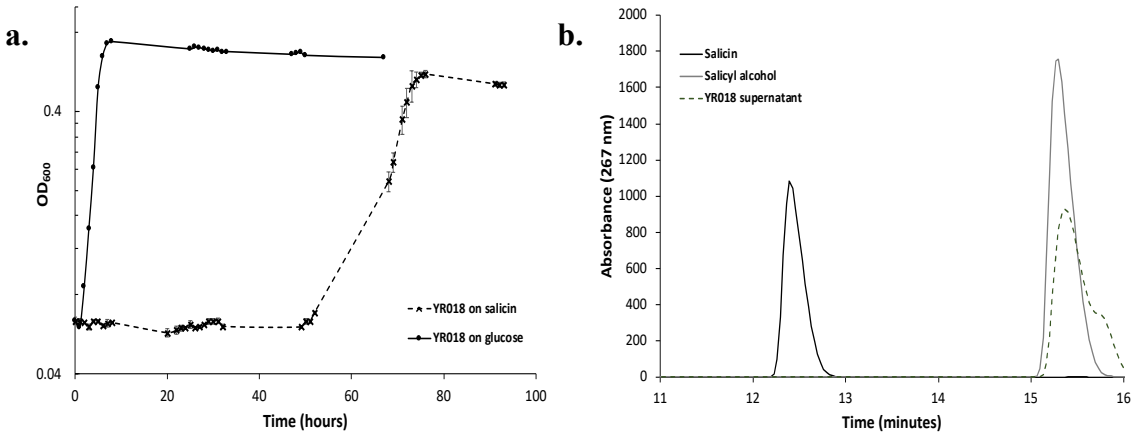


Figure 2.4: *Citrobacter* sp. YR018 grew on salicin, but only after prolonged incubation.
a) YR018 grew on salicin after 50 hours of incubation whereas it was able to grow on glucose right away. b) The supernatants from the salicin cultures were analyzed using HPLC, and it was determined that salicin was utilized by the strain with the release of salicyl alcohol. Salicylate/salicylaldehyde might also have been released in the process as indicated by a second curve right after the salicyl alcohol standard curve.

Growth of *Citrobacter* on salicin is most likely due to mutational activation

Because of the unusual nature of the growth of *Citrobacter* sp. YR018 on salicin, it was investigated further using reintroduction assays to determine whether the growth was due to physiological adaptation or mutational activation. First of all, YR018 was cultured on salicin minimal media, and cells from the stationary phase cultures were subsequently reintroduced in fresh salicin media. In the fresh media, the previously observed lag time was non-existent (figure 2.5a) supporting the case that the "adapted" strain did not require any time to adjust to the new media.

YR018 cells adapted to salicin growth from the previous experiment was preserved as a glycerol stock and stored at -80°C and designated YR018-SAL. Then, the original YR018 strain and YR018-SAL stocks were re-streaked on R2A plates and subsequently sub-cultured in salicin minimal media. A 50-hour lag was observed in the YR018 cultures whereas no lag was seen with

the YR018-SAL strain (figure 2.5b). This result suggested that YR018-SAL did not require any adjustments to grow on salicin compared to YR018 even though both were subjected to similar conditions. The 16S rRNA sequences of both YR018 and YR018-SAL had high sequence identity for trimmed sequences indicating that YR018-SAL strain was not a contaminant but potentially a *Citrobacter* sp. YR018 mutant that had acquired the ability to grow on salicin (see supplementary text).

Possible mutational activation sites on *Citrobacter* are proposed

Nine possible genes that belong to either GH1 or GH3 family have been predicted to be present in *Citrobacter* sp. YR018 (table S2.2). Out of these, three candidates stand out because they contain exactly one gene belonging to respective clusters *bglGFB* or *salRCAB* near GH1 or GH3 homologs. Therefore, these genes/clusters were further investigated. Ga0215739_10977 (probable *bglB* homolog) has a PTS transporter encoding neighboring gene which is flanked by a LacI family transcriptional regulator gene. Blastp was unable to predict it as an antiterminator gene

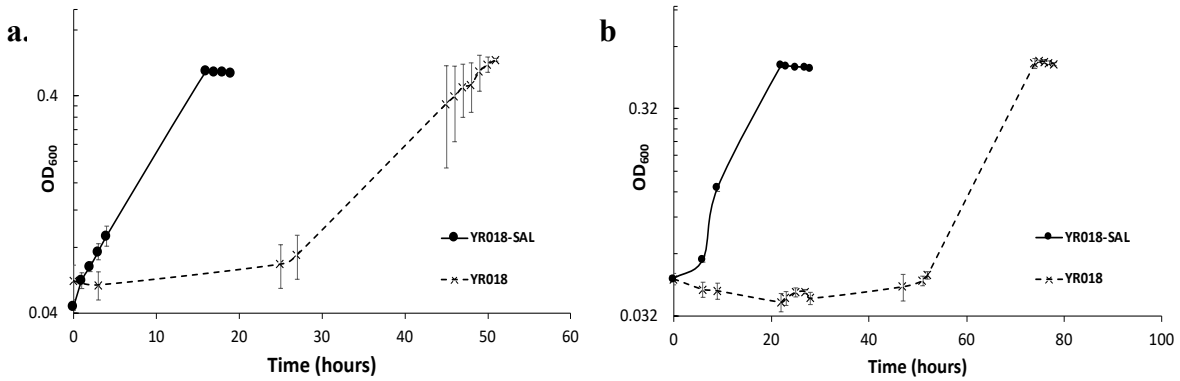


Figure 2.5: Growth of *Citrobacter* on salicin appeared to be due to mutational activation.
a) YR018 was first of all grown on salicin, and then reintroduced in fresh salicin media. YR018-SAL ("salicin-growing" YR018 culture) grew on the fresh media without any lag compared to YR018 on salicin. b) YR018-SAL was stocked in -80°C. Then, it was subcultured on salicin minimal media along with fresh YR018 stock. Their growths were compared, and it was observed that the lag remained for YR018 whereas was non-existent in the case of YR018-SAL. These results support the hypothesis that the growth of *Citrobacter* on salicin is likely due to mutational activation.

(a *bglG* homolog) because of low identity and low query coverage. Regardless, the gene cluster is similar to *bgl* operon. Ga0215739_101543 (another possible *bglB* homolog), on the other hand, has a PTS encoding gene four adjacent genes away. This PTS gene is neighbored by LacI family transcriptional regulator. Finally, *salA* homolog – Ga0215739_101614 has a TonB-dependent receptor protein gene (Ga0215739_101619) five genes away although it did not pass the criteria applied in this study. Overall, mutations in the genes belonging to cluster Ga0215739_10975 - Ga0215739_10977 (*bgl*) or its promoter sequence might be the most likely cause for *Citrobacter* sp. YR018 to have acquired salicin degradation ability. Other possibilities, however, also remain to be determined.

Discussion

An important factor in determining colonization of plant roots by soil microorganisms are metabolic interactions. de Weert and colleagues [107] demonstrated that *Pseudomonas fluorescens* showed a chemotactic response towards tomato root exudates. Similarly, using radiolabeled CO_2 studies, it has previously been demonstrated that the root exudates are utilized and incorporated by soil microbiome members [108]. In a study conducted by Broeckling [109], it was observed that plant exudates are crucial factors in regulating the fungal community. Therefore, root exudates have an important role in microbial colonization.

Populus along with other members of the *Salicaceae* family produces a variety of phenolic compounds that contain a salicin core structure [7]. These phenolic glycosides (PGs) can be up to 30% of foliar dry weight in some of the *Populus* clones [44]. *Populus* secretes PGs such as salicortin along with hydrolases that readily degrade those metabolites to simpler forms such as salicin [7, 55, 60]. Salicin and other complex PGs are potential carbon source for microorganisms

that colonize *Populus* and a possible factor in determination of the *Populus* microbiome structure. In fact, a recent study by Veach et al. [69] demonstrated that PGs such as salicortin and tremulacin can influence *Populus* microbiome.

In this study, we utilized genomics, proteomics, metabolite assays and other quantitative methods for determination of distinct salicin degradation pathways in diverse organisms that associate with *Populus*. Even though degradation of salicin has been reported in a number of studies [2, 73, 75, 86], this is the first study to have investigated these pathways in a systematic fashion for a large number of bacterial strains.

First of all, we mined the genomic data of 407 bacterial isolates from *Populus* in order to predict the presence or absence of four different systems that have been described in the literature to be associated with salicin utilization. In our pipeline, we have used less rigorous parameters for determination of the gene clusters. For instance, even if the genes did not meet the custom additional criteria (alignment length and percent identity) for an individual gene, these genes were reported as possible homologs of the neighboring gene of a reference gene. Similarly, neighboring genes were searched over five genes away from a reference gene because of the less-known nature of these gene clusters. Nonetheless, the overall process is highly rigorous because likelihood of appearance of such cluster is not a random chance and is supported by the fact that only 45 out of 407 strains were predicted to contain salicin clusters— *bgl*, *sal* or *lac*.

The Cbg system consisting of only one gene could not be confirmed to be a likely indicator of presence or absence of the salicin degrading ability. In our analysis, it appeared that considering all four systems would overestimate the number of salicin degraders (~83.5%) and removing Cbg system most likely underestimates the number (~11.8%). Therefore, true number of salicin degraders is predicted to lie somewhere in between. Within the 45 strains that were predicted to

utilize salicin through either Bgl, Lac or Sal system, we observed that Sal and Lac systems are found only in the class *Alphaproteobacteria* (23/23 total strains that contain Sal and Lac systems) whereas the Bgl system is more diverse and found in classes *Gammaproteobacteria* (8/22 total Bgl strains), *Bacilli* (11/22 total Bgl strains), *Actinobacteria* (2/22 total Bgl strains) and *Alphaproteobacteria* (1/22 total Bgl strains). Therefore, results suggest that Sal/Lac system and Bgl system are independently evolved.

We followed the prediction results by experimentally screening all 407 isolates for their ability to grow on salicin minimal media. The results demonstrated that 184 out of 407 strains were able to utilize salicin. All salicin utilizing strains belonged to four phyla – *Proteobacteria* (137 strains), *Bacteroidetes* (17 strains), *Actinobacteria* (16 strains) and *Firmicutes* (14 strains). Within *Proteobacteria*, three classes *Alphaproteobacteria* (49 strains), *Gammaproteobacteria* (50 strains) and *Betaproteobacteria* (38 strains) are represented. This observation is consistent with a few studies that have demonstrated that PGs can influence the bacterial population of phyla *Actinobacteria, Bacteroidetes, Firmicutes* and *Proteobacteria* [56, 69].

Comparing the predicted 45 salicin degraders (containing either Bgl, Sal or Lac systems) with experimental data, we found a ~71% match. It should also be noted that growth of some of the strains on salicin minimal media might not be possible because of other nutrient requirements. Furthermore, some strains require longer incubation times to grow on salicin. Another consideration is the cutoff used for assigning growth to the strains. Therefore, the experimental data should be used as a guide for future use but with caution. Moreover, the observation that the strains that did not possess any GH1 or GH3 homologs but were able to grow on salicin minimal media suggests that these strains might have other mechanisms of degrading salicin and should be investigated further to find novel salicin hydrolases.

Following experimental screening, we identified *Caulobacter* sp. AP07 and *Citrobacter* sp. YR018 for further study. We chose *Caulobacter* sp. AP07 because it was predicted to possess both Lac and Sal systems that have been demonstrated to be involved in salicin utilization. *Caulobacter crescentus* CB15 has been studied for salicin degradation in one study in which researchers demonstrated that salicin induced the expression of *lac* genes, and lac genes were required for growth on agar plates containing salicin as sole carbon source [73]. Other strains of *Caulobacter* have also been reported to have salicin degradation ability, but the potential system required for salicin degradation was not identified in those studies [110, 111]. The molecular mechanism of salicin degradation through lactose dehydrogenase has not been fully elucidated yet, and such endeavor is beyond the scope of this study.

Another reason for choosing *Caulobacter* sp. AP07 was the presence of an incomplete Sal system. *Azospirillum irakense*, a strain taxonomically close to *Caulobacter,* consists of *salRCAB* operon containing functionally redundant genes *salA* and *salB* that encode for GH3 family proteins required for salicin cleavage [72]. Therefore, presence of either gene should provide the strain with salicin degradation ability, and it appears to be the case for *Caulobacter* sp. AP07. An untargeted proteomics comparing glucose and salicin cultures of this strain demonstrated that the proteins of both Sal and Lac pathways were significantly upregulated in salicin cultures than those in glucose cultures suggesting that both pathways are active when salicin is provided.

Genomic and experimental data for *Caulobacter* pose a significant question – do these two pathways work independently or together? We checked all the genomes predicted to contain either Sal or Lac system to identify if there were instances of these genomes containing both systems that we missed. We found three additional genomes (including *Caulobacter* sp. AP07) other than the previously identified four that might contain both systems (see table S2.3). Therefore, it is likely

that the presence of Lac system is required for strains containing Sal system to degrade salicin. The opposite, however, might not be true. In strains in which Sal system is absent (i.e. only Lac system present) such as *Rhizobium* sp. PDC82 (a salicin degrader), other glycosyl hydrolases might replace the function of SalA or SalB.

Lactose dehydrogenase complex consists of enzymes that are not traditionally associated with bond cleavage activity. New research, however, has demonstrated that dehydrogenases could be important in the production and function of beta-glucosidases that act on compounds such as salicin [112]. Studies have demonstrated that cellobiose dehydrogenases can oxidize cellobiose in order to reduce the end product inhibition of cellulases [113, 114]. Additionally, a few studies have demonstrated that salicin can be oxidized at the C3, C4 or both positions by pyranose dehydrogenase [115, 116]. Therefore, it is possible that lactose dehydrogenase oxidizes salicin which can then be acted upon by glycosyl hydrolases such as SalA or SalB enzymes. This hypothesis is further corroborated by a recently published work by Miyazaki and colleagues [117]. In this study, it was demonstrated that the lactose dehydrogenase complex acts in the periplasmic space where it oxidizes the substrate. The Sal system has also been proposed to be active in the periplasm [86]. Arellano and colleagues [73] have suggested that the *Lac* system works in tandem with ketoglucosidases to degrade salicin. Therefore, lactose dehydrogenases might play an important role in salicin degradation by acting in conjunction with hydrolases such as those in the Sal system, but more work is required to test this hypothesis.

Citrobacter sp. YR018 was another organism that was characterized in this study. This strain grew on salicin but only after 2 days. Therefore, we investigated whether the growth was due to adaptation or mutation. The results from reintroduction assays and 16S rDNA sequencing supported the mutational activation hypothesis. Spontaneous mutations in bacteria leading to

salicin utilization is not uncommon [1, 3, 70, 118, 119]. In *E. coli*, the most widely studied system is *bglGFB* whose promoter sites have been observed to be mutated to lead to the gain of salicin utilization function. In *Shigella sonnei,* multiple mutational steps in the *bglB* gene were required for it to be able to grow on salicin [3]. Similarly, mutations in *cel* and *sac* loci have also been proposed to lead to salicin degradation in *E. coli* [118, 119]. Therefore, it is difficult to identify the exact loci that was mutated in *Citrobacter*. We found three candidates, Ga0215739_10977 (*bglB*), Ga0215739_101543 (*bglB*) and Ga0215739_101614 (*salA*) that could potentially be involved. The gene cluster that Ga0215739_10977 is a part of is a *bgl* operon-like structure and might be the most likely candidate that should be pursued first. Overall, this part of the study reveals that bacteria that associate with *Populus* might acquire salicin degradation ability through mutations over their evolutionary history. Therefore, this might be an example of host-mediated selection pressure that forces microbial isolates to adapt in order to colonize their host over long periods of time.

In this study, we have presented a comprehensive guide to identify and study salicin degrading strains associated with *Populus*. Four mechanisms of salicin degradation were investigated. Using both genomic and screening data, we have demonstrated that significant number of *Populus*-associated strains have salicin degradation ability. We chose to focus on two different strains – *Caulobacter* and *Citrobacter* and were able to define two distinct salicin degradation systems, one composed of multiple systems and the other acquired through mutations. Out of these four systems, we could not confirm whether one pathway has distinct advantage over the other or if two different pathways can act together in microbes. These questions are beyond the scope of this study. Regardless, this is one of the first studies to have systematically

investigated distinct salicin degradation mechanisms in *Populus*-associated strains and could provide valuable insights for future research.

Chapter 3

The salicin degradation pathway in *Populus*-isolate *Agrobacterium rhizogenes* OK036

Introduction

Numerous research investigations have demonstrated that the microbiome can help the host plant by extending its abilities such as increasing drought tolerance, nutrient availability and pathogen resistance [120-122]. Therefore, understanding microbial association with its host is an important area of study. In particular, metabolic interactions have been proposed to be crucial in the initiation and stabilization of the host-microbe relationship [123].

Populus and members of the *Salicaceae* family produce and release salicylate-containing compounds called phenolic glycosides (PGs) such as salicortin and tremulacin. Some *Populus* clones have been reported to produce these metabolites at concentrations as high as 30% of dry foliar biomass [44]. Therefore, in order to fully comprehend the microbial colonization process of *Populus*, it is important to understand how these compounds are utilized by *Populus*-associated microbes.

Salicin is a glucosylated form of salicyl alcohol and is the simplest PG. The two functional groups are bonded by β-1,4-D-glycosidic bond. The enzymatic hydrolysis of this compound occurs primarily at this bond through the action of glycosyl hydrolase 1 (GH1) and glycosyl hydrolase 3 (GH3) enzymes [1, 72, 124]. In the present study, we characterize the salicin degradation pathway of *Agrobacterium rhizogenes* OK036, a previously isolated *Populus* root endophyte.

Agrobacterium strains are gram-negative, rod-shaped members of *Alphaproteobacteria* class. Members of the genus *Agrobacterium* are commonly plant-associated and contain both non-pathogenic and pathogenic species [125]. Strains such as *A. tumefaciens*, *A. vitis* and *A. rhizogenes* are well-known plant pathogens causing neoplastic diseases like crown gall and hairy root diseases [126]. Their pathogenic activity is facilitated by the insertion of bacterial genes into plants which causes uncontrolled cell proliferation and production of compounds that are utilized by the bacteria

[127]. Chemotaxis has been proposed to be an important feature of these bacteria which are attracted to plant chemicals such as sugars and phenolic compounds [125, 128]. Therefore, host-*Agrobacterium* metabolic interaction appears to play a key factor in the colonization of the host.

PG degradation in *Agrobacterium* and other organisms has been investigated in a few studies. Coniferin hydrolase (Cbg1) of *Agrobacterium tumefaciens* is a GH3 family enzyme and has demonstrated activity against a number of substrates including salicin [75, 88]. Similarly, GH1 enzymes such as phosphoglucosidase (BglB) have been reported to be involved in the degradation of salicin in organisms such as *Escherichia coli* and *Klebsiella aerogenes* [1, 2]. In our previous work (described in chapter 2), we learned that both BglB (GH1) and Cbg1 (GH3) orthologs are present in OK036 strain. Therefore, both types of enzymes were investigated in this organism.

In the current project, we performed growth and supernatant metabolite analyses of *Agrobacterium rhizogenes* OK036 culture in salicin minimal media to demonstrate that this strain is able to utilize both the glucosyl and salicyl moieties of salicin. We applied bioinformatics and comparative proteomics to propose a pathway for the degradation of salicin by OK036. The enzymes of the proposed pathway are encoded by an eleven member gene cluster. In the proposed pathway, salicin is first cleaved into salicyl alcohol and glucose. Salicyl alcohol is then converted to salicylate which is channeled to the gentisate pathway through salicylyl-CoA and gentisyl-CoA intermediates. Metabolite data examining the inner metabolites of the salicin cultures corroborated the pathway indicated by proteome data. The mechanism described in this study is a novel salicin degradation system that consumes salicin completely. The salicin degradation pathways that were studied in the past to the best of our knowledge involve utilization of only the glucosyl moiety of salicin [1, 2, 71, 129, 130]. Another important finding of the study is a unique enzyme annotated

as benzoate-CoA ligase family protein which is further characterized using bioinformatic approaches.

Methods and materials

Bacterial strains and culture medium

The bacterial strain (*Agrobacterium rhizogenes* OK036) used in this study was isolated from the endosphere of *Populus* root tissue and its genome was sequenced, assembled, and annotated by the Department of Energy (DOE) Joint Genome Institute (JGI) [92]. The genome is available in both National Center for Biotechnology Information (NCBI) (RefSeq: NZ_JQJX00000000.1) and Integrated Microbial Genomes (IMG) (Genome ID: 2582581316) databases. The strain was stored in freezer glycerol stock and streaked onto R2A (Franklin Lakes, BD Difco) plates whenever necessary.

For determination of growth, R2A liquid medium was inoculated with individual colonies picked from the R2A plates. The cultures were incubated at 25°C with gentle shaking at 200 rpm. The overnight cultures were subcultured in 3-(N-morpholino)propanesulfonic acid (MOPS) minimal media (pH: 7.2) with the desired carbon substrate (glucose, salicin, salicyl alcohol, salicylaldehyde, salicylate, gentisate and catechol) at 3 mM concentration [94]. For growth measurements, optical density at wavelength 600 nm (OD_{600}) was used in the spectrophotometer – GENESYS 30 (Thermo Fisher Scientific). The starting OD_{600} of subcultures was between 0.03 – 0.07.

High performance liquid chromatography (HPLC) metabolite analysis

Cultures growing in triplicates in MOPS minimal medium containing salicin as sole carbon

source were harvested at late log phase and pelleted by centrifugation at speed 10621 x g for 5 minutes in Eppendorf centrifuge 5417R. Culture supernatants were collected and stored at -20ºC. Frozen samples were thawed and vortexed before 200 µl being diluted with 500 µl purified water in scintillation vials. HPLC metabolite analysis was performed as described in Shivatare et al. [95]. Briefly, Agilent 1260 Infinity HPLC with C18 column (Thermo Scientific Hypersil Gold 200/4.6 mm, 5 micron) was used for chromatography. For mobile phase, 0.1% (v/v) orthophosphoric acid (pH 2.5) and 100% (v/v) acetonitrile were used. The flow rate was maintained at 1 ml/min and wavelength was set at 267 nm. For identification purposes, standard solutions of salicin and salicyl alcohol were also measured.

Sample preparation for proteome measurements

Individual colonies were picked from the R2A plates and grown overnight on R2A liquid medium. Then, the cultures were subcultured in triplicates on glucose and salicin (3 mM) minimal media. When the OD_{600} reached close to 0.7, the cultures were harvested by centrifugation at 11952 x g (15 mins, 4ºC), flash frozen in liquid N_2 and stored at -80ºC. For reducing proteins, pellets were thawed, washed and resuspended in lysis buffer (500 µL of 4% sodium deoxycholate (SDC) in 100 mM ammonium bicarbonate (ABC, pH 8.0) and 10mM dithiothreitol (DTT)). Lysate was bead-beaten by transferring to eppendorf tube containing a scoop of zirconium oxide beads followed by centrifugation at 21000 x g for 10 minutes and transferred to a new tube. Following measurement of crude protein concentration, alkylation was carried out with 30 mM indole acetic acid (IAA), and then samples were incubated in the dark for 30 minutes. Sample loading (total 200 µg of protein) was done on 10Kda filter (Vivaspin 500) and washed with 500 µL of 100 mM ABC by spinning down at 10000 x g (~2% SDC concentration). Proteins were digested to peptides using

two sequential aliquots of sequencing-grade trypsin (Promega Corp.) using trypsin-protein ratio (w/w) of 1:75 followed by an overnight incubation and another 4 hours of incubation at room temperature. Using molecular weight cut-off (MWCO) membrane, the tryptic peptide solution was spun-filtered. In the same step, 1% formic acid was used to precipitate residual SDC. The SDC precipitate was then removed from the peptide solution using water-saturated ethyl acetate extraction. SpeedVac (Thermo Fisher Scientific) was applied to concentrate peptide samples which were quantified by nanodrop (Thermo Fisher Scientific) using Scopes method [96]. Peptides were finally adjusted to a final concentration of 0.5 µg/µl before being loaded in the autosampler vial for MS analysis.

Peptide samples were analyzed by automated 1D LC-MS/MS analysis. Vanquish UHPLC was plumbed directly in-line with a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) outfitted with a back column (10 cm of 5 µm Kinetex C_{18} reverse phase (RP) resin (Phenomenex)) coupled to an in-house nanospray emitter packed with 30 cm of 5 µm Kinetex C_{18} RP resin (Phenomenex). 3 µg of peptides was loaded onto the column, desalted and separated using 150-minute organic gradient as described elsewhere [97]. Eluted samples were measured and sequenced in a data-dependent approach on the Q Exactive MS.

MS data analysis and evaluation

Proteome Discoverer, (v.2.2, PD) (Thermo Fisher Scientific) with MS Amanda (v.2.2) [98] and Percolator [99] were utilized to collect and process all the MS/MS spectra. First of all, the spectra were recalibrated to improve the feature mapping. An *Agrobacterium rhizogenes* OK036 database was created to which data for common laboratory contaminants were added. Spectral data were searched against this database. To identify the tryptic peptides, these parameters were set in

MS Amanda: MS1 tolerance = 5 ppm; MS2 tolerance = 0.02 Da; missed cleavages = 2; Carbamidomethyl (C, + 57.021 Da) as static modification; and oxidation (M, + 15.995 Da) and carbamylation (N-terminus, + 43.006 Da) as dynamic modifications. The FDR threshold was set at 1% at both peptide-to-spectrum matches (PSMs) and peptide level. Both unique and/or razor peptides were required for quantification of a protein. For grouping proteins, Strict Parsimony Principle was used, and proteins were normalized using PD's normalization feature. Finally, a student t-test (p-value < 0.05) was performed between glucose and salicin groups to determine significantly abundant proteins in either condition.

Preparation of cell lysates for metabolite detection

Cell lysates were prepared using the protocol described in study elsewhere [131]. Briefly, strains were subcultured in salicin and glucose minimal media (3mM) in 300 ml volumes to early – mid log phase. The cultures were pelleted by centrifugation at 11952 x g for 15-20 minutes. Pellets were washed twice with S30 buffer (14 mM magnesium acetate, 60 mM potassium glutamate, 1 mM dithiothreitol (DTT) and 10mM Tris-acetate, pH 8.2). Following final wash step, pellets were weighed, flash frozen in liquid N_2 and stored at -80°C.

For cell-free extracts, the pellets were thawed and resuspended in 0.8 µl S30 buffer per mg of wet cell pellet weight. After resuspension, samples were sonicated at 530 J per ml of suspension at 50% tip amplitude on ice water. Cell-slurry was centrifuged once for 10 minutes at maximum speed at 4°C before extraction and aliquoting of supernatant. The sample aliquots were flash frozen and stored at -80°C. These samples were either used for cell-free reaction or used directly for metabolite analysis.

For cell-free reaction, the extracts were thawed and then 10 mM of either salicin or gentisate was added to each sample. Samples were incubated in a shaking incubator (200 rpm) at 25°C for either 1 hour or 24 hours before being flash frozen in liquid N_2. The samples were stored at -80°C until they were used for metabolite analysis.

Sample preparation for metabolite analysis

To the cell extracts (approximately 50 µL in volume each), 50 µL of extraction solvent (40:40:20 acetonitrile (ACN): methanol (MeOH): water (H_2O) with 0.1 M formic acid (FA) and 25 µM phenylhydrazine (PH)) was added. Addition of PH was to allow for improved signal quality of certain α-keto acid species of interest in the cell extracts, as per the cited method [132]. Samples were shaken on ice for an hour before centrifugation (13000 x g at 4°C for 15 minutes). The supernatant was removed and stored separately at 4°C until LC-MS analysis.

Nano-LC MS/MS

Data was obtained using a Dionex UltiMate 3000 HPLC pump with an auto sampler (Thermo Fisher Scientific) coupled to a Linear Trap Quadrupole (LTQ)-Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific) with a nano-electrospray ionization source (Proxeon). Measurements were carried out in negative-ion mode under the supervision of the XCalibur software package (Thermo Fisher Scientific).

Separations were performed on 75 µm inner diameter fused-silica (Polymicro Technologies) columns with electrospray tips laser-pulled in-house by a Sutter Instrument P-2000 micropipette puller. Columns were pressure packed under helium gas to approximately 30 cm with Kinetex C18 resin (5 µm, 100 Å, Phenomenex). Mobile phase solvents were selected for optimal

negative-mode separations: mobile phase A (95:5 H2O: isopropyl alcohol (IPA) + 1 mM acetic acid (AA)) and B (65:30:5 ACN:H2O:IPA + 1 mM AA), with acetic acid being used as a favorable negative mode ionizing agent as per the cited method [133]. A simple gradient from 100% A to 100% B over 20 minutes was used, holding for 2 minutes at 100% B before returning to 100% A over 3 minutes and column equilibration for 15 minutes. 5 µL of the sample were injected via auto sampler onto the column with a split-flow setup allowing for nano-flow rates for nano-electrospray analysis. The HPLC pump flow rate was set to 0.1 mL/min, which translated to approximately 250 nL/min at the column tip.

For measurement metrics, the instrument method was set up and calibrated as follows. The electrospray ionization (ESI) source capillary temperature and voltage were monitored and optimized in conjunction with calibration of a new column, with optimal conditions being set to 250°C and 4.0 kV, respectively. Full precursor scan measurements were acquired in centroid mode, with a resolving power of 30,000 over a mass range of 50 to 1800 m/z obtained via the Orbitrap. Fragmentation on the top ten ions of each full scan was performed via collision-induced ionization with helium gas in the ion trap with an isolation width of 2 m/z and a 30% normalized collision energy. Precursor ions were placed onto dynamic exclusion lists for two minutes.

LC-MS/MS metabolite data analysis

Raw files were analyzed using XCalibur's Qual Browser. Targeted compounds of interest include pyruvic acid and related derivatives (fumaric acid, fumarylpyruvic acid, and maleylpyruvic acid), salicin and related derivatives (salicylaldehyde, salicyl alcohol, and salicylic acid), glucose, gentisic acid, and coenzyme A (CoA) species (salicylyl-CoA and gentisyl-CoA). Monoisotopic masses for each of the compounds of interest were calculated, and the extracted ion chromatograms

for each compound were examined to determine the presence (or lack thereof) of the desired compounds of interest. As noted earlier, for α-keto acid species of interest, a mass addition of 90.0582 Da is needed to be taken into consideration as a result of the phenylhydrazine reaction.

Genomic data analysis

Glycosyl hydrolase 1 (GH1) or 3 (GH3) genes were identified using local blastp tool [101] on Biopython (v 1.71) [102] using the protocol described elsewhere [129]. For each gene, length of alignment was set to 30% and percent identity was set to 25% as added parameters. If the two thresholds were passed, a blastp gene hit was considered a homolog.

Blastp was also used to identify putative homologs of genes encoding the proposed salicin pathway enzymes using the amino acid sequences of proteins involved in salicylate degradation in *Streptomyces* sp. WA46 as queries. Functional domains (Protein Familes (pfam)) were also searched for each individual protein of the proposed pathway using IMG database.

For salicylate 5-hydroxylase (CoA forming) or S5HCoA homolog analysis, blastp was used as described before using *Agrobacterium* protein sequence as query against all 407 *Populus*-associated microbial genomes. Additional criteria – length of alignment ≥ 50% of length of query sequence and identity ≥ 30% were also used to identify a gene hit as a homolog. Following this, each probable homolog was searched on IMG database for protein families (TIGRFAM). The strains possessing proteins that contained the right TIGRFAM domains were considered further to narrow the search criteria. A few representative strains were investigated manually on the IMG database to identify the neighboring genes of their S5HCoA homolog.

Results

Agrobacterium rhizogenes OK036 is able to completely utilize salicin for growth

When *Agrobacterium* was grown on salicin minimal media, it was able to utilize salicin at a slow growth rate (~ 0.068 hr^{-1}) but was able to grow to a higher OD_{600} (0.763) on salicin than on glucose (0.487) minimal media (see figure 3.1a). This observation was consistent with *Agrobacterium* in multiple growth experiments (see figure S3.1) suggesting the complete utilization of the salicin molecule instead of just the glucose moiety. Therefore, *Agrobacterium* was chosen for further characterization. Metabolite profiling of supernatants of late-log phase OK036 cultures grown on salicin minimal media was performed using HPLC. When growing on salicin, *Agrobacterium* appeared to not release salicyl alcohol as seen with most other strains that have been analyzed in our lab (figure 3.1b). Taken together, these results strongly suggest that *A. rhizogenes* OK036 utilizes both the glucosyl and salicyl moieties of salicin.

Salicin cleavage can occur through either GH1 (BglB) or GH3 (Cbg1) enzymes (see table S3.1). One GH1 homolog was identified through blastp – EW94DRAFT_05840 (percent identity:

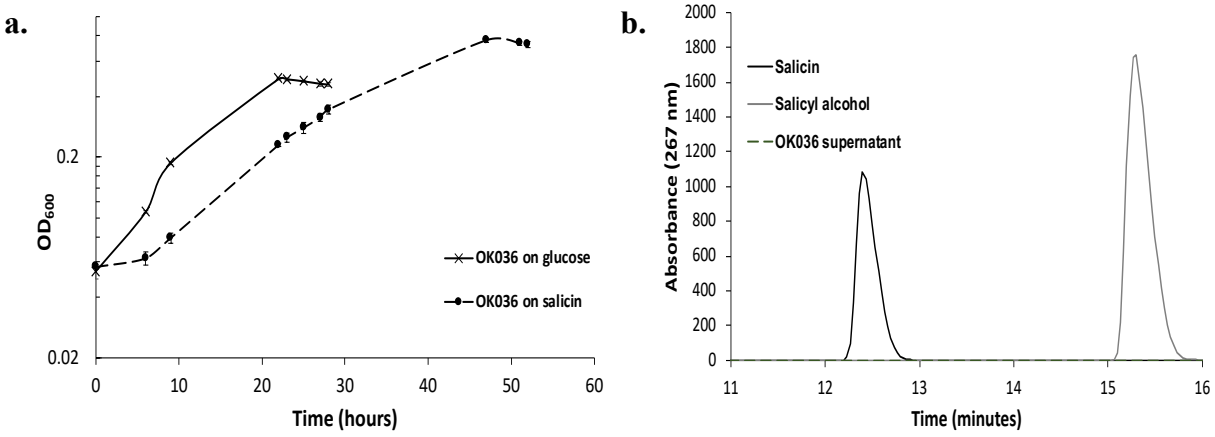


Figure 3.1: *Agrobacterium rhizogenes* OK036 utilized salicin completely.
a) *Agrobacterium rhizogenes* OK036 was cultured in both glucose and salicin minimal media. OK036 strain was able to grow on both media, but its OD_{600} in salicin was almost double that in glucose. b) When the supernatant from salicin culture at the late-log phase was extracted and analyzed using HPLC, no peaks for either salicin or salicyl alcohol was observed. This indicates that *Agrobacterium* utilizes salicin completely by using both the glucosyl and salicyl alcohol moieties for energy generation.

~32%). Likewise, three genes – EW94DRAFT_06426, EW94DRAFT_05475, EW94DRAFT_03783 were predicted to encode GH3 enzymes. However, only EW94DRAFT_06426 out of these three had a complete match (query coverage: 100%) to the *cbg1* amino acid sequence of *Agrobacterium tumefaciens* (now *Rhizobium radiobacter*) with ~77% amino acid identity. See table S3.1 and attachment S3.1 for details on each gene.

A unique salicin degradation pathway is hypothesized for *Agrobacterium rhizogenes* OK036 through proteomic measurements

For determination of probable proteins involved in salicin degradation, an untargeted shotgun proteomic analysis was performed on both salicin and glucose cultures. The analysis of proteins with the highest fold change between the two substrates direct to a gene cluster EW94DRAFT_05420 - EW94DRAFT_05430 as eight enzymes encoded from this cluster were significantly upregulated with fold change > 3 (p-value < 0.05) for each protein (Table 3.1). The glycosyl hydrolases – EW94DRAFT_05840 (GH1) and EW94DRAFT_06426 (GH3) were detected but not significantly upregulated in either culture. Other GH3 proteins EW94DRAFT_05475 and EW94DRAFT_03783 were not detected at all. See attachment S3.2 for complete set of proteome data and S3.3 for annotation data for each gene in the gene cluster identified.

The aforementioned gene cluster (EW94DRAFT_05420 - EW94DRAFT_05430) appears to be involved in the conversion of salicyl alcohol all the way to fumarate and pyruvate through salicylate and gentisate intermediates (figure 3.2). The gene cluster is divided into two subclusters: EW94DRAFT_05427 - EW94DRAFT_05430 (SC1) and EW94DRAFT_05421 - EW94DRAFT_05426 (SC2). SC1 is possibly involved in the conversion of salicyl alcohol to

Table 3.1: Comparison of abundance of proteins between salicin and glucose cultures of *Agrobacterium rhizogenes* OK036 demonstrated involvement of a cluster of eleven genes most likely required for salicin degradation.

Proteins EW94DRAFT_05421 - EW94DRAFT_05428 were significantly expressed in salicin cultures and were also detected in glucose cultures. EW94DRAFT_05429 and EW94DRAFT_05430 were not detected in either condition. The proteins are encoded by genes belonging to an eleven-member gene cluster which probably is involved in salicyl alcohol degradation. The GH1 and GH3 homologs were detected but not significantly expressed in either culture suggesting the hydrolase encoding genes are not induced by salicin.

Locus tag (Homolog)	Number of peptides identified	Detected in glucose	p-value (adjusted)	Fold change
EW94DRAFT_05840 (GH1)	13	Yes	0.01	-0.73
EW94DRAFT_06426 (GH3)	20	Yes	0.02	0.18
EW94DRAFT_05420 (IclR family regulator)	2	Yes	0.06	0.61
EW94DRAFT_05421 (4- Hydroxybenzoyl-CoA thioesterase)	5	Yes	0.01	3.43
EW94DRAFT_05422 (Benzoate-CoA ligase family)	36	Yes	0.01	7.94
EW94DRAFT_05423 (Fumarylpyruvate hydrolase)	11	Yes	0.00	8.48
EW94DRAFT_05424 (Maleylpyruvate isomerase)	10	Yes	0.00	7.67
EW94DRAFT_05425 (Gentisate 1,2-dioxygenase)	16	Yes	0.00	9.32
EW94DRAFT_05426 (Maleylpyruvate isomerase)	8	Yes	0.00	7.28
EW94DRAFT_05427 (Choline dehydrogenase)	32	Yes	0.00	5.67
EW94DRAFT_05428 (Acyl-CoA reductase)	34	Yes	0.00	8.70

salicylate whereas SC2 seems to be involved in salicylate to fumarate/pyruvate transformation. SC2 has been previously characterized in *Streptomyces* sp. WA46 strain for the degradation of salicylate [134]. Blastp was also used to identify the potential homologs (in *Agrobacterium*) of genes encoding proteins of salicylate pathway in *Streptomyces* (table S3.2). Using the information from the proteome data, blastp data and literature, a salicin degradation pathway in *A. rhizogenes* OK036 is proposed in figure 3.2c.

One of the significant differences between *Agrobacterium* and *Streptomyces* pathways is a protein encoded by the gene EW94DRAFT_05422 and annotated as benzoate-CoA ligase family protein. When the amino acid sequence of gene (*sdgA*) encoding salicylate-CoA ligase (AMP forming) enzyme in *Streptomyces* which catalyzes the conversion of salicylate to salicylyl-CoA was queried against the *Agrobacterium* genome, the sequence aligned with C-terminus of EW94DRAFT_05422 at 24% identity. Surprisingly, *sdgC* (encoding salicylyl-CoA 5-hydroxylase catalyzing hydroxylation of salicylyl-CoA to gentisyl-CoA) amino acid sequence aligned with N-terminus of EW94DRAFT_05422 at 38% identity. Upon further inspection, pfam domains – FAD_binding_3 (in SdgC), AMP-binding and AMP-binding_C (in SdgA) were all detected in EW94DRAFT_05422 (figure 3.2b). Therefore, genomic data indicates that protein encoded by EW94DRAFT_05422 possesses dual function: CoA ligase and hydroxylase/monooxygenase. This suggests that in contrast to *Streptomyces* sp. WA46 which requires two enzymes for the conversion of salicylate to gentisyl-CoA, *A. rhizogenes* OK036 encodes a single polypeptide containing both enzyme activities. Thus, protein encoded by EW94DRAFT_05422 could be annotated as salicylate 5-hydroxylase (CoA forming) or S5HCoA.

Another difference between the *Streptomyces* and *Agrobacterium* pathways involves the third reaction (in *Streptomyces* salicylate pathway) in which the thioester group of gentisyl-CoA

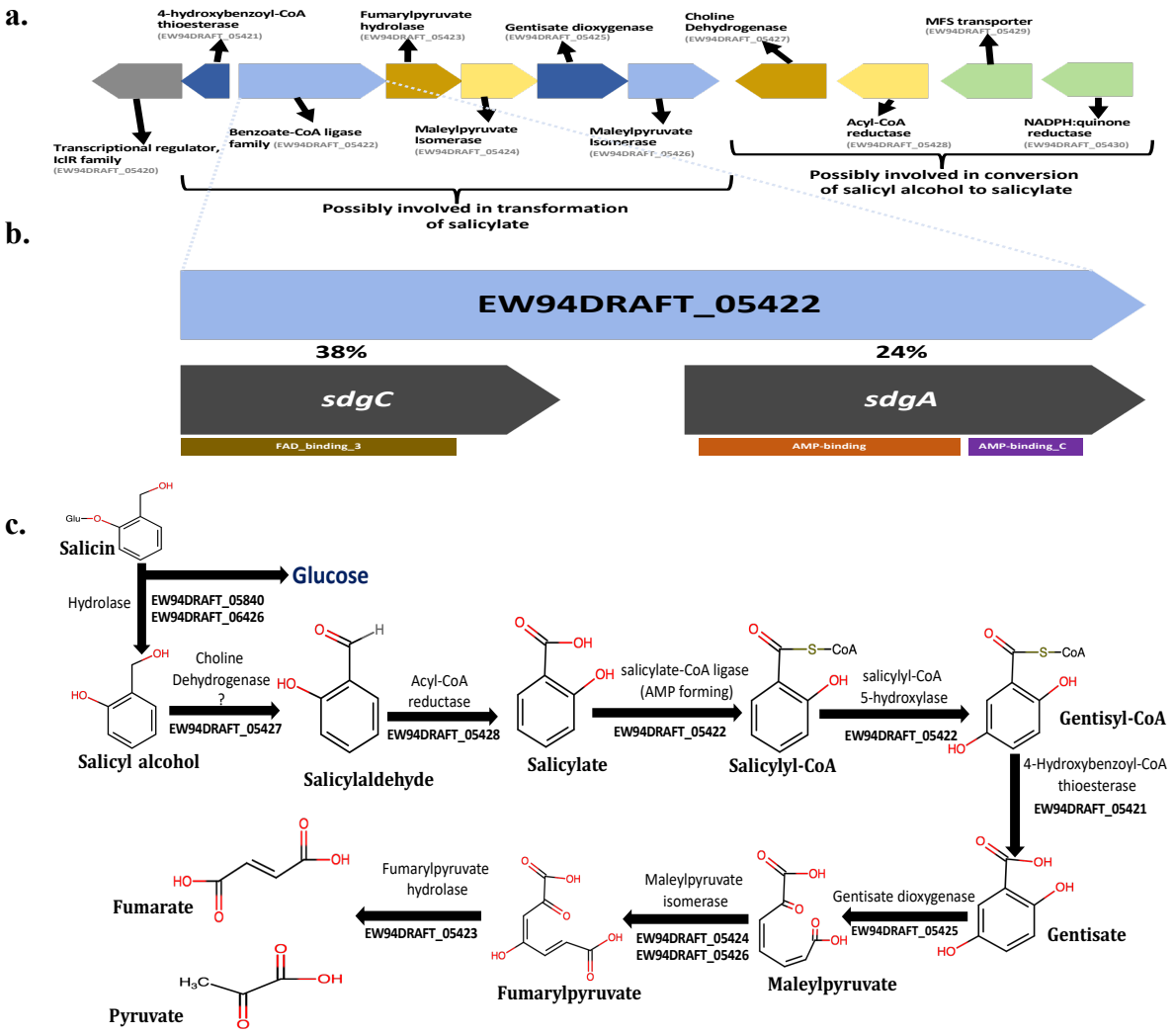


Figure 3.2: Proposed pathway of salicin degradation in *Agrobacterium rhizogenes* OK036.
a) Gene cluster identified using proteomics data that possibly encodes part of the proposed pathway. b) The protein sequences of *sdgA* (encoding salicylate-CoA ligase (AMP forming)) and *sdgC* (encoding salicylyl-CoA 5-hydroxylase) aligned to C-terminus and N-terminus of EW94DRAFT_05422, respectively. The blast identity percentages are shown in the figure. The combined pfam domains of SdgA and SdgC were the same as those of EW94DRAFT_05422 protein suggesting dual function of *Agrobacterium* protein. c) Based on the proteome comparison data and literature, a salicin degradation pathway has been hypothesized for *A. rhizogenes* OK036. In the proposed pathway, salicin is first converted to glucose and salicyl alcohol by the action of glycosyl hydrolases. Salicyl alcohol is oxidized to salicylate, which then proceeds to be transformed to gentisate via CoA analog intermediates. Finally, gentisate ring is cleaved and the product is ultimately hydrolyzed to fumarate and pyruvate. Chemical structures used in this figure were created in ChemSpace.

is hydrolyzed. In *Streptomyces*, this reaction has been designated as a possible spontaneous reaction whereas in *Agrobacterium*, the reaction is hypothesized to be catalyzed by the enzyme 4-hydroxybenzoyl-CoA thioesterase (EW94DRAFT_05421). This protein was significantly upregulated (fold change > 3) in salicin cultures. Furthermore, the enzyme belongs to protein family – 4HBT which has been demonstrated to be involved in thioester cleavage activity (see attachment S3.3) [135]. Zhuang and colleagues [136] were able to characterize the enzyme gentisyl-CoA thioesterase that catalyzes this reaction in the strain *Bacillus halodurans* C-125. Its amino acid sequence was aligned with that of EW94DRAFT_05421 (see supplementary text), and relatively high percent identity of 32% (at 79% query coverage) was observed. Therefore, EW94DRAFT_05421 is most likely a gentisyl-CoA thioesterase. Other differences between proposed *Agrobacterium* and *Streptomyces* pathways also exist and are described in the discussion section.

Agrobacterium rhizogenes OK036 cannot grow on intermediates of the proposed pathway

Agrobacterium was subcultured on the minimal media containing intermediates of the hypothesized salicin pathway as sole carbon sources. The substrates used were – salicyl alcohol, salicylaldehyde, salicylate, gentisate and catechol (negative control). For catechol, since it is easily oxidized causing optical density changes, a standard curve of only catechol minimal media was created first and then for each time point, the difference OK036-OD_{600} – Catechol-OD_{600} was calculated (see supplementary text for details). *Agrobacterium* apparently does not utilize any of the substrates (figure 3.3). This result indicates that the proposed pathway is not present or more likely that the OK036 strain does not possess transporters to import these substrates for utilization.

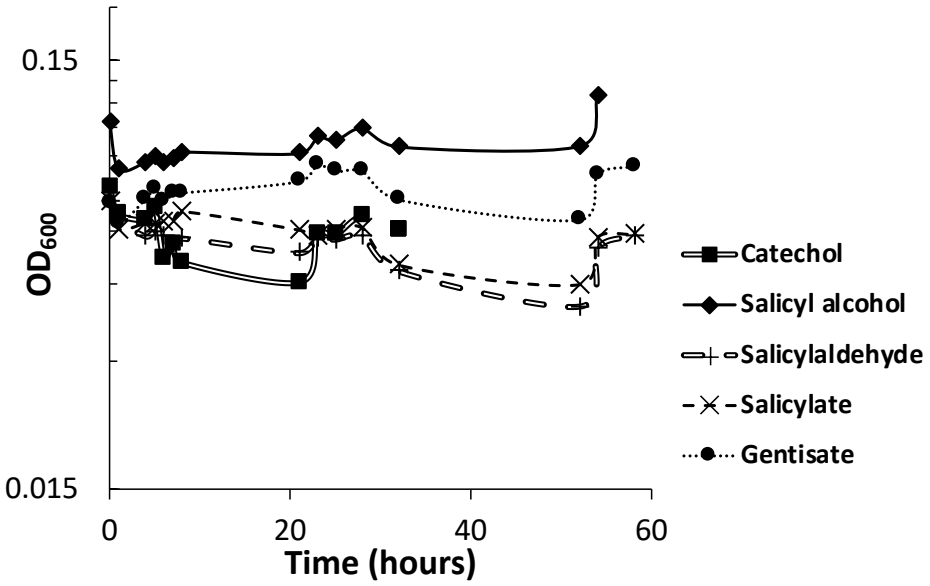


Figure 3.3: *Agrobacterium rhizogenes* OK036 demonstrated no growth on the intermediates of the proposed salicin pathway.
Agrobacterium was subcultured on MOPS minimal media containing desired substrate (3 mM). Catechol was used as negative control. Based on the time-course OD_{600} reading, *Agrobacterium* does not appear to grow on any of the intermediates which is likely due to the absence of the transporters of these compounds. See text for more details.

Metabolomics assay corroborates the presence of the proposed pathway

Metabolomics assay was also performed on the cell-pellets of *Agrobacterium* salicin and glucose cultures to identify the pathway of salicin degradation. For the assay, cells were grown to early log phase followed by lysis using the protocol described already. The cell lysates were then subjected to LC-MS. The intermediates pyruvate, fumarate, salicylaldehyde, salicylate and glucose were all identified in all three replicates of salicin cultures (figure 3.4a). Salicylaldehyde was also identified in glucose cultures.

Then, cell-free reaction assay was performed followed by LC-MS analysis. For the cell-free reaction, cell-lysate samples (both of glucose and salicin cultures) were spiked with either salicin or gentisate (10 mM concentration) and incubated for two different time duration – 1 hour and 24 hours. Metabolite analysis of the spiked samples suggested presence of various

intermediates of the proposed pathway, but glucose culture extracts were also enriched in these intermediates. Intermediates of the pathway such as pyruvate, fumarate, salicylaldehyde, salicylate, glucose were detected in all of the samples (figure 3.4b). The amounts detected were variable for each spiked sample (see attachment S3.4). Salicyl alcohol was detected in the salicin spiked lysates of glucose and salicin cultures. Gentisate was tentatively detected (refers to peak heights that showed enough pattern across the triplicates to be assigned as detected, but not conclusive) in salicin spiked sample of glucose cultures. Finally, fumarylpyruvate was detected in gentisate spiked samples. None of the CoA analogs were detected. Overall, the data indicates the presence of hypothesized salicin pathway in *Agrobacterium rhizogenes* OK036. However, additional work is required to confirm the proposed pathway.

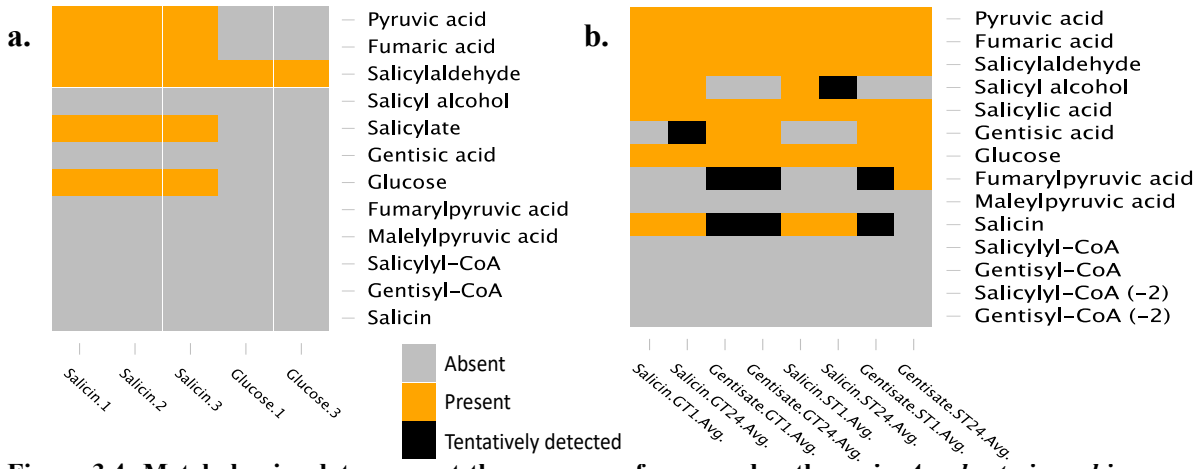


Figure 3.4: Metabolomics data support the presence of proposed pathway in *Agrobacterium rhizogenes* OK036.

Mid-log cultures (salicin and glucose substrate) in triplicates were pelleted, washed and lysed before analyzing using LC-MS. a) Direct analysis of each replicate is shown (only two replicates for glucose). Pyruvate, fumarate, salicylate and glucose were detected in salicin cultures only. Salicylaldehyde was detected in both glucose and salicin cultures. b) Cell lysates were spiked with salicin or gentisate followed by incubation for either 1 hour or 24 hours. Each column represents average of triplicate of glucose or salicin samples. Results were similar to previous experiment. Salicyl alcohol was detected in both salicin and glucose samples spiked with salicin only. Gentisate was tentatively identified in one of the salicin spiked samples (black box). Fumarylpyruvate was identified in gentisate spiked salicin lysate samples (24 hours incubation), and also tentatively detected in other gentisate spiked samples. None of the CoA analogs were detected.

Identification of the prevalence of the proposed OK036 salicin degradation pathway in other *Populus*-associated strains

The amino acid sequence of S5HCoA was searched against the 407 available *Populus*-associated microbial genomes. Using the thresholds – e-value: 1E-05, percent amino acid identity ≥ 30% and alignment length ≥50% in blastp, 252 hits were found representing 153 strains including *Agrobacterium rhizogenes* OK036. After this, TIGRFAM domains were searched for all the gene hits using the IMG database. The OK036 protein contains one TIGRFAM domain. 114 strains belonging to genera *Rhizobium*, *Agrobacterium, Massilia, Duganella, Burkholderia, Massilia, Paraburkholderia, Streptomyces, Tardiphaga and Variovorax* were found to contain S5HCoA homolog with the same TIGRFAM domain (attachment S3.5). These strains were considered for further analysis.

The neighboring genes of the S5HCoA homolog from 15 representative genomes were then manually searched on IMG database (figure 3.5). The gene cluster structures were strikingly similar between *Agrobacterium* and *Rhizobium* representative strains. Similarly, *Bradyrhizobium*, *Bosea* and *Tardiphaga* strains had similar genetic architecture. This analysis indicates that the strains from these five genera might be able to utilize salicin or intermediates of hypothesized salicin pathway by using the mechanism proposed in this study.

Discussion

Plant-microbe metabolic interactions have been identified as an important process in microbial associations with plant. Studies have shown that plants can release 40% or more of their fixed carbon through the process of rhizodeposition depending on the plant species, developmental stage and environment [137]. These compounds stimulate growth and interactions with the soil

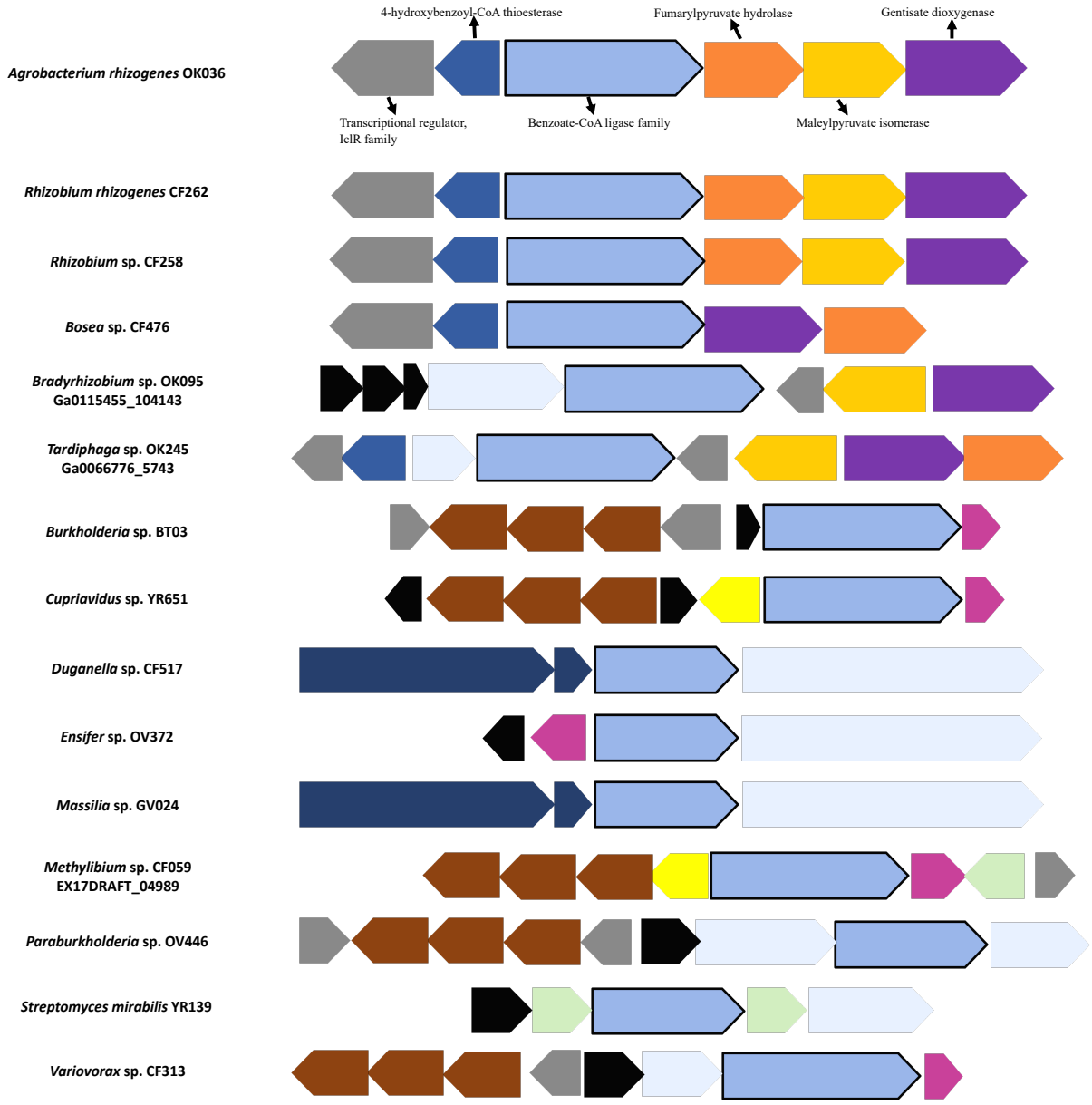


Figure 3.5: Gene cluster identification in other *Populus*-associated strains.
Gene clusters of representative strains of 15 diverse genera were manually searched to identify the neighboring genes of the S5HCoA homolog. Out of 15 strains, three possessed strikingly same gene cluster whereas *Bosea*, *Bradyrhizobium* and *Tardiphaga* had a similar cluster structure. More than one gene was identified as S5HCoA homolog in some organisms such as *Methylibium* sp. CF059. Only one is shown in this picture. Color code: black- hypothetical protein, brown- genes of *box* pathway, light yellow- shikimate kinase, navy blue- related to signal response, light blue- monooxygenase/dioxygenase/dehydrogenase/hydroxylase, lavender- esterase, light green- synthetase.

microbiome [40]. In a study involving $^{13}CO_2$ pulse labeling field experiment, the carbon from plants was observed to be assimilated into the microbiome associated with the grassland vegetation [108, 138]. Likewise, Badri and colleagues [139] have observed that root exudates specifically phenolic compounds could influence the composition of the rhizosphere microbial community.

Populus synthesizes a variety of chemicals and one of the most significant groups of compounds is the phenolic glycosides (PGs) which is produced at concentrations of up to 30% of foliar dry weight in certain *Populus* clones [44]. Because perennial plants (such as *Populus*) have been reported to contribute significant amounts of fixed carbon to their roots, *Populus* likely secretes large amounts of PGs such as salicortin and salicin through its roots [7, 55]. Thus, to understand the microbial colonization of *Populus,* it is important to understand the PG/salicin degradation mechanisms deployed by *Populus*-associated microbial strains such as *Agrobacterium rhizogenes* OK036.

Since members of *Agrobacterium* are known to be able to inject their genetic content into the dicotyledonous plants, these strains are rich in hydrolytic enzymes that can degrade plant cell wall materials such as glucans, xylan, arabinose, polygalacturonases and cellobiose [140, 141]. These enzymes are promiscuous in nature and are demonstrated to catalyze reactions with multiple substrates. For instance, Cbg1 (GH3) protein isolated from *Agrobacterium tumefaciens* has demonstrated moderate activity on salicin but high activity on coniferin and arbutin [75, 88]. Similarly, two bacterial GH1 enzymes were isolated in a metagenomic screening study involving winter wheat soil and both enzymes demonstrated activity against salicin, arbutin and esculin [142]. Therefore, *Agrobacterium rhizogenes* OK036 might have an arsenal of glycosyl hydrolases that might be involved in salicin hydrolysis.

In our preliminary growth analysis, strain OK036 was observed to be able to grow on

salicin minimal media to one of the highest observed ODs (see figure S3.1) for salicin degraders making it a suitable candidate to study a unique salicin utilization mechanism. Most of the salicin degraders consume only the glucosyl part of salicin and secrete salicyl alcohol [2, 129, 130]. In the case of *Agrobacterium*, it was likely consuming both functional groups for its growth which was supported by the absence of salicyl alcohol in the supernatant of salicin cultures (figure 3.1b).

For identifying the salicin utilization pathway in *A. rhizogenes* OK036, a proteomics study was employed. The comparison of proteome data between salicin and glucose cultures suggested significant upregulation of proteins encoded by a gene cluster EW94DRAFT_05420 - EW94DRAFT_05430 as seven of the proteins encoded from this cluster were in the top ten significantly abundant proteins in terms of fold change (p-value < 0.05). Neither GH1 nor GH3 enzymes were significantly upregulated in either cultures suggesting these are not induced by salicin or glucose. The gene cluster can be divided into two subclusters: SC1 – potentially involved in converting salicyl alcohol to salicylate and SC2 – required for salicylate transformation. This gene cluster was further investigated.

The subcluster SC2 has been identified in *Streptomyces* sp. WA46 strain for salicylate utilization. In the *Streptomyces* pathway, salicylate is ultimately converted to fumarate and pyruvate through gentisate intermediate [134]. Salicylate is first converted to salicylyl-CoA through the action of salicylate-CoA ligase (AMP forming). This protein has been identified in strains of *Pseudomonas*, *Mycobacterium* and *Rhodococcus* [143-145]. Ishiyama and colleagues [134] were able to demonstrate that the protein (encoded by *sdgA*) converts salicylate to salicylyl-CoA using cell-free extract assays. Similar enzymes were first characterized in *Azoarcus evansii* and coined as benzoate-CoA ligase (AMP forming) and 3-hydroxybenzoate-CoA ligase (AMP forming) as they could convert benzoate and 3-hydroxybenzoate to benzoyl-CoA and 3-

hydroxybenzoyl-CoA, respectively [146]. We determined that *sdgA* has homology to the C-terminal portion of EW94DRAFT_05422 gene of *Agrobacterium* at 24% amino acid identity.

The second step in the proposed pathway in *Streptomyces* is hydroxylation of salicylyl-CoA to gentisyl-CoA by the action of salicylyl-CoA 5-hydroxylase [134, 145, 147]. Surprisingly, the best blastp hit of this protein-encoding gene (*sdg*C) matched to the N-terminal portion of EW94DRAFT_05422 gene at 38% identity. The *pfam* domains of *sdgA* – *sdgC* and EW94DRAFT_05422 were same indicating that EW94DRAFT_05422 has a dual function such that it is both a CoA ligase and a hydroxylase. This is a distinction between *Streptomyces* and *Agrobacterium* pathways because one single polypeptide can convert salicylate to gentisyl-CoA in *Agrobacterium* whereas in *Streptomyces,* two enzymes are required.

The second difference between the *Streptomyces* pathway and our proposed pathway involves the hydrolysis of thioester group of gentisyl-CoA which in *Agrobacterium* is not a spontaneous but an enzyme-catalyzed reaction. Using the genomic, proteomic and literature data, we conclude that the gene EW94DRAFT_05421 most likely encodes for a gentisyl-CoA thioesterase.

In subsequent steps of the proposed pathway, gentisate is converted to pyruvate analogs. This gentisate pathway has been previously identified in strains of *Rhodococcus, Moraxella, Corynebacterium* and *Ralstonia* [147-150]. In the subcluster SC2, two genes are annotated as maleylpyruvate isomerase which catalyzes the conversion of maleylpyruvate to fumarylpyruvate. Both proteins were upregulated significantly in salicin cultures of *Agrobacterium rhizogenes* OK036. Shen and colleagues [149] have reported that two different types of maleylpyruvate isomerases exist depending on their requirement for glutathione (GSH) as cofactor. *A. rhizogenes* appears to contain both type of enzymes as suggested by PFAM domains present in each protein

– EW94DRAFT_05424 (GST_N_3 and GST_C_3) and EW94DRAFT_05426 (MDMPI_N). This appears to be another unique feature in *Agrobacterium*. See attachment S3.3 for further details.

We also propose that the other half of the gene cluster comprising of EW94DRAFT_05427 - EW94DRAFT_05430 (subcluster SC1) is required for the conversion of salicyl alcohol to salicylate. The gene acyl-CoA reductase (EW94DRAFT_05428) belongs to aldehyde dehydrogenase protein family suggesting that it might be involved in the reduction of salicylaldehyde to salicylate. EW94DRAFT_05428 was 61% identical to the gene encoding salicylaldehyde dehydrogenase of *Pseudomonas putida* when both amino acid sequences were aligned [150]. Oxidation of salicyl alcohol to salicylaldehyde probably is catalyzed by either choline dehydrogenase or reduced nicotinamide adenine dinucleotide phosphate (NADPH):quinone reductase (see attachment S3.3). Salicin hydrolysis to salicyl alcohol and glucose is carried out by either GH1 or GH3 enzyme which do not appear to be induced by salicin.

Metabolomics data corroborate the presence of proposed pathway in *Agrobacterium rhizogenes* OK036. However, this pathway appears to be partially active even in the glucose cultures as several intermediates of the proposed pathway were detected in the cell-free reactions of glucose culture lysates spiked with salicin and gentisate. Proteome data supported this observation as the proteins of the salicin pathway were also detected in glucose cultures. Furthermore, it appears that *Agrobacterium* does not contain transporters of several intermediates of the pathway based on the lack of growth on minimal media containing the intermediates of proposed pathway as sole carbon sources. Overall, the proposed pathway is likely present in *Agrobacterium rhizogenes* OK036.

Finally, the protein sequence of *A. rhizogenes* OK036 salicylate 5-hydroxylase (CoA forming)/S5HCoA was searched against 407 available *Populus*-associated microbial genomes

(which also includes *A. rhizogenes* OK036). Our analysis resulted in 140 gene hits representing 114 genomes. Fifteen representative strains were manually searched in the IMG database for determination of the neighboring genes of S5HCoA homolog. Out of 15 strains, three including *Agrobacterium rhizogenes* OK036 contained the same gene cluster structure, and three including *Tardiphaga* contained a similar cluster. These strains most likely can grow on salicin or intermediates of the proposed pathway using the mechanism described in this study. Overall, the data suggests that the hypothesized pathway might be functional in strains other than those belonging to the genus *Agrobacterium.*

In conclusion, we have identified a novel mechanism of salicin degradation via gentisate in *Agrobacterium rhizogenes* OK036. The proteins of the proposed pathway appear to belong to an eleven-member gene cluster. Within this gene cluster, first half (SC2) is involved in salicylate transformation whereas second half (SC1) appears to be required for converting salicyl alcohol to salicylate. S5HCoA protein, encoded by EW94DRAFT_05422 gene (of subcluster SC2) is proposed to be a dual-function protein consisting of both CoA ligase and monooxygenase activities and likely converts salicylate to gentisyl-CoA. This is the first study to have identified the novel salicin utilization pathway and will provide new avenues in research on salicin degradation in the future.

Chapter 4

Mechanism of unidirectional by-product cross-feeding of salicyl alcohol between *Populus* microbiome isolates

Introduction

In natural context, microorganisms interact extensively either through competitive or cooperative means which shape the microbial community of a specific niche [151-154]. In recent times, studies on the microbial metabolic interactions that drive the mechanistic processes defining the microbiome features have been given much attention. Within this area, cross-feeding metabolic interactions involving diverse bacterial species have been demonstrated in several studies [155-157]. D'Souza et al. [158] have classified cross-feeding interactions using two parameters – degree of reciprocity between the partners and cost to each microbe involved in the interaction. Based on these two factors, metabolic exchanges between microbes can be subdivided into five categories –

Unidirectional by-product cross-feeding: In this type of interaction, the by-product released by one organism is taken up and utilized by another. For instance, in a co-culture of *Bifidobacterium adolescentis* L2-32 and *Eubacterium hallii* on starch, it was found that *B. adolescentis* provided *E. hallii* with lactate so that the latter could grow in the co-culture [159].

Bidirectional by-product cross-feeding: Both organisms in the interaction benefit from each other by the release and subsequent utilization of by-products. LaSarre and colleagues [160] demonstrated that *E. coli* and a mutant strain of *Rhodopseudomonas palustris* cross-feed organic acid and ammonium to each other respectively so that both can cooperatively grow in glucose culture.

By-product reciprocity: This is a special type of cross-feeding interaction in which one member produces costly metabolite that benefits its partner but only receives metabolic by-product in return. For instance, two termite hindgut bacteria – *Treponema azotonutricium* and *T. primitia* are proposed to exchange folate and hydrogen between them in co-culture [161].

Bidirectional cooperative cross-feeding: In this cross-feeding interaction, both members cross-

feed each other costly metabolites. For instance, in a study involving two mutant strains of *Saccharomyces cerevisiae* – one leucine auxotroph-tryptophan overproducer and the other leucine overproducer-tryptophan auxotroph, it was demonstrated that a stable mutualistic relationship could be achieved between them because of the cross-feeding of the amino acids [162].

Unidirectional cooperative cross-feeding: There is a theoretical possibility that one organism provides costly metabolite to another but does not receive anything in return. This type of interaction is extremely unstable and most likely leads to other types of interactions discussed before.

Metabolic interactions between microbial partners are constrained by the niche they inhabit [154]. In fact, studies have implied that hosts such as plants play an active role in the colonization of endophytes by attracting specific bacteria through the release of certain compounds from their roots [107, 163]. *Populus*, a woody perennial native to North America, produces numerous compounds specifically phenolic glycosides (PGs), which have been reported at concentrations of up to 30% of foliar biomass in certain *Populus* clones [44]. PGs consist of a salicin sub-structure attached to various functional moieties such as benzyl-, acetyl- and/or hydrocyclohexene group by an esterification process [7]. Even though their role as herbivore-deterrent is well-studied, studies demonstrating their function in shaping plant microbiome is scarce [56, 63, 69, 164]. Understanding how microbiome is influenced by PGs is an important question especially in *Populus* and for that, we need to study how PGs such as salicin are degraded by *Populus* isolates.

A few studies have demonstrated microbial utilization of salicin, and it was first described in *Escherichia coli* whose wild-type strains do not grow on salicin, but spontaneous mutants arise in the salicin culture that can consume salicin for growth [81]. The *bgl* operon of *E. coli* responsible for salicin utilization has been studied in other organisms including *Erwinia chrysanthemi* and

Klebsiella aerogenes [2, 71]. The *bglGFB* gene cluster (Bgl system) encodes proteins BglG, a positive regulator, BglF, phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS) transporter and BglB, a β-phosphoglucosidase enzyme [1, 70].

Salicin metabolism has been reported to result in the release of salicyl alcohol [2, 78]. Utilization of benzyl alcohol (a compound related to salicyl alcohol) through naphthalene degradation pathway has been studied in *Pseudomonas putida* CSV86, and the enzymes – benzyl alcohol dehydrogenase and benzaldehyde dehydrogenase required for the transformation of benzyl alcohol to benzoate have been characterized [165]. Interestingly, these enzymes have also demonstrated activity against salicyl alcohol and salicylaldehyde, respectively [165, 166]. In the proposed pathway in *P. putida*, benzyl alcohol is first transformed to catechol which is then directed to the ortho-cleavage pathway of β-ketoadipate channel [166-168].

In the present investigation, we elucidate the molecular mechanism of salicyl alcohol cross-feeding between two *Populus* microbiome isolates, *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16, when inoculated on salicin together. By utilizing bioinformatics and quantitative approaches such as GC-MS (metabolomics) and LC-MS (shotgun proteomics), we demonstrate that *Rahnella* utilizes Bgl system to grow on salicin and releases salicyl alcohol. Salicyl alcohol then can be utilized by other member of the co-culture – *Pseudomonas* sp. GM16 using the pathway proposed for the degradation of benzyl alcohol in *P. putida* CSV80 strain. The presence of *Pseudomonas* sp. GM16 does not appear to benefit *Rahnella* but might inhibit its growth. Therefore, the results from this study suggest the presence of a unidirectional cross-feeding interaction between the two organisms. The current research provides insights into potential multi-trophic interactions between *Populus* and its microbiota directly and indirectly through PGs such

as salicin. This relationship is possibly important for the establishment and stability of the *Populus* microbiome.

Materials and methods

Bacterial strains and culture medium

The γ-proteobacteria *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16 were previously isolated from *Populus* root tissues. The genomes were sequenced and assembled as described elsewhere [91]. The genomes are available in both National Center for Biotechnology Information (NCBI) (*Rahnella*: GCA_000799835.1; *Pseudomonas*: GCA_000282155.1) and Integrated Microbial Genomes (IMG) (*Rahnella*: 2585427591; *Pseudomonas*: 2511231031) databases. The bacterial strains were maintained on R2A (Franklin Lakes, BD Difco, NJ, USA) agar plates.

For growth assays, individual colonies were picked from the agar plates and grown in liquid R2A medium overnight in the incubator at 25°C and 250 rpm. The samples were subcultured in 3-(N-morpholino)propanesulfonic acid (MOPS) minimal medium (pH: 7.2) [94] containing the desired carbon source (salicin, salicyl alcohol and glucose) at concentrations of either 3 mM or 5 mM. Growth was measured using optical density (OD) at wavelength 600 nm on Spectronic 20D+ (Thermo Fisher Scientific). The starting OD_{600} of subcultures were between 0.07 - 0.1.

Primer design, DNA isolation and microbial quantification

Cell pellets harvested from bacterial cultures (1:1 starting OD_{600} ratio for co-culture) at different time points during growth were lysed and total DNA was extracted using QIAgen DNeasy Blood and Tissue kit (Qiagen). Total DNA was stored at -20°C for future use. Quantitative PCR (qPCR) was used for detection and quantification of microbes. qPCR was performed using

iTaq Universal SYBR Green One-Step Kit (Bio-Rad Laboratories). The primers were designed to be strain-specific for *Rahnella* sp. OV744 (forward: TGCCTTTCACGCCGATTAT and reverse: TACGGTTACGCACGGTTTC) and *Pseudomonas* sp. GM16 (forward: TGCTTCGGCAGCGATTTA and reverse: CACGGAGATGAAGTTCTCGATAG). The qPCR reaction was run on a CFX96 machine (Bio-Rad Laboratories) and was monitored using Bio-Rad CFX Manager v3.1. Following the run, the qPCR data was imported and analyzed in Microsoft Excel . See supplementary text for calculation protocol.

For determination of stable state of the co-culture, qPCR was performed on samples over several generations. *Rahnella* and *Pseudomonas* cells were co-cultured in salicin minimal medium. Then, the overnight samples were diluted in the ratio of 1:10 into fresh salicin media for another growth cycle. This dilution process was repeated for several days and the culture samples were extracted between every dilution. Following the isolation of cell pellets and lysis, qPCR was performed for every sample and the data was analyzed as described previously.

Gas chromatography - mass spectrometry (GC-MS) metabolome analysis

Secreted metabolites from *Rahnella* and *Pseudomonas* cultured on salicin (3 mM) minimal medium in mono- and co-cultures were determined by GC-MS method. 1 ml of culture supernatants were isolated at different time points during growth and stored at -20°C. Frozen supernatant samples were thawed and vortexed. 100 µl of sample was transferred to a scintillation vial and 15 µl of sorbitol (1 mg/mL) was added as an internal standard. After drying under a nitrogen stream, the samples were dissolved in 0.5 mL acetonitrile (ACN), and silylated to generate trimethylsilyl (6TMS) derivatives, as described elsewhere [169]. After 2 days, 1 µl aliquots were injected into an Agilent 5975C inert XL gas chromatograph-mass spectrometer (GC-MS). The

standard quadrupole GC-MS was operated in electron impact (70 eV) ionization mode, targeting 2.5 full-spectrum (50-650 Da) scans per second, as described previously [169].

Metabolite peaks were extracted using a key selected ion, characteristic m/z fragment to minimize integration of co-eluting metabolites. The extracted peaks of known metabolites were scaled back to the total ion current (TIC) using previously calculated scaling factors. Peaks were quantified by area integration and normalized to the quantity of internal standard recovered, sample volume analyzed, derivatized and injected. A large user-created database and the Wiley Registry 10th Edition/ National Institute of Standards and Technology (NIST) 2014 Mass Spectral Library was used to identify and quantify the metabolites of interest. Unidentified metabolites were represented by their retention time and key m/z ratios.

Sample preparation for proteome measurements

Bacterial strains were grown in triplicates in three different culture conditions in MOPS minimal medium – *Rahnella* sp. OV744 in pure culture on glucose or salicin (5 mM), *Pseudomonas* sp. GM16 in pure culture on glucose or salicyl alcohol (3 mM), and *Rahnella-Pseudomonas* in co-culture on salicin (3 mM). After the cultures reached an OD_{600} of 0.7 or higher, cell pellets were harvested by centrifugation at 11952 x g (15 mins, 4°C) and flash frozen in liquid N_2 and stored at -80°C. Triplicate samples were processed separately for shotgun proteomics analysis [170]. Cell pellets were thawed and then suspended in lysis buffer (5% sodium dodecyl sulfate (SDS), 50 mM Tris-HCl, 0.15M NaCl, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM $MgCl_2$, pH 8.5) and lysed by immersing sample tubes in a boiling water bath for 20 min. Chilled (-4°C) trichloroacetic acid (Fisher Scientific, 99.8%) was added to a final concentration of 25%, and protein was precipitated by storage overnight at -20°C. Samples were then thawed, and

the supernatant discarded. The remaining pellets were washed twice, by vortexing briefly with chilled (-80°C) acetone, centrifuging at 21000 x g for 20 min, and discarding the supernatant. The pellets were air dried and then dissolved in 50 mM Tris, 10 mM $CaCl_2$ at pH 7.8. For protein denaturation, samples were incubated for 3 hours at 60°C in 6M guanidine, 10 mM dithiothreitol (DTT). Cell lysates were then cooled to room temperature and diluted 6-fold in 50 mM Tris, 10 mM $CaCl_2$ at pH 7.8. Trypsin (proteomics grade, Sigma) was added to each lysate at a rate of 20 μg/mg protein and incubated at 37°C overnight. Then, an identical amount of trypsin was added, followed by incubation at 37°C for an additional 4 hours. The resulting digests were adjusted to 0.1% formic acid (EMD Millipore, Suprapur 98-100%). The removal of particulates and remaining cell debris was carried out by centrifugation through 0.45 μm pore filters (Ultrafree-MC, Millipore). Samples were frozen at −80°◦C until LC-MS/MS analysis.

Liquid chromatography/tandem mass spectrometry (LC MS/MS) analysis

Two-dimensional LC-MS/MS using the multidimensional protein identification technology (MudPIT) approach [171, 172] was utilized to analyze the tryptic peptide mixtures. The protocol is described in detail elsewhere [173]. For each growth condition, three biological replicates were analyzed. Using a pressure cell (New Objective), digests corresponding to digestion of ~100 μg protein were loaded onto a precolumn fabricated from 150 μm internal diameter (ID) fused silica tubing (Polymicro Technologies). The precolumn was packed with a ~4 cm long bed of reverse-phase chromatographic phase (Jupiter C18, 3 μm particle size, Phenomenex) upstream of a ~ 4cm bed of strong cation exchange material (5 μm particle size SCX, Phenomenex).

After loading, digests were desalted by attaching the precolumn to an HPLC system (see

below) and subjecting to several short reverse-phase gradients to wash away salts, and to elute the peptides from the reverse-phase bed to the SCX bed. The precolumn was then attached to an analytical column through a filter union (Upchurch Scientific). Analytical columns were fabricated from 100 μm ID PicoTip Emitters (New Objective) packed with a ~14 cm bed of reverse-phase material (Jupiter C18, 3 μm particle size, Phenomenex). The pre- and analytical column assembly was attached to an Accela HPLC system (Thermo Fisher Scientific), and two-dimensional LC was affected through application of 11 short step gradients of increasing salt (ammonium acetate) concentration, each followed by a separate reverse-phase gradient. The eluent was interfaced by a nanospray source (Proxeon) with a linear-geometry quadrupole ion trap mass spectrometer (LTQ-XL, Thermo Fisher Scientific), operating under control of XCalibur software for data acquisition in data-dependent mode. Tandem mass spectra were acquired via collision-induced dissociation (CID) from up to 5 of the most intense ions in each full-scan mass spectrum. Precursor isolation width for CID was 3 m/z units, with normalized collision energy set at 35%. Dynamic exclusion was invoked with repeat count of 1, exclusion list size set to 300, repeat duration of 60 seconds, and exclusion duration of 180 seconds.

Proteomics data analysis

Myrimatch (v. 2.1.138) [174] was used to obtain peptide identifications from tandem mass spectra. IDPicker [175] was used to compile the peptide identifications to determine protein identifications. Protein FASTA files for *Rahnella* (downloaded February 16, 2017 as 2585427591.genes.faa, containing 4962 proteins) and *Pseudomonas* (downloaded June 29, 2016 as 2511231031.genes.faa, containing 5888 proteins) were obtained from the DOE (Department of Energy) Joint Genome Institute Integrated Microbial Genomes (JGI/IMG) database [100], and

concatenated with sequences of 44 common contaminant proteins to provide a database for Myrimatch searches. Myrimatch estimated false discovery rates for peptide identification using a sequence-reversed analog of each protein [176, 177]. Peptide identification criteria set for Myrimatch required precursor m/z tolerance of 1.5, fragment m/z tolerance of 0.5, charge states up to +4, TIC cutoff of 98%, cleavage rule "Trypsin/P", fully tryptic peptides only, with a maximum of two missed trypsin cleavage sites per peptide. A protein was assumed to be detected if the number of tryptic peptides identified was ≥ 2 for at least one LC-MS/MS analysis. The false discovery rate for peptide-spectrum matches was set in Myrimatch at $\leq 2\%$, and the resulting peptide and protein false discovery rates were $\leq 0.8\%$ and $\leq 2.5\%$, respectively.

Further analyses of proteome data were performed using custom scripts in R[178]. Proteins that shared all their identified peptides were combined into protein groups. Protein abundances were estimated using the spectrum count (number of tandem mass spectra assigned to a protein)[179], corrected for shared peptides [180]. To compare protein abundances across experimental treatments, normalized spectral abundance factors (NSAF) were calculated. NSAFs were calculated in two ways. "Combined" NSAFs summed to 1 for all proteins from both *Pseudomonas* sp. GM16 and *Rahnella* sp. OV744, while "by-organism" NSAFs summed to 1 for GM16, and also to 1 for OV744. Two approaches were used to identify differentially abundant proteins in pairwise comparisons, based on NSAF. T-tests were performed, assuming unequal variances between two treatments. Benjamini-Hochberg (BH) correction was applied to t-test p-values. The fold change for a protein between the two conditions was calculated by determining the ratios of the NSAF values (by-organism), and then log transforming using base 2.

Recombinant plasmid construction for complementation assay

First of all, primers were designed to PCR amplify the gene encoding the *Rahnella* sp. OV744 phosphoglucosidase using OneTaq Quick-Load 2X Master Mix kit (New England Biolabs (NEB) Inc.). The primer sequences are – forward: CGCGAGCTCATGAACAAACAATTTCCA and reverse: GGCGGTACCTCAGGACGTTAATGAC. These sequences contain SacI and KpnI (New England Biolabs (NEB) Inc.) restriction sites at their 5' end right after 3 bp overhangs. The vector plasmid (pSRK.km) [181] was maintained in *E. coli* (strain TOP10) on Luria-Bertani (LB) [182] plates containing kanamycin antibiotic (50 µg/ml concentration). The plasmid was extracted using QIAquick Spin Miniprep Kit (Qiagen). Concentrations of both the vector plasmid and the insert (PCR product) were measured using nanodrop (Thermo Fisher Scientific). Then both were restriction digested using SacI and KpnI (New England Biolabs (NEB) Inc.) by incubating at 37°C for two hours. Following the digestion, the samples were run on 1% agar gel containing ethidium bromide. For controls, single digested plasmids and undigested plasmids were also added to the gel run. After this, appropriate gel bands (containing the double digested plasmid and insert) were cut and gel extracted using QIAquick gel extraction kit (Qiagen). Ligation of insert to vector was performed using DNA Ligase Kit (New England Biolabs (NEB) Inc.) and by incubating at room temperature for 10-30 minutes. As negative control for ligation, only digested plasmid (no insert) was used in a separate reaction mix. Following the incubation, 3 µl of ligase reaction was used to transform 40 µl NEB 10β competent cells (*E. coli* DH5α). First the cell samples (along with ligase reaction) were mixed and incubated on ice for 30 minutes. This was followed by 30 second heat shock in 42°C in hot water bath and subsequent five-minute incubation on ice. Then, 950 µl LB liquid media was added to the cell mix and incubated at 37°C with gentle shaking (200 rpm) in the incubator for 1 hour. Finally, 100 µl of the culture was directly plated onto LB plates containing

kanamycin (50 µg/ml concentration) and incubated at 37°C overnight. Next day, selected individual colonies were picked and grown in LB liquid media containing kanamycin overnight in the shaking incubator (200 rpm) at 37°C. Then plasmids were extracted from the cultures using QIAquick Spin Miniprep Kit (Qiagen). Plasmids were then restriction digested with SacI and KpnI as described before and run on gel and visualized on Benchtop 3UV Transilluminator connected to MultiDoc-It Digital Imaging System (UVP). One of the colonies containing the recombinant plasmid with the required insert was then chosen for sequencing for verification purpose. This plasmid has been designated as pSRK.km_OVCompYR.

Preparation of electrocompetent cells for complementation assay

Pantoea sp. YR343 phosphoglucosidase mutant (ΔGH1) was a gift from Dr. Jennifer Morrell-Falvey's lab in Oak Ridge National Laboratory, Oak Ridge, TN. This mutant strain was maintained on a R2A agar plate. The electrocompetent cells of ΔGH1 were prepared following a standard protocol. Briefly, an overnight R2A culture of ΔGH1 was added to 50 ml R2A fresh liquid media and grown to OD_{600} close to 0.4 by incubating at 25°C and 200 rpm. Following this, cells were recovered by centrifugation at 11952 x g (10 minutes, 4°C). Supernatant was decanted and pellets were resuspended in 50 ml chilled sterile distilled water. The cultures were pelleted and washed again using chilled distilled water. After this, the pellets were resuspended in 10 ml of chilled sterile 10% glycerol and extracted by centrifugation followed by decantation. The pellets were then finally suspended in 1 ml of chilled 10% glycerol and aliquoted in chilled microfuge tubes. Aliquots were flash frozen in liquid N_2 and stored at -80°C until required.

Complementation assay

The complementation of ΔGH1 with pSRK.km_OVCompYR was performed using electroporation method using Gene Pulser (Bio-Rad Laboratories). For controls, ΔGH1 competent cells were transformed with either pSRK.km plasmid (insert control) or sterile water (negative control). Approximately 10 ng of DNA was added to 40 µl ΔGH1 competent cells. 0.2 cm cuvettes were used and following settings were applied: voltage – 2.5 kV, capacitance – 25 µFD and resistance – 200 Ω. After each pulse, 1 ml of R2A liquid media was added to the cell mix. The cultures were pipetted to 1.5 ml eppendorf tubes which were sealed with parafilm. Cultures were grown at 25°C with shaking at 200 rpm overnight. Following this, cultures were spun down at 2655 x g for 2 minutes, and pellets were resuspended in 100 µl R2A liquid medium. Entire suspension was then spread on the R2A plates containing kanamycin (50 µg/ml concentration), and the plates were incubated for 1-2 days at 25°C until colonies were fully visible. The colonies were then streaked onto R2A plates (with kanamycin) and MOPS plates (containing either salicin (1 mM) or glucose (1 mM) as sole carbon source and kanamycin as the selection antibiotic) and grown overnight at 25°C. Selected colony that grew on all three types of plate conditions was then subcultured on liquid minimal medium containing salicin (3 mM) as sole carbon source. The growth was monitored using optical density (OD_{600}) over thirty-hour period.

Bioinformatic analysis

The tools in JGI/IMG [100], UniProt-based Families (UniFam) [183], and KEGG (Kyoto Encyclopedia of Genes and Genomes) [184] were used for identifying the pathways for substrate metabolism. First, the FASTA files for both strains were extracted from JGI/IMG database [100], and entered into UniFam genome annotation pipeline [183]. Next, all the possible reactions were

determined using the pathway maps in KEGG for all the substrates/intermediates of salicin or salicyl alcohol pathway in *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16 respectively. Then, the enzyme commission (EC) numbers for the enzymes catalyzing each reaction were searched against the results obtained from UniFam annotation. Each of the possible genes corresponding to the EC number was then searched in the JGI/IMG server's "Find Genes" module in order to corroborate the UniFam annotation. If the EC number could not be matched against the genome annotation, the missing enzyme was predicted based on the operon information because most of the genes involved occurred in operons in both strains. InterPro annotations found in the JGI/IMG server were also utilized for further assurance that the genes encoded for the missing enzymes. Finally, the protein sequences of the enzymes in other set of studies were used as queries against the genomes of *Rahnella* and *Pseudomonas* in blastp tool to further confirm the proteins and pathways that are proposed to be involved in the metabolism of salicin and salicyl alcohol, respectively.

Results

Rahnella sp. OV744 grows on salicin by utilizing the glucosyl moiety, and in the process secretes salicyl alcohol

Rahnella was observed to grow in minimal medium containing salicin or glucose but not salicyl alcohol (figure 4.1a). Using the *E. coli* phosphoglucosidase (glycosyl hydrolase family 1 – GH1 which cleaves salicin at its β-glycoside bond) protein sequence, blastp was used to predict GH1 homologs in *Rahnella* (figure S4.1). Out of five predicted genes, gene EX29DRAFT_02290 (EC 3.2.1.86) was proposed to be the best candidate because the genes flanking EX29DRAFT_02290 were annotated as a phosphotransferase system (PTS) based

transporter/permease (EX29DRAFT_02289) and an anti-terminator regulator (EX29DRAFT_02291) (see attachment S4.1). The cluster architecture is similar to the *bgl* operon in *E. coli*. Based on prior knowledge of the system, it was hypothesized that *Rahnella* sp. OV744 imports and hydrolyzes salicin to glucose 6-phosphate and salicyl alcohol and then uses glucose 6-phosphate for energy generation. In the proposed model, *Rahnella* releases salicyl alcohol into the media.

To determine whether the predicted pathway for the conversion of salicin was utilized by *Rahnella* with subsequent release of salicyl alcohol, a time-course metabolomic study was carried out with focus on salicin and salicyl alcohol. *Rahnella* sp. OV744 was grown on salicin minimal media and at different time points, the supernatant was extracted and analyzed using GC-MS. As shown in figure 4.1b, *Rahnella* consumed salicin over time and secreted salicyl alcohol to a final concentration of 190 μg/mL (sorbitol equivalent). This experiment provided clear evidence that *Rahnella* sp. OV744 only utilizes the glucosyl moiety of salicin and secretes salicyl alcohol into the media.

Proteomics and metabolomic investigation of salicin degradation in *Rahnella*

To support the annotation-based prediction that *bgl* operon is present in *Rahnella* for salicin utilization, a proteomics study was performed. In the study, the protein expression profiles between cells grown on salicin or glucose and harvested at mid-exponential phase were compared. The hypothesized proteins, PTS transporter/permease (EX29DRAFT_02289) and 6-phospho-β-glucosidase (EX29DRAFT_02290) were detected in salicin culture but were absent in all of the replicates in glucose cultures (table 4.1). The protein encoded by the anti-terminator gene

(EX29DRAFT_02291) could not be detected in treatment groups. Complete proteome data have been provided in attachment S4.2.

Proteome comparisons were also used to investigate the mechanism of export of salicyl alcohol. Due to the lack of prior knowledge of exporters of salicyl alcohol, blastp (e-value ≤ 1E-05) was used to query the protein sequences of known exporters of related compounds such as toluene against the *Rahnella* sp. OV744 genome [185-188]. The gene hits from blastp were then matched against the proteomics data and several candidate genes/proteins have been identified. Further investigation is required to determine if any of these transporters or other mechanisms are utilized by *Rahnella* for releasing salicyl alcohol (see table S4.1).

The metabolomic data was further investigated for identification of intermediate metabolites of the proposed pathway. PTS-based transporters are known to phosphorylate their substrates which means salicin should be imported as salicin 6-phosphate. Spectral data from GC-MS analysis of both internal and secreted metabolites were analyzed. Salicin 6-phosphate was

Table 4.1: Probable proteins involved in salicin degradation for *Rahnella* sp. OV744 were detected in salicin cultures only.
Except for EX29DRAFT_02291, other proteins encoded by the *Rahnella bgl* operon were detected only in salicin cultures. All the proteins detected in the proteomics experiment are provided in attachment S4.2.

Locus Tag	Annotation	Present in glucose	BH corrected p-value	Fold change
EX29DRAFT_02289	PTS system beta-glucoside-specific IIA component, Glc family (TC 4.A.1.2.2)/PTS system beta-glucoside-specific IIB component, Glc family (TC 4.A.1.2.2)/PTS system beta-glucoside-specific IIC component, Glc family (TC 4.A.1.2.2)	No	0.011	7.570
EX29DRAFT_02290	6-phospho-beta-glucosidase	No	0.002	7.634
EX29DRAFT_02291	transcriptional antiterminator, BglG family	No	NA	NA

detected in the supernatant samples of salicin monocultures. The electron ionization (EI) fragmentation pattern of 6TMS derivative of salicin 6-phosphate was discovered and is presented in figure 4.1c. The identification of this metabolite further corroborates the presence of the proposed pathway in *Rahnella* sp. OV744.

Rahnella **sp. OV744 phosphoglucosidase can functionally replace the phosphoglucosidase function of** ***Pantoea*** **sp. YR343 mutant**

Rahnella phosphoglucosidase gene was PCR amplified and cloned into pSRK.km vector (see figure S4.2). This recombinant plasmid is designated as pSRK.km_OVCompYR and was transformed into *Pantoea* phosphoglucosidase mutant (ΔGH1) using electroporation method. For control, pSRK.km plasmid only and sterile water were also used separately for transformation. *Pantoea* was selected because it is proposed to contain a similar *bgl* operon and the same mechanism of salicin degradation as that of *Rahnella* sp. OV744 (unpublished work) (see figure 4.2a).

Following transformation and subsequent steps, colonies were streaked onto R2A plates (with kanamycin). Individual colonies from R2A plates were picked and grown in MOPS minimal media containing salicin as sole carbon source and kanamycin as the selection antibiotic. As can be seen in figure 4.2b, cultures of mutant containing pSRK.km_OVCompYR (recombinant vector) were able to grow on salicin minimal media to an OD comparable to that of the wild type. On the other hand, cultures of strains containing only pSRK.km vector demonstrated no growth on salicin minimal media. The results indicate that *Rahnella* phosphoglucosidase can complement the *Pantoea* ΔGH1 mutant and further support the proposed pathway of salicin degradation in *Rahnella*.

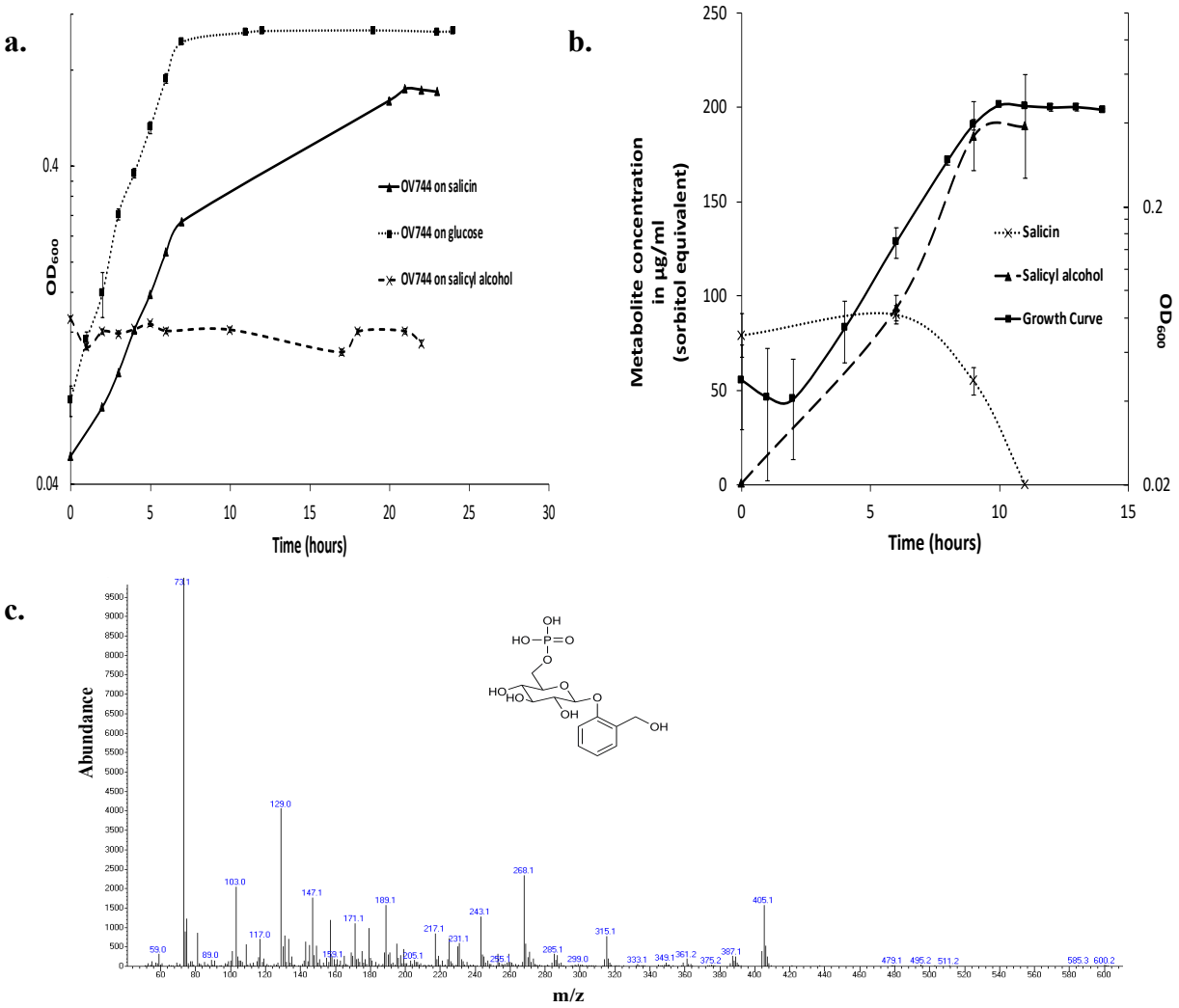


Figure 4.1: *Rahnella* sp. OV744 utilizes salicin and secretes salicyl alcohol in the process.
a) When *Rahnella* was inoculated with salicin, glucose and salicyl alcohol, it was able to grow on salicin and glucose, but not on salicyl alcohol minimal media. b) When a time-course metabolite analysis was performed on supernatant of OV744 salicin culture, the reduction in the levels of salicin was clearly visible along with the rise in salicyl alcohol levels. c) The electron ionization (EI) fragmentation pattern shown here represents the 6TMS derivative of salicin 6-phosphate which was identified from the supernantant metabolome of *Rahnella* salicin cultures. Salicin 6-phosphate is an intermediate of the proposed salicin degradation pathway.

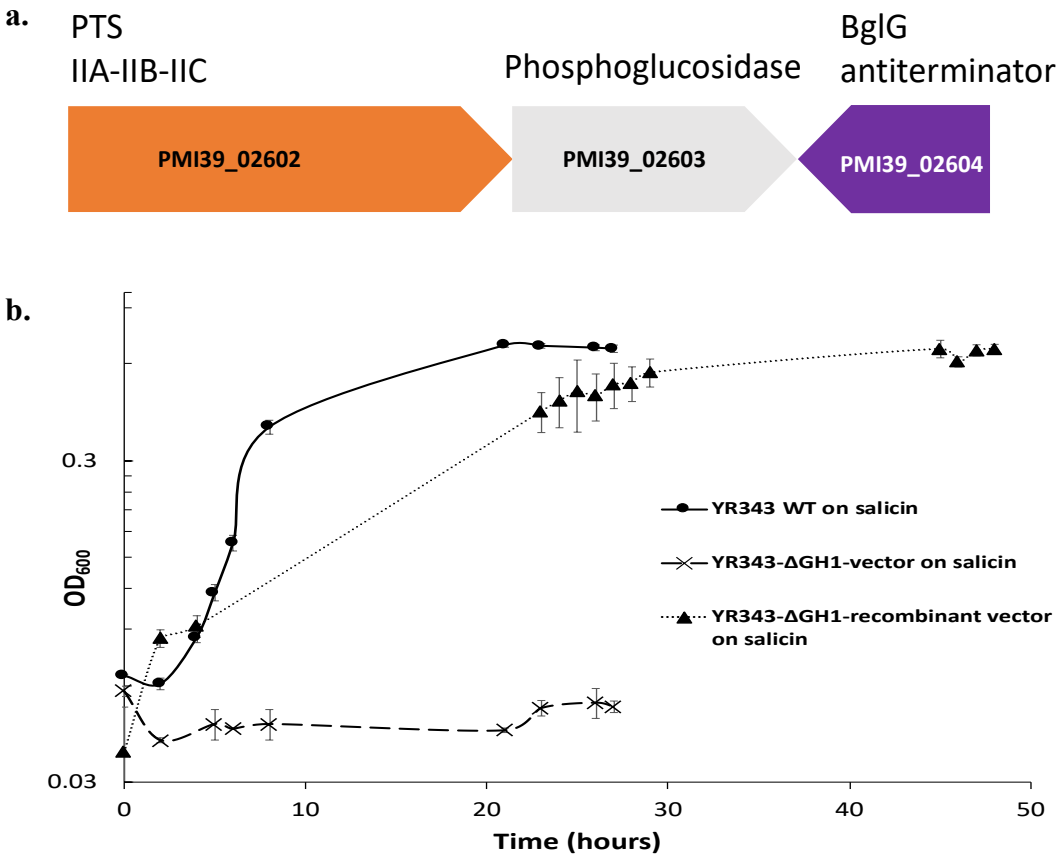


Figure 4.2: *Rahnella* sp. OV744 phosphoglucosidase gene complements the *Pantoea* sp. YR343 ΔGH1 mutant.

a) Structures of *bgl* operon of *Pantoea* and *Rahenlla* are strikingly similar (see figure S4.1) and are proposed to utilize salicin using same mechanism (unpublished work). b) *Rahnella* phosphoglucosidase gene complements the ΔGH1 mutant as the transformed strains (mutant with recombinant vector, ▲) are able to grow on MOPS minimal liquid media containing salicin to an OD_{600} comparable to that of YR343 wild type (WT, ●). It should be noted that it took the transformed mutant longer time to reach the equivalent OD. The mutant transformed with pSRK.km plasmid only (vector, ×) did not grow on the media at all. All the strains in this experiment were able to grow on minimal media containing glucose as sole carbon source (data not shown).

***Pseudomonas* sp. GM16 utilizes salicyl alcohol**

Pseudomonas can utilize and grow on salicyl alcohol as well as glucose minimal media, but it cannot use salicin (figure 4.3a). The growth curve (in figure 4.3a) demonstrates that there is a considerable lag during the growth of *Pseudomonas* in salicyl alcohol minimal media. To eliminate the possibility that the growth could be due to the appearance of mutants, overnight salicyl alcohol subculture of *Pseudomonas* was subcultured again in glucose minimal media before a final subculture in salicyl alcohol minimal media. The longer lag time persisted even in the final salicyl alcohol culture (figure S4.3) which indicates that the lag in growth is due to physiological adaptation, and not because of the emergence of mutants.

For predicting the pathway for degradation of salicyl alcohol, protein sequences of *Pseudomonas* sp. GM16 were annotated using UNIFAM tool followed by comparison of UNIFAM annotations with KEGG database to find the genes that encode proteins for a possible salicyl alcohol degradation pathway. Enzymes that could not be identified in the previous steps were then determined by using the protein family annotations and operon structure information in JGI/IMG database. Finally, the protein sequences of benzyl alcohol dehydrogenase, benzaldehyde dehydrogenase, salicylate hydroxylase and catechol dioxygenase from *P. putida* CSV80 were used for determination of homologs in *Pseudomonas* sp. GM16. All these data led to an operon PMI19_01133 – PMI19_01136 which is proposed to be involved in the conversion of salicyl alcohol to muconate using the aforementioned enzymes (figure 4.3b). All the knowledge acquired from the previous steps led to the hypothesis that *Pseudomonas* sp. GM16 degrades salicyl alcohol through catechol intermediate which is then channeled into the β-ketoadipate pathway. Other proteins involved in the ortho-cleavage pathway of the β-ketoadipate channel were also identified using the process outlined in methods section (see attachment S4.3).

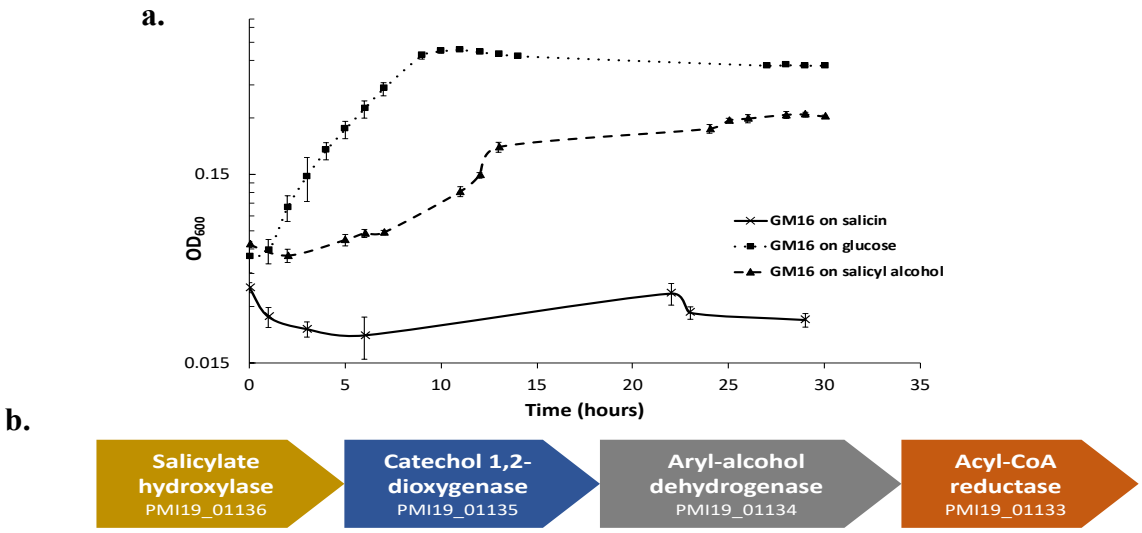


Figure 4.3: *Pseudomonas* sp. GM16 can utilize salicyl alcohol but not salicin.
a) Growth experiments were performed on GM16 to determine the substrates it can utilize. As a positive control, glucose was used. *Pseudomonas* was able to grow on glucose and salicyl alcohol minimal media but could not metabolize salicin. The initial long lag phase when GM16 was inoculated with salicyl alcohol was because of the metabolic adaptation of the strain to salicyl alcohol, and not due to emergence of mutants (see figure S4.3). b) The genes of the operon provided in the figure are proposed to encode proteins involved in salicyl alcohol degradation such that salicyl alcohol is oxidized to salicylate first before being converted to catechol by the action of salicylate hydroxylase.

Proteomics data support the presence of the predicted salicyl alcohol degradation pathway

To support the proposed mechanism of salicyl alcohol degradation, a proteomics study on the glucose and salicyl alcohol cultures harvested at mid-log phase was performed. The comparison of proteome profiles between salicyl alcohol and glucose cultures confirmed that the predicted proteins involved in salicyl alcohol pathway were significantly expressed during growth on salicyl alcohol (table 4.2). Only a few probable proteins, including aryl-alcohol dehydrogenase (PMI19_01134) and 3-ketoacyl-CoA thiolase (PMI19_03256 and PMI19_01217) did not meet the cutoff ($p < 0.05$ and/or protein abundance fold-change between the two substrates ≥ 2). Complete proteomics dataset has been provided in the attachment S4.4. This proteome dataset was further examined for identifying importer of salicyl alcohol, but it could not be determined. For additional support of the proposed salicyl alcohol pathway in *Pseudomonas*, growth assays on the minimal

Table 4.2: Proteins of probable pathway of salicyl alcohol degradation were significantly upregulated in *Pseudomonas* sp. GM16.

Proteome profiles of salicyl alcohol and glucose cultures of *Pseudomonas* were compared. Out of 18 proteins predicted to be involved in the proposed pathway, 16 were detected, and eleven of those were detected only in salicyl alcohol culture. Two proteins that were not detected are highlighted in grey. If a step in the pathway is predicted to be catalyzed by two or more proteins and were detected in the proteome data, they are highlighted in magenta, blue and orange. The detected proteins were significantly expressed in salicyl alcohol cultures than in glucose cultures except for proteins PMI19_01134, PMI19_01184, PMI19_03943, PMI19_03944, PMI19_01217, PMI19_03256 (corrected $p < 0.05$). All the proteins detected in the proteomics experiment are provided in attachment S4.4.

Locus Tag	Protein Name	Present in glucose	BH corrected p-value	Fold change
PMI19_01133	Acyl-CoA reductase	Yes	0.022	6.204
PMI19_01134	aryl-alcohol dehydrogenase	Yes	0.108	6.094
PMI19_01136	salicylate hydroxylase	No	3.82E-06	8.913
PMI19_01135	catechol 1,2-dioxygenase	No	0.0002	8.389
PMI19_01184	catechol 1,2-dioxygenase	Yes	0.0707	6.777
PMI19_03945	catechol 1,2-dioxygenase	No	0.003	3.604
PMI19_01186	muconate cycloisomerase	No	3.11E-05	6.753
PMI19_03943	muconate cycloisomerase	No	0.591	2.533
PMI19_01185	muconolactone delta-isomerase	No	0.008	6.918
PMI19_03944	muconolactone delta-isomerase	No	0.586	2.0191
PMI19_04396	3-oxoadipate enol-lactonase	No	0.006	5.202
PMI19_04846	3-oxoadipate enol-lactonase	No	NA	NA
PMI19_01175	3-oxoadipate CoA-transferase, alpha subunit	No	0.001	5.657
PMI19_01176	3-oxoadipate CoA-transferase, beta subunit	No	0.002	3.828
PMI19_01217	3-ketoacyl-CoA thiolase	Yes	0.812	0.250
PMI19_03256	3-ketoacyl-CoA thiolase	Yes	0.586	-0.728
PMI19_00175	acetyl-CoA acyltransferase	No	NA	NA
PMI19_04401	3-oxoadipyl-CoA thiolase	Yes	0.002	4.602

media containing the intermediates of the pathway as sole carbon sources were also performed. *Pseudomonas* was able to grow on salicylate and catechol further supporting the hypothesized pathway of salicyl alcohol utilization (figure S4.4).

Both *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16 grow in co-culture in salicin by unidirectional cross-feeding mechanism

When monitoring the growth of *Rahnella-Pseudomonas* co-culture in salicin minimal media, a spike in OD_{600} compared to that of *Rahnella* monoculture in salicin was observed (figure 4.4a). Since *Rahnella* can only utilize salicin releasing salicyl alcohol whereas *Pseudomonas* can only degrade salicyl alcohol, it was hypothesized that the higher OD_{600} was observed due to the growth of both strains in salicin such that *Rahnella* sp. OV744 cross-feeds salicyl alcohol to *Pseudomonas*. To test this hypothesis, an experiment was performed to determine if *Pseudomonas* could grow on spent medium of *Rahnella* salicin cultures. The results indicated growth of *Pseudomonas* on the salicin-spent medium, but no growth was observed on the glucose-spent medium (figure 4.4b).

qPCR experiments were also performed to quantify the growth of both strains on salicin co-culture. The qPCR data clearly confirmed that both strains were able to grow on salicin co-culture as suggested by the increase in DNA copy count of each strain over time. Remarkably, it was discovered that the final gene copy count of *Pseudomonas* was higher than that of *Rahnella* even though *Rahnella* sp. OV744 is the provider in the predicted cross-feeding interaction (see figure 4.4c). Since it was an unusual phenomenon, another set of qPCR experiments to determine the stable state of the co-culture was carried out. In this case, the growth of both strains in the co-

culture in salicin was monitored over several generations. The results from this experiment demonstrated that the growth of *Pseudomonas* takes over the growth of *Rahnella* in co-culture. In fact, the ratio between *Pseudomonas* and *Rahnella* DNA copy count in the co-culture was consistently at the ratio ~ 2:1 over ten generations (figure S4.5). Furthermore, when the DNA copy counts of *Rahnella* in both salicin monoculture and co-culture were compared, the difference was found to be significant. Even when the initial co-culture copy count number was doubled (to account for half the number of cells used in the beginning of the co-culture) and then used to calculate the copy count numbers at different time points, the difference between two conditions appeared significant (see figure S4.6). The qPCR dataset and the calculated DNA copy count for each experiment (co-culture qPCR, stable state, monoculture qPCR) are presented in attachments S4.5, S4.6 and S4.7. See supplementary text for calculation method.

Metabolomics and proteomics data elucidate the mechanism of metabolic cross-feeding interaction between *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16

For confirmation of the cross-feeding of salicyl alcohol between the two strains, a time-course metabolite study was performed on the supernatant of the co-culture. *Rahnella* and *Pseudomonas* were grown together on salicin minimal media and supernatant samples were extracted at different time points during growth. GC-MS analyses of the supernatant samples demonstrated that over time salicin was utilized by the co-culture, but salicyl alcohol levels remained negligible (0 – 12 μg/ml, sorbitol equivalent) throughout (figure 4.4d). This experiment further supported the hypothesis that *Rahnella*-secreted salicyl alcohol is utilized by *Pseudomonas* in co-culture.

To further corroborate the salicyl alcohol cross-feeding hypothesis and to determine

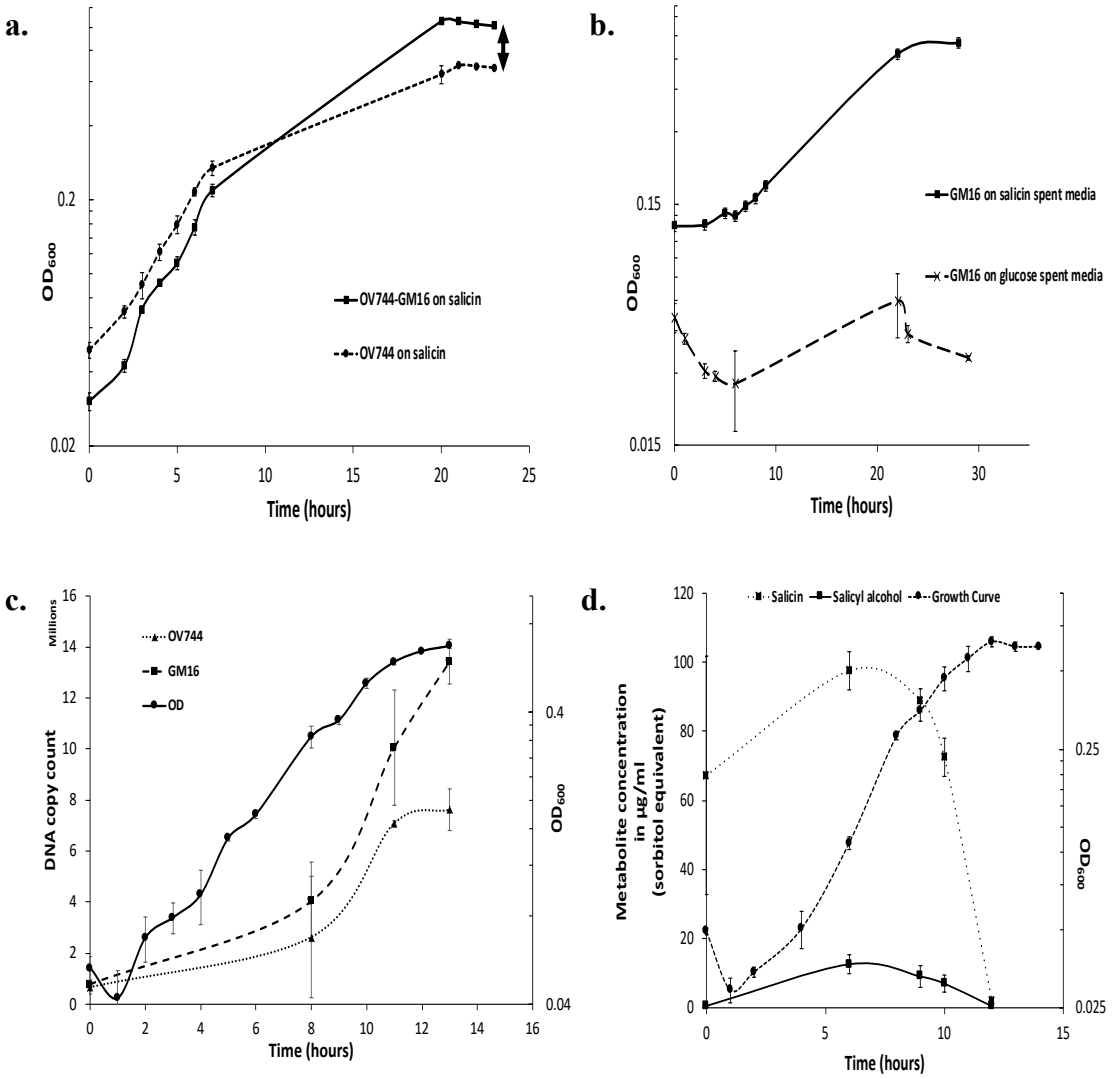


Figure 4.4: *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16 can both grow on salicin co-culture by cross-feeding salicyl alcohol.

a) When *Rahnella* was co-cultured with *Pseudomonas* on salicin, a spike in OD_{600} (visualized by double-arrowed line) was observed compared to the OD_{600} of *Rahnella* salicin monoculture. b) To demonstrate that *Pseudomonas* utilized salicyl alcohol secreted by *Rahnella*, *Pseudomonas* was inoculated with *Rahnella* grown salicin and glucose spent media. *Pseudomonas* was not able to utilize glucose spent media, but it could use salicin culture spent media. c) A qPCR experiment was performed to monitor the growth of individual strains in salicin co-culture. As can be seen in the figure, as the co-culture grew, both *Rahnella* and *Pseudomonas* were able to multiply as demonstrated by the increase in DNA copy count of both strains. d) When a time-course metabolite analysis was performed on the supernatant of salicin co-culture, salicin levels decreased over time whereas salicyl alcohol levels remained low throughout suggesting salicyl alcohol utilization by *Pseudomonas*.

whether the proteins detected in monocultures were also present in salicin co-culture, a proteomics study comparing proteome of *Rahnella* and *Pseudomonas* glucose monocultures and salicin co-culture was carried out. For both *Rahnella* and *Pseudomonas*, proteins of the salicin and salicyl alcohol pathways were significantly upregulated in co-culture compared to those in glucose monocultures. In the case of *Pseudomonas*, only PMI19_03944 (3-oxoadipate enol-lactonase) could not be detected in either substrate conditions. Therefore, the proteomic study further established the proposed model of unidirectional salicyl alcohol cross-feeding between *Rahnella* and *Pseudomonas* in the co-culture (figure 4.5). The proteome data comparing co-culture and glucose monocultures have been provided in the attachment S4.8.

Nature of interaction between *Pseudomonas* sp. GM16 and *Rahnella* sp. OV744 in the salicin co-culture is proposed using growth assays and proteomic analysis

More experiments were performed to determine the nature of interaction between *Pseudomonas* sp. GM16 and *Rahnella* sp. OV744 in salicin co-culture. Association of *Pseudomonas* to *Rahnella* did not provide any growth advantage to *Rahnella* because salicyl alcohol did not inhibit growth of *Rahnella* appreciably at concentrations as high as 3 mM (figure S4.7). Similarly, experiments were performed to determine whether secreted metabolites of *Pseudomonas* sp. GM16 were able to inhibit growth of *Rahnella* sp. OV744. Neither the salicyl alcohol culture nor the salicin co-culture spent media inhibited growth of *Rahnella* (figure S4.8). Furthermore, no inhibiting/toxic compound was detected in the supernatant metabolome of salicin co-culture (data not shown).

To further elucidate the interaction between *Rahnella* and *Pseudomonas*, proteome comparisons between co-culture, *Rahnella* salicin and *Pseudomonas* salicyl alcohol monocultures

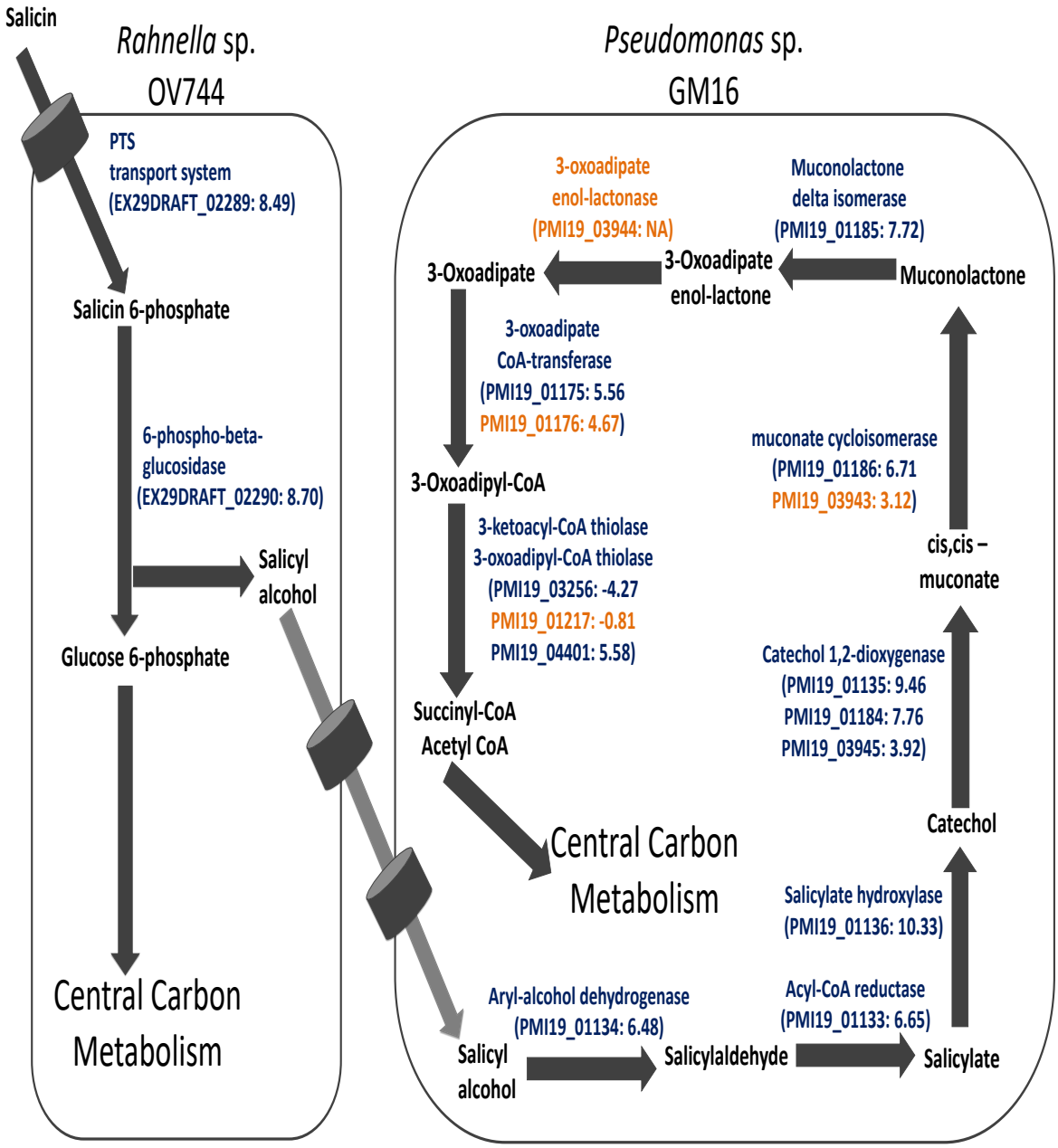


Figure 4.5: Proteomics data support the salicyl alcohol cross-feeding hypothesis between *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16 in the salicin co-culture.
LC-MS/MS was utilized for proteomics on the *Rahnella-Pseudomonas* salicin co-culture and were compared to proteome data of glucose monocultures of each strain. For *Rahnella*, proteins of proposed salicin degradation pathway were upregulated in co-culture than in glucose monoculture. Similarly, protein expression of *Pseudomonas* was higher for proteins in co-culture for the hypothesized salicyl alcohol metabolic pathway than those in glucose monoculture. In the figure, the values within the bracket represent fold change. A BH corrected p-value < 0.05 was used as cutoff and the ones that passed the cutoff are in blue whereas the ones that did not are in orange. Some predicted proteins were not detected in the proteomics data and are omitted from this figure except PMI19_03944 because it is the only protein representing its particular function. The light grey arrow represents yet to be determined transporters of salicyl alcohol.

were performed. The iron-related metabolism was found to be upregulated in *Rahnella* in co-culture compared to salicin monoculture indicating that iron-deficiency could be one of the reasons the growth of *Rahnella* is inhibited in salicin co-culture. Similarly, co-culture and salicyl alcohol proteome comparisons allude to growth inhibition by *Pseudomonas* possibly by more than one mechanism (see attachment S4.9). For instance, hydrogen cyanide synthase proteins *hcnB* and *hcnC* were significantly upregulated in co-culture. These proteins were also abundant in glucose monocultures (see attachment S4.4) suggesting that *Pseudomonas* releases toxic hydrogen cyanide regardless of the presence of *Rahnella*. Similarly, proteins encoded from a gene cluster and related to fatty acid/lipoprotein metabolism were significantly abundant in co-culture compared to salicyl alcohol monocultures of *Pseudomonas*. Numerous studies have demonstrated that lipids are involved in growth inhibition [189, 190]. Finally, type VI secretion system secreted protein VgrG was also upregulated and has been demonstrated to inhibit microbial growth [191, 192].

Chemotaxis assays suggest that salicin might be involved in mediating the interaction between *Populus*, *Rahnella* and *Pseudomonas*

Chemotaxis assays were performed on soft agar plates (0.3%) containing several substrates including salicin and salicyl alcohol (see supplementary text for the method). The results suggest that at least *Pseudomonas* demonstrates attraction towards salicyl alcohol. *Rahnella* demonstrated relatively weak motility towards salicin compared to other substrates (see figure S4.9). Overall, the chemotaxis assays support the hypothesis that salicin might be involved in facilitating the interaction between *Populus*, *Rahnella* and *Pseudomonas*.

Discussion

In this study, we have described a cross-feeding interaction between two *Populus* associates – *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16. Both *Rahnella* and *Pseudomonas* display plant growth promoting properties (data not shown). Likewise, being in association with plants provides numerous benefits to the bacteria such as availability of carbon sources. But the nutrients are never always constant with shifts between favorable and unfavorable sources. Plants undergo not just above-ground, but also below-ground physical stress caused by root disease causing fungi [193] and pests such as gophers, moth larvae and beetles [194-196]. Similarly, throughout the year, the nutrient levels undergo changes in the plants. Studies have revealed that these factors cause fluctuations in phenolic glycoside (PG) production not only in the above-ground tissues, but also in the roots with reports suggesting a swing of ~ 21% between low and high nutrient conditions [197, 198].

During times of higher production, plants release more salicin, and higher phenolic glycosides such as salicortin and tremulacin can also be readily degraded to salicin [199, 200]. Therefore, plant- and soil-associated bacteria that have evolved with a system that can readily degrade phenolic glycoside have a survival advantage and increased growth in plant-related niches. *Rahnella* sp. OV744 was able to successfully colonize *Populus* rhizosphere possibly because of its salicin degrading ability. It is also likely that *Rahnella* could have acquired this system over time because of its association with *Populus*. In *E. coli*, it has been proposed that mutations in the promotor region of silent *bgl* operon or in genes encoding other glucosidases are required for the activation of salicin-degradation ability [1, 4]. A *bgl*-like operon is found in *Rahnella* and most likely uses a similar mechanism of salicin degradation as in *E. coli* as was corroborated by not only proteomics but also metabolomics data. The intermediate of the proposed mechanism, salicin

6-phosphate was detected in the metabolomic analysis of supernatant of salicin cultures. To the best of our knowledge, this is the first time this metabolite has been identified. The metabolite was not detected in the metabolomic study involving inner metabolites possibly because salicin 6-phosphate is immediately consumed within the cell. It was detected in supernatant likely because it was released due to cell lysis and probably is not consumed as it never gets in contact with the right enzyme.

In addition to the support from both the proteomic and metabolomic approaches, molecular assay using complementation experiment was also utilized to further validate the proposed pathway. This approach was carried out by Raghunand and Mahadevan [2] to determine the *bgl* pathway in *Klebsiella aerogenes*. In our case, *Pantoea* sp. YR343 mutant strain (in which phosphoglucosidase (*bgl*B) gene has been deleted) was used. *Pantoea* is taxonomically close to *Rahnella*, and it contains a *bgl* like operon similar to that of *Rahnella* and *E. coli*. Likewise, *Pantoea* has been proposed to utilize the same pathway as *Rahnella* to degrade salicin (unpublished). Therefore, this mutant strain was used to determine whether *Rahnella* phosphoglucosidase gene could functionally replace the *bgl*B of *Pantoea* mutant strain. The complementation assay results demonstrate that *Rahnella bgl*B is able to successfully complement the lost phosphoglucosidase function in *Pantoea* mutant indicating that both organisms have the same pathway of salicin degradation. Taken together with results from other experiments, proposed pathway of salicin utilization exists in *Rahnella*.

The by-product of the salicin metabolism in *Rahnella* is salicyl alcohol which is released into the media as suggested by the supernatant metabolite analysis of the *Rahnella* salicin cultures (figure 4.1b). *Pseudomonas* sp. GM16 can use salicyl alcohol for growth as we have demonstrated in this study. Salicyl alcohol is not generally considered a growth substrate. In fact, salicyl alcohol

has been demonstrated to inhibit microbial growth [201]. Furthermore, the intermediates of the proposed salicyl alcohol pathway are also toxic to microbes. Salicylic acid is known to be a growth-inhibiting compound of *P. synringae* [202]. It is an uncoupler of oxidative phosphorylation [203]. Salicylaldehyde, another intermediate, is also proven to be antimicrobial [204]. Catechol has also been demonstrated to inhibit growth of microbes [205] because it can be oxidized to quinones [206, 207], which are potent electrophiles and can crosslink or degrade amino acids [208, 209]. On the other hand, studies have demonstrated that few strains of *Pseudomonas* are able to utilize not only benzyl alcohol and similar compounds but also the intermediates of proposed pathway using distinct branches of naphthalene degradation channel [165-168, 210]. *Pseudomonas* sp. GM16 possesses a salicyl alcohol pathway which appears to require several enzymes of the naphthalene degradation pathway. The salicyl alcohol degradation capability likely provides *Pseudomonas* sp. GM16 distinct advantage to colonize *Populus* roots. Another possibility is that this pathway might be a result of evolutionary specialization because of *Pseudomonas*'s association with salicin-degrading organisms in the root microbiome.

In this study, multiple proteome comparisons were performed and were found to be helpful in determination of most likely proteins involved in the degradation of respective substrates. For instance, one of the probable proteins involved in salicin transport could be EX29DRAFT_03584 based on genomic data (see figure S4.1). The proteome comparison between glucose and salicin monocultures of *Rahnella* did not conclusively indicate whether the protein was involved or not. When *Rahnella* proteome was analyzed for the co-culture, however, the protein was not detected which suggests that EX29DRAFT_03584 is likely not involved in salicin transport (see attachments S4.2 and S4.8). The exact mechanism of export of salicyl alcohol, however, could not be characterized because no prior knowledge of the exporters is found in literature. Protein

sequences of known exporters of related compounds such as toluene were queried against the *Rahnella* sp. OV744 genome, and then gene hits were matched against the proteome datasets [185-188]. Several candidates have been identified, but more work is required to determine the mechanisms of salicyl alcohol export in *Rahnella* (see table S4.1). The importer of salicyl alcohol in *Pseudomonas* sp. GM16 could not be determined conclusively using the proteomics data either. Therefore, at this point, we cannot predict the exact mechanism of salicyl alcohol uptake in *Pseudomonas*, and it could very likely be a diffusion process which is the case for related compound 4-hydroxybenzoic acid [211]. It is also possible that the proteomics method was not able to separate the transporters in detectable amount. Similarly, we could not confirm whether single or multiple proteins are involved in each step of the salicyl alcohol degradation pathway.

In regard to co-culture, the proteomics and metabolite data provide direct evidence that this two-member pathway is utilized by both strains to grow cooperatively on salicin by cross-feeding salicyl alcohol. Since both strains are morphologically similar and because of quick dilution of fluorescence, microscopic techniques were not feasible making qPCR the technique of choice to elucidate the growth of both strains in co-culture. Surprisingly, the data also showed that *Pseudomonas* was able to take over the growth in the co-culture as seen in figure 4.4c. This is an interesting phenomenon because *Rahnella* is the provider of carbon source in the proposed unidirectional cross-feeding interaction. The qPCR plots suggest that *Pseudomonas* either has a faster growth rate than that of *Rahnella* or that the *Pseudomonas* suppresses the growth of the *Rahnella* in co-culture. Of the two cases, the latter might be a better explanation because *Pseudomonas* grows slowly in salicyl alcohol monoculture (figure 4.3). More assays and analyses were performed in this study to determine the nature of interaction between *Rahnella* and *Pseudomonas* in salicin co-culture, but so far nothing conclusive can be assumed. We were not

able to determine any toxic/inhibiting compound in the supernatant metabolite analysis of co-culture which indicates that *Pseudomonas* potentially inhibits growth by using either volatile compounds or by contact inhibition. Another possibility is that the *Pseudomonas*-released compound is utilized rapidly by *Rahnella,* and hence not detected in metabolomic study. We also analyzed the proteome data and compared protein profiles between salicin co-culture and *Rahnella* salicin and *Pseudomonas* salicyl alcohol monocultures. These comparisons suggest multiple causes, such as iron deficiency or production of inhibiting compounds (volatile or related to contact) by *Pseudomonas,* that could be related to growth inhibition of *Rahnella* in salicin co-culture. More work will be required to conclusively determine the exact process.

It should be understood that laboratory and natural conditions might not be same. Moreover, these two strains do not occur in the same niche (*Rahnella* - rhizosphere and *Pseudomonas* - endosphere) but do co-exist with similar organisms in nature. If the results are to be taken at the face value, we hypothesize that *Pseudomonas* might be a selfish cheater such that this interaction neither harms nor helps *Rahnella* in natural system. Such "cheating" interactions have been widely reported in the literature [151, 212, 213]. Furthermore, another possible reason these two strains can co-exist in nature could be that both organisms benefit from being in contact with *Populus*.

Our study has demonstrated a mechanism by which phenolic glycoside such as salicin could be involved in microbial recruitment in the *Populus* microbiota. These compounds are used by salicin-degrading organisms such as *Rahnella* which utilizes *bgl* system to consume salicin with subsequent release of salicyl alcohol as a by-product. The advantage of such system is poorly understood, but one study has demonstrated that this aglycone molecule can act in bacterial defense by killing bacterivores [79]. We propose another possible advantage of this system which is to

attract other microorganisms toward *Populus* microbiome (as suggested by chemotaxis assays). Therefore, we offer a model of tripartite relationship in which *Rahnella* acts as mediator between *Populus* and *Pseudomonas*, but only benefits from its association with *Populus*, not with *Pseudomonas* (figure 4.6).

In summary, we utilized multi-data approaches to demonstrate that salicin-degrading strains (such as *Rahnella* sp. OV744) and salicyl alcohol degrading strains (such as *Pseudomonas* sp. GM16) could co-exist in the same niche because of by-product cross-feeding interaction. Therefore, phenolic glycosides could be used for attracting at least two different types of organisms into the *Populus* microbiome. The co-culture model used in this project could be further studied in depth to understand the nature of interaction between the two members. This type of study could be crucial in elucidating plant colonization by multiple bacteria and shedding light on the evolution of relationship between *Populus* and its microbiome.

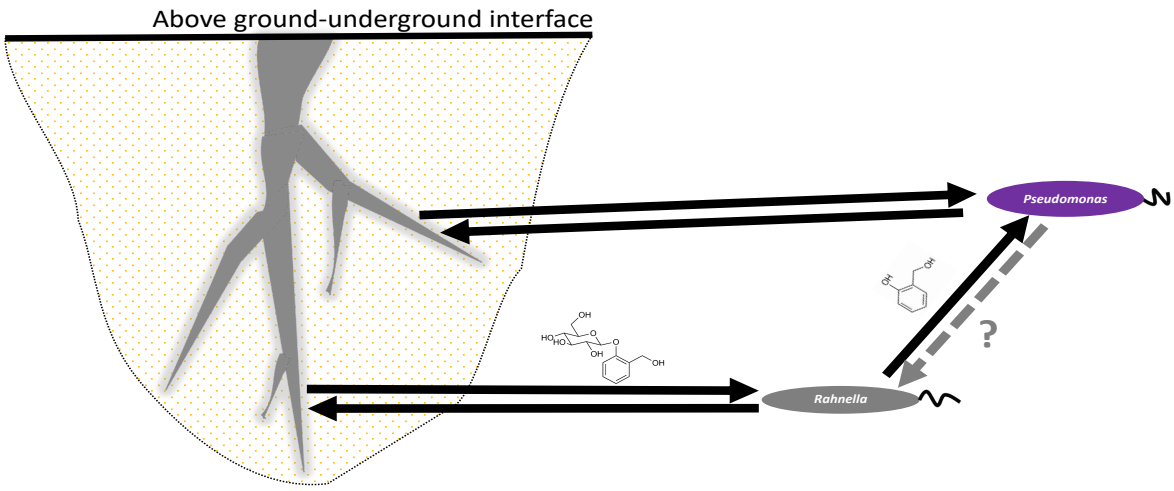


Figure 4.6: Proposed model of mechanism of microbial recruitment of *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16 employed by *Populus* through salicin/PGs.
In the suggested model, plants provide carbon substrates such as PGs to *Rahnella* and *Pseudomonas*. *Rahnella* can utilize PGs such as salicin and releases salicyl alcohol as by-product. Salicyl alcohol is then utilized by *Pseudomonas*. In return, no benefit providing activity has yet been observed. Instead, *Pseudomonas* appears to inhibit the growth of *Rahnella*, but exact mechanism of the interaction is not established yet. Both *Pseudomonas* and *Rahnella* have demonstrated plant-growth promoting properties suggesting mutually beneficial relationship between these microbes and *Populus*. In the figure, dark arrows suggest beneficial interaction and grey, dotted arrow represents possible harmful activity.

Chapter 5

Conclusions and future directions

Populus is an important commodified plant that has been extensively studied for the generation of biofuels. For robust growth of such plants, it is crucial to understand their microbiome because the microbiome provides distinct capabilities to plants such as scavenging nutrients, fixing atmospheric nitrogen and combating pathogens. To understand the microbial association with plants, it is quintessential to determine metabolic interactions that occur between the host and its microbiome. Therefore, we have focused our attention on one of the relatively lesser known class of metabolites – phenolic glycosides (PGs) that could be involved in inter-kingdom communications. Since *Populus* commits significant amount of carbon to the production of PGs, we seek to understand the mechanisms of PG degradation in *Populus* isolates and to propose that PGs might be important for *Populus* to fine-tune its microbiome.

In this investigation, we used genomic and experimental data to identify salicin degrading strains that associate with *Populus*. Using the literature-guided knowledge of these salicin utilization systems and experimental screening process, we were able to demonstrate that salicin degradation is a prominent feature of *Populus*-associated strains. *Populus* contains a diverse microbiome consisting of strains belonging to multiple families. Regardless, salicin utilization appears to be widespread and not consistent with the microbial genotype excluding a few cases. In fact, as the study suggests, bacterial degradation of salicin is possibly partially an effect of host selective pressure. The presence of separate mechanisms of salicin degradation in *Populus*-associated bacteria suggest that host-metabolic interactions likely leads phylogenetically diverse bacteria to evolve relatively distinct mechanisms to utilize PGs. This proposal was evident in four different bacteria that were studied as part of this thesis.

One of the first bacteria that is described in this study is *Caulobacter* sp. AP07. *Caulobacter* strains have not been well-characterized as salicin degraders. Only one study on Lac

system has been described in literature to explain salicin degradation by *Caulobacter*. But our study has proposed that *Caulobacter* strains actually possess two different systems – Lac and Sal to degrade salicin. Both genomic and proteomic data suggest that these two systems are not only present, but in fact, active in *Caulobacter* sp. AP07 when grown on salicin minimal media. Present study cannot unequivocally conclude that these two systems act independently or in conjunction with each other, but future study could focus on that aspect. If the two systems act together, it would suggest that lactose dehydrogenases are required to facilitate the cleavage activity of glycosyl hydrolases such as those of Sal system.

In the case of *Citrobacter* sp. YR018, data suggests a working hypothesis that host might be able to drive the metabolic evolution of its microbial partner. This study does not provide a direct evidence that in nature host can provide selective pressure to force bacteria to evolve into a salicin utilizer, but it appears that bacteria can adjust to metabolic cues and evolve to grow on salicin. It should be noted that *Citrobacter* gained the salicin growth function because it probably already has the machinery to transport and cleave molecules similar to salicin and requires mutations in the genes or promoter regions in order to utilize salicin. Future work on this strain might be focused on determining the exact loci that are mutated in order for *Citrobacter* to have gained the salicin utilization ability. In this study, we have provided potential targets that could be investigated to understand this phenomenon.

As part of the systematic study on the salicin degradation processes in *Populus*-associated microbial strains, we identified a unique organism that is able to utilize salicin entirely. We were able to demonstrate through several quantitative methods that *Agrobacterium rhizogenes* OK036 possesses a novel salicin degradation mechanism in which the salicyl alcohol moiety is channeled into the gentisate pathway. There were a few unique features of this pathway that have been

elaborated in chapter 3. Salicylate 5-hydroxylase (CoA forming) or S5HCoA was identified in this study as a protein containing both ligase and monooxygenase activities such that it could transform salicylate (an intermediate of salicin degradation) to gentisyl-CoA by itself. In the future, this pathway could be further characterized by studying other organisms, and we have provided a few candidate strains that could be investigated.

Finally, we also investigated a co-culture model of a salicin degrader – *Rahnella* sp. OV744 and a salicyl alcohol degrader – *Pseudomonas* sp. GM16 in which it was demonstrated that *Rahnella* cross-fed salicyl alcohol to *Pseudomonas* in the salicin co-culture in order for both of these strains to grow. The study has huge implications in the area of multitrophic interactions between plant and its microbiome and in between the members of the microbiome. PG such as salicin are released by plant and such complex metabolites could be one of the mechanisms utilized by host such as *Populus* to shape its microbiome. Organisms that are able to utilize these metabolites or by-products of these metabolites are probably more equipped to sustain and thrive as members of the *Populus* microbiome. Since the scope of this study was limited to determination of cross-feeding interaction between these two strains, the nature of interaction is only briefly discussed. In the future, it will be interesting to determine what other kind of interactions occur between the two members of the co-culture model. The experiments so far indicate that *Pseudomonas* is an opportunist and a social cheater, but more experiments are required to prove this hypothesis. Future endeavors might include a time-course proteomics study, membrane-separated growth assays or microfluidic assays in order to further understand the interactions that occur between these two *Populus* isolates. Investigations into determination of mechanisms of salicyl alcohol transport could also be performed and proteomic data might be able to help identify targets. Finally, the proteomic data itself could be further analyzed in order to fully understand the

changes that occur when *Rahnella*/*Pseudomonas* are grown on salicin/salicyl alcohol versus on glucose. Moreover, the comparison of proteome data of co-culture and monocultures in salicin and salicyl alcohol could also be utilized to further our understanding of the changes between separate growth conditions.

Overall, this study is solely focused on the salicin degradation mechanisms of the *Populus* isolates, and how such mechanisms are important to bacterial strains that associate with *Populus*. Salicin degradation is possibly essential to defining the structure of the *Populus* microbiome, and we have tried to identify and characterize diverse strains in this study to propose that salicin systems have evolved independent of the strain type. This study also demonstrates that salicin not only acts as a carbon source but also as a metabolic means by which *Populus* can shape its microbiome. Therefore, we believe this study will not only provide a guide for future studies of similar ilk, but also promote the knowledge of salicin utilization systems in bacteria associated with *Populus*.

Bibliography

1. Schnetz K, Toloczyki C, Rak B: **Beta-glucoside (bgl) operon of Escherichia coli K-12: nucleotide sequence, genetic organization, and possible evolutionary relationship to regulatory components of two Bacillus subtilis genes**. *Journal of Bacteriology* 1987, **169**(6):2579-2590.
2. Raghunand TR, Mahadevan S: **The Î²-glucoside genes of Klebsiella aerogenes: conservation and divergence in relation to the cryptic bgl genes of Escherichia coli**. *FEMS Microbiology Letters* 2003, **223**(2):267-274.
3. Desai SK, Nandimath K, Mahadevan S: **Diverse pathways for salicin utilization in Shigella sonnei and Escherichia coli carrying an impaired bgl operon**. *Arch Microbiol* 2010, **192**(10):821-833.
4. Zangoui P, Vashishtha K, Mahadevan S: **Evolution of aromatic beta-glucoside utilization by successive mutational steps in Escherichia coli**. *J Bacteriol* 2015, **197**(4):710-716.
5. Babst BA, Harding SA, Tsai C-J: **Biosynthesis of phenolic glycosides from phenylpropanoid and benzenoid precursors in Populus**. *Journal of chemical ecology* 2010, **36**(3):286-297.
6. Bulgarelli D, Schlaeppi K, Spaepen S, van Themaat EVL, Schulze-Lefert P: **Structure and functions of the bacterial microbiota of plants**. *Annual review of plant biology* 2013, **64**:807-838.
7. Boeckler GA, Gershenzon J, Unsicker SB: **Phenolic glycosides of the Salicaceae and their role as anti-herbivore defenses**. *Phytochemistry* 2011, **72**(13):1497-1509.
8. Bulgarelli D, Rott M, Schlaeppi K, van Themaat EVL, Ahmadinejad N, Assenza F, Rauf P, Huettel B, Reinhardt R, Schmelzer E: **Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota**. *Nature* 2012, **488**(7409):91.
9. Grayer RJ, Kokubun T: **Plant–fungal interactions: the search for phytoalexins and other antifungal compounds from higher plants**. *Phytochemistry* 2001, **56**(3):253-263.
10. Knief C, Delmotte N, Chaffron S, Stark M, Innerebner G, Wassmann R, von Mering C, Vorholt JA: **Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice**. *ISME J* 2012, **6**(7):1378-1390.
11. Zaitlin M, Hull R: **Plant virus-host interactions**. *Annual Review of Plant Physiology* 1987, **38**(1):291-315.
12. Bodenhausen N, Horton MW, Bergelson J: **Bacterial communities associated with the leaves and the roots of Arabidopsis thaliana**. *PLoS One* 2013, **8**(2):e56329.
13. Kim M, Singh D, Lai-Hoe A, Go R, Rahim RA, Ainuddin A, Chun J, Adams JM: **Distinctive phyllosphere bacterial communities in tropical trees**. *Microbial Ecology* 2012, **63**(3):674-681.
14. Lambais MR, Lucheta AR, Crowley DE: **Bacterial community assemblages associated with the phyllosphere, dermosphere, and rhizosphere of tree species of the Atlantic forest are host taxon dependent**. *Microbial ecology* 2014, **68**(3):567-574.
15. Ottesen AR, Peña AG, White JR, Pettengill JB, Li C, Allard S, Rideout S, Allard M, Hill T, Evans P: **Baseline survey of the anatomical microbial ecology of an important food plant: Solanum lycopersicum (tomato)**. *BMC microbiology* 2013, **13**(1):114.
16. Iniguez AL, Dong Y, Triplett EW: **Nitrogen fixation in wheat provided by Klebsiella pneumoniae 342**. *Molecular Plant-Microbe Interactions* 2004, **17**(10):1078-1085.
17. 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. *Biology and Fertility of Soils* 2015, **51**(4):403-415.
18. Schuhegger R, Ihring A, Gantner S, Bahnweg G, Knappe C, Vogg G, Hutzler P, Schmid M, Van Breusegem F, Eberl LEO *et al*: **Induction of systemic resistance in tomato by N-acyl-L-homoserine lactone-producing rhizosphere bacteria**. *Plant, Cell and Environment* 2006, **29**(5):909-918.
19. Morrissey JP, Dow JM, Mark GL, O'Gara F: **Are microbes at the root of a solution to world food production?: Rational exploitation of interactions between microbes and plants can help to transform agriculture**. *EMBO reports* 2004, **5**(10):922-926.
20. Gottel NR, Castro HF, Kerley M, Yang Z, Pelletier DA, Podar M, Karpinets T, Uberbacher E, Tuskan GA, Vilgalys R: **Distinct microbial communities within the endosphere and rhizosphere of Populus deltoides roots across contrasting soil types**. *Appl Environ Microbiol* 2011, **77**(17):5934-5944.
21. Doty SL: **Functional Importance of the Plant Endophytic Microbiome: Implications for Agriculture, Forestry, and Bioenergy**. 2017:1-5.
22. Hurek T, Handley LL, Reinhold-Hurek B, Piché Y: **Azoarcus grass endophytes contribute fixed nitrogen to the plant in an unculturable state**. *Molecular Plant-Microbe Interactions* 2002, **15**(3):233-242.
23. Molina-Favero C, Creus CM, Simontacchi M, Puntarulo S, Lamattina L: **Aerobic nitric oxide production by Azospirillum brasilense Sp245 and its influence on root architecture in tomato**. *Molecular plant-microbe interactions* 2008, **21**(7):1001-1009.
24. Chiou TJ, Liu H, Harrison MJ: **The spatial expression patterns of a phosphate transporter (MtPT1) from Medicago truncatula indicate a role in phosphate transport at the root/soil interface**. *The Plant Journal* 2001, **25**(3):281-293.
25. Campestre MP, Castagno LN, Estrella MJ, Ruiz OA: **Lotus japonicus plants of the Gifu B-129 ecotype subjected to alkaline stress improve their Fe2+ bio-availability through inoculation with Pantoea eucalypti M91**. *Journal of plant physiology* 2016, **192**:47-55.
26. Dobbelaere S, Croonenborghs A, Thys A, Broek AV, Vanderleyden J: **Phytostimulatory effect of Azospirillum brasilense wild type and mutant strains altered in IAA production on wheat**. *Plant and soil* 1999, **212**(2):153-162.
27. Leveau JH, Lindow SE: **Utilization of the plant hormone indole-3-acetic acid for growth by Pseudomonas putida strain 1290**. *Appl Environ Microbiol* 2005, **71**(5):2365-2371.
28. Glick BR, Jacobson CB, Schwarze MM, Pasternak J: **1-Aminocyclopropane-1-carboxylic acid deaminase mutants of the plant growth promoting rhizobacterium Pseudomonas putida GR12-2 do not stimulate canola root elongation**. *Canadian Journal of Microbiology* 1994, **40**(11):911-915.
29. Ryu CM, Farag MA, Hu CH, Reddy MS, Wei HX, Pare PW, Kloepper JW: **Bacterial volatiles promote growth in Arabidopsis**. *Proc Natl Acad Sci U S A* 2003, **100**(8):4927-4932.
30. Blom D, Fabbri C, Connor E, Schiestl F, Klauser D, Boller T, Eberl L, Weisskopf L: **Production of plant growth modulating volatiles is widespread among rhizosphere bacteria and strongly depends on culture conditions**. *Environmental microbiology* 2011, **13**(11):3047-3058.

31. Combes-Meynet E, Pothier JF, Moënne-Loccoz Y, Prigent-Combaret C: **The Pseudomonas secondary metabolite 2, 4-diacetylphloroglucinol is a signal inducing rhizoplane expression of Azospirillum genes involved in plant-growth promotion**. *Molecular plant-microbe interactions* 2011, **24**(2):271-284.
32. Felix G, Duran JD, Volko S, Boller T: **Plants have a sensitive perception system for the most conserved domain of bacterial flagellin**. *The Plant Journal* 1999, **18**(3):265-276.
33. Lopez-Gomez M, Sandal N, Stougaard J, Boller T: **Interplay of flg22-induced defence responses and nodulation in Lotus japonicus**. *Journal of experimental botany* 2011, **63**(1):393-401.
34. Schuhegger R, Ihring A, Gantner S, Bahnweg G, Knappe C, Vogg G, Hutzler P, Schmid M, Van Breusegem F, Eberl L: **Induction of systemic resistance in tomato by N-acyl-L-homoserine lactone-producing rhizosphere bacteria**. *Plant, Cell & Environment* 2006, **29**(5):909-918.
35. Doebley JF, Gaut BS, Smith BD: **The molecular genetics of crop domestication**. *Cell* 2006, **127**(7):1309-1321.
36. Stokstad E: **The nitrogen fix**. In.: American Association for the Advancement of Science; 2016.
37. Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrektson A, Kunin V, Del Rio TG: **Defining the core Arabidopsis thaliana root microbiome**. *Nature* 2012, **488**(7409):86.
38. Tuskan GA, DiFazio SP, Teichmann T: **Poplar genomics is getting popular: the impact of the poplar genome project on tree research**. *Plant Biol (Stuttg)* 2004, **6**(1):2-4.
39. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A: **The genome of black cottonwood, Populus trichocarpa (Torr. & Gray)**. *science* 2006, **313**(5793):1596-1604.
40. Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM: **The role of root exudates in rhizosphere interactions with plants and other organisms**. *Annu Rev Plant Biol* 2006, **57**:233-266.
41. Compant S, Clément C, Sessitsch A: **Plant growth-promoting bacteria in the rhizo- and endosphere of plants: their role, colonization, mechanisms involved and prospects for utilization**. *Soil Biology and Biochemistry* 2010, **42**(5):669-678.
42. Julkunen-Tiitto R: **A chemotaxonomic survey of phenolics in leaves of northern Salicaceae species**. *Phytochemistry* 1986, **25**(3):663-667.
43. Lindroth R, Pajutee M: **Chemical analysis of phenolic glycosides: art, facts, and artifacts**. *Oecologia* 1987, **74**(1):144-148.
44. Donaldson JR, Stevens MT, Barnhill HR, Lindroth RL: **Age-related shifts in leaf chemistry of clonal aspen (Populus tremuloides)**. *J Chem Ecol* 2006, **32**(7):1415-1429.
45. Forster N, Ulrichs C, Zander M, Katzel R, Mewis I: **Factors influencing the variability of antioxidative phenolic glycosides in Salix species**. *J Agric Food Chem* 2010, **58**(14):8205-8210.
46. Lindroth RL, Hwang S-Y: **Clonal variation in foliar chemistry of quaking aspen (Populus tremuloides Michx.)**. *Biochemical Systematics and Ecology* 1996, **24**(5):357-364.

47. Erwin EA, Turner MG, Lindroth RL, Romme WH: **Secondary Plant Compounds in Seedling and Mature Aspen (Populus tremuloides) in Yellowstone National Park, Wyoming**. *The American Midland Naturalist* 2001, **145**(2):299-308.
48. Kopper BJ, Lindroth RL: **Effects of elevated carbon dioxide and ozone on the phytochemistry of aspen and performance of an herbivore**. *Oecologia* 2003, **134**(1):95-103.
49. Hale BK, Herms DA, Hansen RC, Clausen TP, Arnold D: **Effects of drought stress and nutrient availability on dry matter allocation, phenolic glycosides, and rapid induced resistance of poplar to two lymantriid defoliators**. *Journal of chemical ecology* 2005, **31**(11):2601-2620.
50. Broberg CL, Inkster JAH, Borden JH: **Phenological and chemical differences among hybrid poplar clones (Salicaceae) varying in resistance to Cryptorhynchus lapathi (L.)(Coleoptera: Curculionidae)**. *Biochemical Systematics and Ecology* 2010, **38**(1):29-48.
51. Clausen TP, Reichardt PB, Bryant JP, Werner RA, Post K, Frisby K: **Chemical model for short-term induction in quaking aspen (Populus tremuloides) foliage against herbivores**. *Journal of Chemical Ecology* 1989, **15**(9):2335-2346.
52. Fields MJ, Orians CM: **Specificity of phenolic glycoside induction in willow seedlings (Salix sericea) in response to herbivory**. *J Chem Ecol* 2006, **32**(12):2647-2656.
53. Rubert-Nason KF, Couture JJ, Major IT, Constabel CP, Lindroth RL: **Influence of genotype, environment, and gypsy moth herbivory on local and systemic chemical defenses in trembling aspen (Populus tremuloides)**. *Journal of chemical ecology* 2015, **41**(7):651-661.
54. Gould GG, Jones CG, Rifleman P, Perez A, Coleman JS: **Variation in eastern cottonwood (Populus deltoides Bartr.) phloem sap content caused by leaf development may affect feeding site selection behavior of the aphid, Chaitophorous populicola Thomas (Homoptera: Aphididae)**. *Environmental entomology* 2014, **36**(5):1212-1225.
55. Morant AV, Jorgensen K, Jorgensen C, Paquette SM, Sanchez-Perez R, Moller BL, Bak S: **beta-Glucosidases as detonators of plant chemical defense**. *Phytochemistry* 2008, **69**(9):1795-1813.
56. Mason CJ, Couture JJ, Raffa KF: **Plant-associated bacteria degrade defense chemicals and reduce their adverse effects on an insect defoliator**. *Oecologia* 2014, **175**(3):901-910.
57. Massad TJ, Trumbore SE, Ganbat G, Reichelt M, Unsicker S, Boeckler A, Gleixner G, Gershenzon J, Ruehlow S: **An optimal defense strategy for phenolic glycoside production in Populus trichocarpa--isotope labeling demonstrates secondary metabolite production in growing leaves**. *New Phytol* 2014, **203**(2):607-619.
58. Lindroth RL, Scriber JM, Hsia MS: **Chemical ecology of the tiger swallowtail: mediation of host use by phenolic glycosides**. *Ecology* 1988, **69**(3):814-822.
59. Reichardt P, Bryant J, Mattes B, Clausen T, Chapin F, Meyer M: **Winter chemical defense of Alaskan balsam poplar against snowshoe hares**. *Journal of Chemical Ecology* 1990, **16**(6):1941-1959.
60. Pentzold S, Zagrobelny M, Rook F, Bak S: **How insects overcome two-component plant chemical defence: plantβ-glucosidases as the main target for herbivore adaptation**. *Biological Reviews* 2014, **89**(3):531-551.

61. Haruta M, Pedersen JA, Constabel CP: **Polyphenol oxidase and herbivore defense in trembling aspen (Populus tremuloides): cDNA cloning, expression, and potential substrates**. *Physiologia Plantarum* 2001, **112**(4):552-558.
62. Barreto G, Madureira D, Capani F, Aon-Bertolino L, Saraceno E, Alvarez-Giraldez LD: **The role of catechols and free radicals in benzene toxicity: an oxidative DNA damage pathway**. *Environ Mol Mutagen* 2009, **50**(9):771-780.
63. Boeckler GA, Paetz C, Feibicke P, Gershenzon J, Unsicker SB: **Metabolism of poplar salicinoids by the generalist herbivore Lymantria dispar (Lepidoptera)**. *Insect biochemistry and molecular biology* 2016, **78**:39-49.
64. Mahdi JG: **Medicinal potential of willow: A chemical perspective of aspirin discovery**. *Journal of Saudi Chemical Society* 2010, **14**(3):317-322.
65. El-Shemy HA, Aboul-Enein AM, Aboul-Enein MI, Issa SI, Fujita K: **The effect of willow leaf extracts on human leukemic cells in vitro**. *BMB Reports* 2003, **36**(4):387-389.
66. Mahdi J, Mahdi A, Mahdi A, Bowen I: **The historical analysis of aspirin discovery, its relation to the willow tree and antiproliferative and anticancer potential**. *Cell proliferation* 2006, **39**(2):147-155.
67. Mahdi J, Al-Musayeib N, Mahdi E, Pepper C: **Pharmacological importance of simple phenolic compounds on inflammation, cell proliferation and apoptosis with a special reference to β-D-salicin and hydroxybenzoic acid**. In.: SAGE Publications Sage UK: London, England; 2013.
68. Tschaplinski TJ, Plett JM, Engle NL, Deveau A, Cushman KC, Martin MZ, Doktycz MJ, Tuskan GA, Brun A, Kohler A: **Populus trichocarpa and Populus deltoides exhibit different metabolomic responses to colonization by the symbiotic fungus Laccaria bicolor**. *Molecular Plant-Microbe Interactions* 2014, **27**(6):546-556.
69. Veach AM, Morris R, Yip DZ, Yang ZK, Engle NL, Cregger MA, Tschaplinski TJ, Schadt CW: **Rhizosphere microbiomes diverge among Populus trichocarpa plant-host genotypes and chemotypes, but it depends on soil origin**. *Microbiome* 2019, **7**(1).
70. Schnetz K, Rak B: **Regulation of the bgl operon of Escherichia coli by transcriptional antitermination**. *The EMBO Journal* 1988, **7**(10):3271-3277.
71. el Hassouni M, Chippaux M, Barras F: **Analysis of the Erwinia chrysanthemi arb genes, which mediate metabolism of aromatic beta-glucosides**. *Journal of bacteriology* 1990, **172**(11):6261-6267.
72. Faure D, Desair J, Keijers V, Bekri MA, Proost P, Henrissat B, Vanderleyden J: **Growth of Azospirillum irakense KBC1 on the aryl β-glucoside salicin requires either salA or salB**. *Journal of bacteriology* 1999, **181**(10):3003-3009.
73. Arellano BH, Ortiz JD, Manzano J, Chen JC: **Identification of a dehydrogenase required for lactose metabolism in Caulobacter crescentus**. *Appl Environ Microbiol* 2010, **76**(9):3004-3014.
74. Charaoui-Boukerzaza S, Hugouvieux-Cotte-Pattat N: **A family 3 glycosyl hydrolase of Dickeya dadantii 3937 is involved in the cleavage of aromatic glucosides**. *Microbiology* 2013, **159**(Pt 11):2395-2404.
75. Watt DK, Ono H, Hayashi K: **Agrobacterium tumefaciensβ-glucosidase is also an effective β-xylosidase, and has a high transglycosylation activity in the presence of alcohols**. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* 1998, **1385**(1):78-88.

76. Shakya M, Gottel N, Castro H, Yang ZK, Gunter L, Labbe J, Muchero W, Bonito G, Vilgalys R, Tuskan G *et al*: **A multifactor analysis of fungal and bacterial community structure in the root microbiome of mature Populus deltoides trees**. *PLoS One* 2013, **8**(10):e76382.
77. Lindroth R, Peterson S: **Effects of plant phenols of performance of southern armyworm larvae**. *Oecologia* 1988, **75**(2):185-189.
78. Pinto SS, Diogo HP: **Calorimetric studies on the phenolic glycoside D(-)-salicin**. *J Pharm Sci* 2008, **97**(12):5354-5362.
79. Sonowal R, Nandimath K, Kulkarni SS, Koushika SP, Nanjundiah V, Mahadevan S: **Hydrolysis of aromatic beta-glucosides by non-pathogenic bacteria confers a chemical weapon against predators**. *Proc Biol Sci* 2013, **280**(1762):20130721.
80. Mason CJ, Rubert-Nason KF, Lindroth RL, Raffa KF: **Aspen defense chemicals influence midgut bacterial community composition of gypsy moth**. *J Chem Ecol* 2015, **41**(1):75-84.
81. Schaefler S, Malamy A: **Taxonomic investigations on expressed and cryptic phospho-β-glucosidases in Enterobacteriaceae**. *Journal of bacteriology* 1969, **99**(2):422-433.
82. Gulati A, Mahadevan S: **Mechanism of catabolite repression in the bgl operon of Escherichia coli: involvement of the anti-terminator BglG, CRP-cAMP and EIIAGlc in mediating glucose effect downstream of transcription initiation**. *Genes to Cells* 2000, **5**(4):239-250.
83. Hayano K, Fukui S: **Purification and properties of 3-ketosucrose-forming enzyme from the cells of Agrobacterium tumefaciens**. *Journal of Biological Chemistry* 1967, **242**(16):3665-3672.
84. Van Beeumen J, De Ley J: **Hexopyranoside: cytochrome c oxidoreductase from Agrobacterium tumefaciens**. *European journal of biochemistry* 1968, **6**(3):331-343.
85. Schuerman P, Liu J, Mou H, Dandekar A: **3-Ketoglycoside-mediated metabolism of sucrose in E. coli as conferred by genes from Agrobacterium tumefaciens**. *Applied microbiology and biotechnology* 1997, **47**(5):560-565.
86. Faure D, Saier Jr MH, Vanderleyden J: **An evolutionary alternative system for aryl beta-glucosides assimilation in bacteria**. *Journal of molecular microbiology and biotechnology* 2001, **3**(3):467-470.
87. Somers E, Keijers V, Ptacek D, Ottoy MH, Srinivasan M, Vanderleyden J, Faure D: **The salCAB operon of Azospirillum irakense, required for growth on salicin, is repressed by SalR, a transcriptional regulator that belongs to the LacI/GalR family**. *Molecular and General Genetics MGG* 2000, **263**(6):1038-1046.
88. Castle L, Smith K, Morris R: **Cloning and sequencing of an Agrobacterium tumefaciens beta-glucosidase gene involved in modifying a vir-inducing plant signal molecule**. *Journal of bacteriology* 1992, **174**(5):1478-1486.
89. Yoshida E, Hidaka M, Fushinobu S, Koyanagi T, Minami H, Tamaki H, Kitaoka M, Katayama T, Kumagai H: **Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 beta-glucosidase from Kluyveromyces marxianus**. *Biochem J* 2010, **431**(1):39-49.
90. Goyal K, Selvakumar P, Hayashi K: **Characterization of a thermostable β-glucosidase (BglB) from Thermotoga maritima showing transglycosylation activity**. *Journal of Molecular Catalysis B: Enzymatic* 2001, **15**(1-3):45-53.

91. Brown SD, Utturkar SM, Klingeman DM, Johnson CM, Martin SL, Land ML, Lu T-YS, Schadt CW, Doktycz MJ, Pelletier DA: **Twenty-one genome sequences from Pseudomonas species and 19 genome sequences from diverse bacteria isolated from the rhizosphere and endosphere of Populus deltoides**. In.: Am Soc Microbiol; 2012.
92. Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F *et al*: **Genomic features of bacterial adaptation to plants**. *Nat Genet* 2017, **50**(1):138-150.
93. Blair PM, Land ML, Piatek MJ, Jacobson DA, Lu T-YS, Doktycz MJ, Pelletier DA: **Exploration of the biosynthetic potential of the Populus microbiome**. *mSystems* 2018, **3**(5):e00045-00018.
94. Neidhardt FC, Bloch PL, Smith DF: **Culture medium for enterobacteria**. *Journal of bacteriology* 1974, **119**(3):736-747.
95. Shivatare RS, Phopase ML, Nagore DH, Nipanikar SU, Chitlange SS: **Development and Validation of HPLC Analytical Protocol for Quantification of Salicin from Salix alba L**. *L, Inventi Rapid: Pharm Analysis & Quality Assurance* 2015, **61**:2015.
96. Scopes R: **Measurement of protein by spectrophotometry at 205 nm**. *Analytical biochemistry* 1974, **59**(1):277-282.
97. Clarkson SM, Giannone RJ, Kridelbaugh DM, Elkins JG, Guss AM, Michener JK: **Construction and optimization of a heterologous pathway for protocatechuate catabolism in Escherichia coli enables bioconversion of model aromatic compounds**. *Appl Environ Microbiol* 2017, **83**(18):e01313-01317.
98. 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**. *Journal of proteome research* 2014, **13**(8):3679-3684.
99. Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ: **Semi-supervised learning for peptide identification from shotgun proteomics datasets**. *Nature methods* 2007, **4**(11):923.
100. Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, Padki A, Zhao X, Dubchak I, Hugenholtz P, Anderson I: **The integrated microbial genomes (IMG) system**. *Nucleic acids research* 2006, **34**(suppl_1):D344-D348.
101. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *Journal of molecular biology* 1990, **215**(3):403-410.
102. Chapman B, Chang J: **Biopython: Python tools for computational biology**. *ACM Sigbio Newsletter* 2000, **20**(2):15-19.
103. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO: **SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB**. *Nucleic Acids Research* 2007, **35**(21):7188-7196.
104. Felsenstein J: **PHYLIP (Version 3.6) Phylogeny Inference Package**. *Cladistics* 1989, **5**:164-166.
105. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY: **ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data**. *Methods in Ecology and Evolution* 2017, **8**(1):28-36.
106. Cheng J, Galili T: **RStudio Inc, Bostock M, Palmer J: d3heatmap: Interactive Heat Maps Using 'htmlwidgets' and 'D3. js'. 2015**. *R package version 06*, **1**.
107. de Weert S, Vermeiren H, Mulders IH, Kuiper I, Hendrickx N, Bloemberg GV, Vanderleyden J, De Mot R, Lugtenberg BJ: **Flagella-driven chemotaxis towards**

exudate components is an important trait for tomato root colonization by Pseudomonas fluorescens. *Molecular Plant-Microbe Interactions* 2002, **15**(11):1173-1180.
108. Rangel-Castro JI, Killham K, Ostle N, Nicol GW, Anderson IC, Scrimgeour CM, Ineson P, Meharg A, Prosser JI: **Stable isotope probing analysis of the influence of liming on root exudate utilization by soil microorganisms**. *Environmental Microbiology* 2005, **7**(6):828-838.
109. Broeckling CD, Broz AK, Bergelson J, Manter DK, Vivanco JM: **Root exudates regulate soil fungal community composition and diversity**. *Appl Environ Microbiol* 2008, **74**(3):738-744.
110. Liu Q-M, Ten LN, Im W-T, Lee S-T, Yoon M-H: **Caulobacter ginsengisoli sp. nov., a novel stalked bacterium isolated from ginseng cultivating soil**. *J Microbiol Biotechnol* 2010, **20**:15-20.
111. Gao J-l, Sun P, Sun X-h, Tong S, Yan H, Han M-l, Mao X-j, Sun J-g: **Caulobacter zeae sp. nov. and Caulobacter radicis sp. nov., novel endophytic bacteria isolated from maize root (Zea mays L.)**. *Systematic and applied microbiology* 2018, **41**(6):604-610.
112. Saibi W, Gargouri A: **Cellobiose dehydrogenase influences the production of S. microspora beta-glucosidase**. *World J Microbiol Biotechnol* 2012, **28**(1):23-29.
113. Turbe-Doan A, Arfi Y, Record E, Estrada-Alvarado I, Levasseur A: **Heterologous production of cellobiose dehydrogenases from the basidiomycete Coprinopsis cinerea and the ascomycete Podospora anserina and their effect on saccharification of wheat straw**. *Applied Microbiology and Biotechnology* 2012, **97**(11):4873-4885.
114. Chen K, Liu X, Long L, Ding S: **Cellobiose dehydrogenase from Volvariella volvacea and its effect on the saccharification of cellulose**. *Process biochemistry* 2017, **60**:52-58.
115. Sedmera P, Halada P, Peterbauer C, Volc J: **A new enzyme catalysis: 3,4-dioxidation of some aryl β-d-glycopyranosides by fungal pyranose dehydrogenase**. *Tetrahedron Letters* 2004, **45**(47):8677-8680.
116. Peterbauer CK, Volc J: **Pyranose dehydrogenases: biochemical features and perspectives of technological applications**. *Appl Microbiol Biotechnol* 2010, **85**(4):837-848.
117. Miyazaki R, Yamazaki T, Yoshimatsu K, Kojima K, Asano R, Sode K, Tsugawa W: **Elucidation of the intra- and inter-molecular electron transfer pathways of glucoside 3-dehydrogenase**. *Bioelectrochemistry* 2018, **122**:115-122.
118. Parker L, Hall B: **A fourth Escherichia coli gene system with the potential to evolve beta-glucoside utilization**. *Genetics* 1988, **119**(3):485-490.
119. Kricker M, Hall BG: **Directed evolution of cellobiose utilization in Escherichia coli K12**. *Molecular biology and evolution* 1984, **1**(2):171-182.
120. Khan Z, Rho H, Firrincieli A, Hung SH, Luna V, Masciarelli O, Kim S-H, Doty SL: **Growth enhancement and drought tolerance of hybrid poplar upon inoculation with endophyte consortia**. *Current Plant Biology* 2016, **6**:38-47.
121. Otieno N, Lally RD, Kiwanuka S, Lloyd A, Ryan D, Germaine KJ, Dowling DN: **Plant growth promotion induced by phosphate solubilizing endophytic Pseudomonas isolates**. *Frontiers in Microbiology* 2015, **6**:745.
122. Weston DJ, Pelletier DA, Morrell-Falvey JL, Tschaplinski TJ, Jawdy SS, Lu T-Y, Allen SM, Melton SJ, Martin MZ, Schadt CW: **Pseudomonas fluorescens induces strain-dependent and strain-independent host plant responses in defense networks,**

primary metabolism, photosynthesis, and fitness. *Molecular Plant-Microbe Interactions* 2012, **25**(6):765-778.
123. Huang X-F, Chaparro JM, Reardon KF, Zhang R, Shen Q, Vivanco JM: **Rhizosphere interactions: root exudates, microbes, and microbial communities**. *Botany* 2014, **92**(4):267-275.
124. Faure D: **The family-3 glycoside hydrolases: from housekeeping functions to host-microbe interactions**. *Appl Environ Microbiol* 2002, **68**(4):1485-1490.
125. Winans SC: **Two-way chemical signaling in Agrobacterium-plant interactions**. *Microbiology and Molecular Biology Reviews* 1992, **56**(1):12-31.
126. Escobar MA, Dandekar AM: **Agrobacterium tumefaciens as an agent of disease**. *Trends in Plant Science* 2003, **8**(8):380-386.
127. Gonzalez-Mula A, Lachat J, Mathias L, Naquin D, Lamouche F, Mergaert P, Faure D: **The biotroph Agrobacterium tumefaciens thrives in tumors by exploiting a wide spectrum of plant host metabolites**. *New Phytol* 2019, **222**(1):455-467.
128. Shaw C, Ashby A, Brown A, Royal C, Loake G, Shaw C: **virA and virG are the Ti-plasmid functions required for chemotaxis of Agrobacterium tumefaciens towards acetosyringone**. *Molecular microbiology* 1988, **2**(3):413-417.
129. Dahal SH, Robert L.; Burdick, Leah; Poudel, Suresh; Blair, Patricia M.; Pelletier, Dale A.: **Phylogenetic and functional diversity in salicin degrading Populus bacterial isolates**. In.
130. Dahal SH, Gregory B.; Chourey, Karuna; Timm, Collin M.; Lu, Tse-Yuan S.; Engle, Nancy L.; Tschaplinski, Timothy J.; Pelletier, Dale A.: **Mechanism of unidirectional by-product cross-feeding of salicyl alcohol between Populus microbiome isolates**. In.
131. Garcia DC, Mohr BP, Dovgan JT, Hurst GB, Standaert RF, Doktycz MJ: **Elucidating the potential of crude cell extracts for producing pyruvate from glucose**. *Synthetic Biology* 2018, **3**(1):ysy006.
132. Zimmermann M, Sauer U, Zamboni N: **Quantification and mass isotopomer profiling of α-keto acids in central carbon metabolism**. *Analytical chemistry* 2014, **86**(6):3232-3237.
133. Wu Z, Gao W, Phelps MA, Wu D, Miller DD, Dalton JT: **Favorable effects of weak acids on negative-ion electrospray ionization mass spectrometry**. *Analytical chemistry* 2004, **76**(3):839-847.
134. Ishiyama D, Vujaklija D, Davies J: **Novel pathway of salicylate degradation by Streptomyces sp. strain WA46**. *Appl Environ Microbiol* 2004, **70**(3):1297-1306.
135. Benning MM, Wesenberg G, Liu R, Taylor KL, Dunaway-Mariano D, Holden HM: **The three-dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from Pseudomonas sp. strain CBS-3**. *Journal of Biological Chemistry* 1998, **273**(50):33572-33579.
136. Zhuang Z, Song F, Takami H, Dunaway-Mariano D: **The BH1999 protein of Bacillus halodurans C-125 is gentisyl-coenzyme A thioesterase**. *J Bacteriol* 2004, **186**(2):393-399.
137. Lynch J, Whipps J: **Substrate flow in the rhizosphere**. *Plant and soil* 1990, **129**(1):1-10.
138. Rangel-Castro JI, Prosser JI, Ostle N, Scrimgeour CM, Killham K, Meharg AA: **Flux and turnover of fixed carbon in soil microbial biomass of limed and unlimed plots of an upland grassland ecosystem**. *Environmental Microbiology* 2005, **7**(4):544-552.

139. Badri DV, Vivanco JM: **Regulation and function of root exudates**. *Plant, Cell & Environment* 2009, **32**(6):666-681.
140. Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, Okura VK, Zhou Y, Chen L, Wood GE, Almeida NF: **The genome of the natural genetic engineer Agrobacterium tumefaciens C58**. *science* 2001, **294**(5550):2317-2323.
141. Mathews SL, Hannah H, Samagaio H, Martin C, Rodriguez-Rassi E, Matthysse AG: **Glycoside Hydrolase Genes Are Required for Virulence of Agrobacterium tumefaciens on Bryophyllum daigremontiana and Tomato**. *Appl Environ Microbiol* 2019.
142. Biver S, Stroobants A, Portetelle D, Vandenbol M: **Two promising alkaline β-glucosidases isolated by functional metagenomics from agricultural soil, including one showing high tolerance towards harsh detergents, oxidants and glucose**. *Journal of industrial microbiology & biotechnology* 2014, **41**(3):479-488.
143. Quadri LE, Sello J, Keating TA, Weinreb PH, Walsh CT: **Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin**. *Chemistry & biology* 1998, **5**(11):631-645.
144. Quadri LE, Keating TA, Patel HM, Walsh CT: **Assembly of the Pseudomonas aeruginosa nonribosomal peptide siderophore pyochelin: In vitro reconstitution of aryl-4, 2-bisthiazoline synthetase activity from PchD, PchE, and PchF**. *Biochemistry* 1999, **38**(45):14941-14954.
145. Di Gennaro P, Terreni P, Masi G, Botti S, De Ferra F, Bestetti G: **Identification and characterization of genes involved in naphthalene degradation in Rhodococcus opacus R7**. *Appl Microbiol Biotechnol* 2010, **87**(1):297-308.
146. Mohamed ME, Zaar A, Ebenau-Jehle C, Fuchs G: **Reinvestigation of a new type of aerobic benzoate metabolism in the proteobacterium Azoarcus evansii**. *J Bacteriol* 2001, **183**(6):1899-1908.
147. Grund E, Denecke B, Eichenlaub R: **Naphthalene degradation via salicylate and gentisate by Rhodococcus sp. strain B4**. *Appl Environ Microbiol* 1992, **58**(6):1874-1877.
148. Crawford RL, Hutton SW, Chapman PJ: **Purification and properties of gentisate 1, 2-dioxygenase from Moraxella osloensis**. *Journal of Bacteriology* 1975, **121**(3):794-799.
149. Shen XH, Jiang CY, Huang Y, Liu ZP, Liu SJ: **Functional identification of novel genes involved in the glutathione-independent gentisate pathway in Corynebacterium glutamicum**. *Appl Environ Microbiol* 2005, **71**(7):3442-3452.
150. Zhou N-Y, Fuenmayor SL, Williams PA: **nag Genes ofRalstonia (Formerly Pseudomonas) sp. Strain U2 Encoding Enzymes for Gentisate Catabolism**. *Journal of Bacteriology* 2001, **183**(2):700-708.
151. Hibbing ME, Fuqua C, Parsek MR, Peterson SB: **Bacterial competition: surviving and thriving in the microbial jungle**. *Nat Rev Microbiol* 2010, **8**(1):15-25.
152. Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M, Gophna U, Sharan R, Ruppin E: **Competitive and cooperative metabolic interactions in bacterial communities**. *Nat Commun* 2011, **2**:589.
153. Kemen E: **Microbe-microbe interactions determine oomycete and fungal host colonization**. *Curr Opin Plant Biol* 2014, **20**:75-81.

154. Stubbendieck RM, Vargas-Bautista C, Straight PD: **Bacterial Communities: Interactions to Scale**. *Front Microbiol* 2016, **7**:1234.
155. Rios-Covian D, Gueimonde M, Duncan SH, Flint HJ, de los Reyes-Gavilan CG: **Enhanced butyrate formation by cross-feeding between Faecalibacterium prausnitzii and Bifidobacterium adolescentis**. *FEMS Microbiol Lett* 2015, **362**(21).
156. Graber JR, Breznak JA: **Folate cross-feeding supports symbiotic homoacetogenic spirochetes**. *Appl Environ Microbiol* 2005, **71**(4):1883-1889.
157. Krause SM, Johnson T, Karunaratne YS, Fu Y, Beck DA, Chistoserdova L, Lidstrom ME: **Lanthanide-dependent cross-feeding of methane-derived carbon is linked by microbial community interactions**. *Proceedings of the National Academy of Sciences* 2017, **114**(2):358-363.
158. D'Souza G, Shitut S, Preussger D, Yousif G, Waschina S, Kost C: **Ecology and evolution of metabolic cross-feeding interactions in bacteria**. *Nat Prod Rep* 2018, **35**(5):455-488.
159. Belenguer A, Duncan SH, Calder AG, Holtrop G, Louis P, Lobley GE, Flint HJ: **Two routes of metabolic cross-feeding between Bifidobacterium adolescentis and butyrate-producing anaerobes from the human gut**. *Appl Environ Microbiol* 2006, **72**(5):3593-3599.
160. LaSarre B, McCully AL, Lennon JT, McKinlay JB: **Microbial mutualism dynamics governed by dose-dependent toxicity of cross-fed nutrients**. *The ISME journal* 2017, **11**(2):337.
161. Rosenthal AZ, Matson EG, Eldar A, Leadbetter JR: **RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture**. *ISME J* 2011, **5**(7):1133-1142.
162. Hoek TA, Axelrod K, Biancalani T, Yurtsev EA, Liu J, Gore J: **Resource Availability Modulates the Cooperative and Competitive Nature of a Microbial Cross-Feeding Mutualism**. *PLoS Biol* 2016, **14**(8):e1002540.
163. Compant S, Reiter B, Sessitsch A, Nowak J, Clement C, Ait Barka E: **Endophytic colonization of Vitis vinifera L. by plant growth-promoting bacterium Burkholderia sp. strain PsJN**. *Appl Environ Microbiol* 2005, **71**(4):1685-1693.
164. Mason CJ, Lowe-Power TM, Rubert-Nason KF, Lindroth RL, Raffa KF: **Interactions between bacteria and aspen defense chemicals at the phyllosphere–herbivore interface**. *Journal of chemical ecology* 2016, **42**(3):193-201.
165. Paliwal V, Raju SC, Modak A, Phale PS, Purohit HJ: **Pseudomonas putida CSV86: a candidate genome for genetic bioaugmentation**. *PLoS One* 2014, **9**(1):e84000.
166. Shrivastava R, Basu A, Phale PS: **Purification and characterization of benzyl alcohol- and benzaldehyde-dehydrogenase from Pseudomonas putida CSV86**. *Archives of microbiology* 2011, **193**(8):553-563.
167. Katagiri M, Takemori S, Nakazawa K, Suzuki H, Akagi K: **Benzylalcohol dehydrogenase, a new alcohol dehydrogenase from Pseudomonas sp**. *Biochimica et Biophysica Acta (BBA)-Enzymology* 1967, **139**(1):173-176.
168. Basu A, Dixit S, Phale P: **Metabolism of benzyl alcohol via catechol ortho-pathway in methylnaphthalene-degrading Pseudomonas putida CSV86**. *Applied microbiology and biotechnology* 2003, **62**(5-6):579-585.
169. Tschaplinski TJ, Standaert RF, Engle NL, Martin MZ, Sangha AK, Parks JM, Smith JC, Samuel R, Jiang N, Pu Y: **Down-regulation of the caffeic acid O-methyltransferase**

gene in switchgrass reveals a novel monolignol analog. *Biotechnology for biofuels* 2012, **5**(1):71.
170. Thompson MR, Chourey K, Froelich JM, Erickson BK, VerBerkmoes NC, Hettich RL: **Experimental approach for deep proteome measurements from small-scale microbial biomass samples**. *Analytical chemistry* 2008, **80**(24):9517-9525.
171. Washburn MP, Wolters D, Yates III JR: **Large-scale analysis of the yeast proteome by multidimensional protein identification technology**. *Nature biotechnology* 2001, **19**(3):242.
172. Wolters DA, Washburn MP, Yates JR: **An automated multidimensional protein identification technology for shotgun proteomics**. *Analytical chemistry* 2001, **73**(23):5683-5690.
173. Estenson K, Hurst GB, Standaert RF, Bible AN, Garcia D, Chourey K, Doktycz MJ, Morrell-Falvey JL: **Characterization of Indole-3-acetic acid biosynthesis and the effects of this phytohormone on the proteome of the plant-associated microbe Pantoea sp. YR343**. *Journal of proteome research* 2018, **17**(4):1361-1374.
174. Tabb DL, Fernando CG, Chambers MC: **MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis**. *Journal of proteome research* 2007, **6**(2):654-661.
175. Ma Z-Q, Dasari S, Chambers MC, Litton MD, Sobecki SM, Zimmerman LJ, Halvey PJ, Schilling B, Drake PM, Gibson BW: **IDPicker 2.0: Improved protein assembly with high discrimination peptide identification filtering**. *Journal of proteome research* 2009, **8**(8):3872-3881.
176. Moore RE, Young MK, Lee TD: **Qscore: an algorithm for evaluating SEQUEST database search results**. *Journal of the American Society for Mass Spectrometry* 2002, **13**(4):378-386.
177. Elias JE, Gygi SP: **Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry**. *Nature methods* 2007, **4**(3):207.
178. Team RC: **R: A Language and Environment for Statistical Computing http://www**. *R-project org* 2014.
179. Liu H, Sadygov RG, Yates JR: **A model for random sampling and estimation of relative protein abundance in shotgun proteomics**. *Analytical chemistry* 2004, **76**(14):4193-4201.
180. Zhang Y, Wen Z, Washburn MP, Florens L: **Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins**. *Analytical chemistry* 2010, **82**(6):2272-2281.
181. Khan SR, Gaines J, Roop RM, Farrand SK: **Broad-host-range expression vectors with tightly regulated promoters and their use to examine the influence of TraR and TraM expression on Ti plasmid quorum sensing**. *Appl Environ Microbiol* 2008, **74**(16):5053-5062.
182. Bertani G: **Studies on lysogenesis i.: The mode of phage liberation by lysogenic escherichia coli1**. *Journal of bacteriology* 1951, **62**(3):293.
183. Chai J, Kora G, Ahn T-H, Hyatt D, Pan C: **Functional phylogenomics analysis of bacteria and archaea using consistent genome annotation with UniFam**. *BMC evolutionary biology* 2014, **14**(1):207.
184. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M: **KEGG: Kyoto encyclopedia of genes and genomes**. *Nucleic acids research* 1999, **27**(1):29-34.

185. Garcia V, Godoy P, Daniels C, Hurtado A, Ramos JL, Segura A: **Functional analysis of new transporters involved in stress tolerance in Pseudomonas putida DOT-T1E**. *Environ Microbiol Rep* 2010, **2**(3):389-395.
186. Fillet S, Daniels C, Pini C, Krell T, Duque E, Bernal P, Segura A, Lu D, Zhang X, Ramos JL: **Transcriptional control of the main aromatic hydrocarbon efflux pump in Pseudomonas**. *Environ Microbiol Rep* 2012, **4**(2):158-167.
187. Segura A, Molina L, Fillet S, Krell T, Bernal P, Munoz-Rojas J, Ramos JL: **Solvent tolerance in Gram-negative bacteria**. *Curr Opin Biotechnol* 2012, **23**(3):415-421.
188. Ramos JL, Sol Cuenca M, Molina-Santiago C, Segura A, Duque E, Gomez-Garcia MR, Udaondo Z, Roca A: **Mechanisms of solvent resistance mediated by interplay of cellular factors in Pseudomonas putida**. *FEMS Microbiol Rev* 2015, **39**(4):555-566.
189. Fay J, Farias R: **The inhibitory action of fatty acids on the growth of**. *Microbiology* 1975, **91**(2):233-240.
190. Velázquez-Becerra C, Macías-Rodríguez LI, López-Bucio J, Altamirano-Hernández J, Flores-Cortez I, Valencia-Cantero E: **A volatile organic compound analysis from Arthrobacter agilis identifies dimethylhexadecylamine, an amino-containing lipid modulating bacterial growth and Medicago sativa morphogenesis in vitro**. *Plant and soil* 2011, **339**(1-2):329-340.
191. Hachani A, Allsopp LP, Oduko Y, Filloux A: **The VgrG proteins are "a la carte" delivery systems for bacterial type VI effectors**. *J Biol Chem* 2014, **289**(25):17872-17884.
192. Basler M, Ho BT, Mekalanos JJ: **Tit-for-tat: type VI secretion system counterattack during bacterial cell-cell interactions**. *Cell* 2013, **152**(4):884-894.
193. Ross WD: **Fungi associated with root diseases of aspen in Wyoming**. *Canadian Journal of Botany* 1976, **54**(8):734-744.
194. Andersen DC, MacMahon JA: **Population dynamics and bioenergetics of a fossorial herbivore, Thomomys talpoides (Rodentia: Geomyidae), in a spruce-fir sere**. *Ecological Monographs* 1981, **51**(2):179-202.
195. Wallner WE, Wagner DL, Parker BL, Tobi DL: **Bioecology of the conifer swift moth, Korscheltellus gracilis, a root feeder associated with spruce-fir decline**. *In: Baranchikov, Yuri N; Mattson, William J; Hain, Fred P; Payne, Thomas L, eds Forest Insect Guilds: Patterns of Interaction with Host Trees; 1989 August 13-17; Abakan, Siberia, USSR Gen Tech Rep NE-153 Radnor, PA: US Department of Agriculture, Forest Service, Northeastern Forest Experiment Station: 199-204* 1991, **153**:199-204.
196. Coyle D, Mattson W, Raffa K: **Invasive root feeding insects in natural forest ecosystems of North America**. *Root feeders: An ecosystem perspective* 2008:134-151.
197. Stevens MT, Gusse AC, Lindroth RL: **Root Chemistry in Populus tremuloides: Effects of Soil Nutrients, Defoliation, and Genotype**. *Journal of Chemical Ecology* 2014, **40**(1):31-38.
198. Randriamanana TR, Nybakken L, Lavola A, Aphalo PJ, Nissinen K, Julkunen-Tiitto R: **Sex-related differences in growth and carbon allocation to defence in Populus tremula as explained by current plant defence theories**. *Tree Physiology* 2014, **34**(5):471-487.
199. Ruuhola T: **Dynamics of salicylates in willows and its relation to herbivory**: University of Joensuu Joensuu, Finland; 2001.

200. Ruuhola T, Julkunen-Tiitto R, Vainiotalo P: **In vitro degradation of willow salicylates**. *Journal of chemical ecology* 2003, **29**(5):1083-1097.
201. Masika P, Sultana N, Afolayan A, Houghton P: **Isolation of two antibacterial compounds from the bark of Salix capensis**. *South African Journal of Botany* 2005, **71**(3-4):441-443.
202. Cameron RK, Zaton K: **Intercellular salicylic acid accumulation is important for age-related resistance in Arabidopsis to Pseudomonas syringae**. *Physiological and molecular plant pathology* 2004, **65**(4):197-209.
203. Terada H: **Uncouplers of oxidative phosphorylation**. *Environmental health perspectives* 1990, **87**:213-218.
204. Gross J, Schumacher K, Schmidtberg H, Vilcinskas A: **Protected by fumigants: beetle perfumes in antimicrobial defense**. *Journal of chemical ecology* 2008, **34**(2):179.
205. Hussain H, Badawy A, Elshazly A, Elsayed A, Krohn K, Riaz M, Schulz B: **Chemical constituents and antimicrobial activity of Salix subserrata**. *Records of Natural Products* 2011, **5**(2):133.
206. Appel HM: **Phenolics in ecological interactions: the importance of oxidation**. *Journal of Chemical Ecology* 1993, **19**(7):1521-1552.
207. Pourcel L, Routaboul J-M, Cheynier V, Lepiniec L, Debeaujon I: **Flavonoid oxidation in plants: from biochemical properties to physiological functions**. *Trends in plant science* 2007, **12**(1):29-36.
208. Bittner S: **When quinones meet amino acids: chemical, physical and biological consequences**. *Amino acids* 2006, **30**(3):205-224.
209. Friedman M: **Food browning and its prevention: an overview**. *Journal of Agricultural and Food chemistry* 1996, **44**(3):631-653.
210. Jiménez JI, Nogales J, García JL, Díaz E: **A Genomic View of the Catabolism of Aromatic Compounds in Pseudomonas**. 2010:1297-1325.
211. Ramos-Gonzalez MI, Godoy P, Alaminos M, Ben-Bassat A, Ramos JL: **Physiological characterization of Pseudomonas putida DOT-T1E tolerance to p-hydroxybenzoate**. *Appl Environ Microbiol* 2001, **67**(9):4338-4341.
212. West SA, Griffin AS, Gardner A, Diggle SP: **Social evolution theory for microorganisms**. *Nat Rev Microbiol* 2006, **4**(8):597-607.
213. Morris JJ, Lenski RE, Zinser ER: **The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss**. *MBio* 2012, **3**(2):e00036-00012.
214. Parales RE, Luu RA, Chen GY, Liu X, Wu V, Lin P, Hughes JG, Nesteryuk V, Parales JV, Ditty JL: **Pseudomonas putida F1 has multiple chemoreceptors with overlapping specificity for organic acids**. *Microbiology* 2013, **159**(Pt 6):1086-1096.
215. Zdobnov EM, Apweiler R: **InterProScan–an integration platform for the signature-recognition methods in InterPro**. *Bioinformatics* 2001, **17**(9):847-848.
216. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR: **HMMER web server: 2015 update**. *Nucleic acids research* 2015, **43**(W1):W30-W38.
217. Zhao M, Chen Y, Qu D, Qu H: **TSdb: a database of transporter substrates linking metabolic pathways and transporter systems on a genome scale via their shared substrates**. *Science China Life Sciences* 2011, **54**(1):60-64.

218. Geertz-Hansen HM, Blom N, Feist AM, Brunak S, Petersen TN: **Cofactory: Sequence-based prediction of cofactor specificity of Rossmann folds**. *Proteins: Structure, Function, and Bioinformatics* 2014, **82**(9):1819-1828.

Appendices

Appendix A: Supplementary data for chapter 2

Table S2.1: Taxonomic classification of strains predicted to contain either Bgl, Sal or Lac system.
45 strains predicted to possess either Bgl, Sal or Lac systems are presented along with their respective phylum, class, order, family and genus. The strains are color-coded by the following predicted systems: grey (Bgl), orange (Lac), blue (Sal) and yellow (both Lac and Sal).

Strain Name	Strain ID	Phylum	Class	Order	Family
Pantoea sp. YR512	2690315618	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Erwiniaceae*
Rahnella aquatilis OV744	2585427591	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Yersiniaceae*
Pantoea sp. OV426	2690315639	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Erwiniaceae*
Paenibacillus sp. OV031	2616644976	*Firmicutes*	*Bacilli*	*Bacillales*	*Paenibacillaceae*
Paenibacillus sp. CF095	2690315662	*Firmicutes*	*Bacilli*	*Bacillales*	*Paenibacillaceae*
Paenibacillus sp. OK003	2687453683	*Firmicutes*	*Bacilli*	*Bacillales*	*Paenibacillaceae*
Pantoea sp. YR525	2690315619	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Erwiniaceae*
Paenibacillus sp. PDC88	2693429827	*Firmicutes*	*Bacilli*	*Bacillales*	*Paenibacillaceae*
Microbacterium sp. CF335	2585428157	*Actinobacteria*	*Actinobacteria*	*Micrococcales*	*Microbacteriaceae*
Pantoea sp. GM01	2511231035	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Erwiniaceae*
Bacillus sp. OV752	2738541358	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*
Paenibacillus sp. OK060	2684623002	*Firmicutes*	*Bacilli*	*Bacillales*	*Paenibacillaceae*

Table S2.1 continued.

Strain Name	Strain ID	Phylum	Class	Order	Family
Enterobacter sp. OV724	2758568622	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Enterobacteriaceae*
Pantoea sp. YR343	2511231025	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Erwiniaceae*
Paenibacillus sp. OK076	2690315677	*Firmicutes*	*Bacilli*	*Bacillales*	*Paenibacillaceae*
Pleomorphomonas sp. CF100	2738543031	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Methylocystaceae*
Bacillus sp. OK048	2687453680	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*
Rahnella aquatilis OV588	2585427592	*Proteobacteria*	*Gammaproteobacteria*	*Enterobacterales*	*Yersiniaceae*
Bacillus sp. OK077	2738543006	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*
Bacillus sp. YR335	2738543010	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*
Microbacterium sp. CF332	2609459766	*Actinobacteria*	*Actinobacteria*	*Micrococcales*	*Microbacteriaceae*
Bacillus sp. YR331	2615840594	*Firmicutes*	*Bacilli*	*Bacillales*	*Bacillaceae*
Rhizobium sp. PDC82	2690315628	*Proteobacteria*	*Alphaproteobacteria*	*Rhizobiales*	*Rhizobiaceae*
Sphingomonas sp. OK281	2690315632	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Caulobacter sp. AP07	2510917020	*Proteobacteria*	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*
Caulobacter henricii CF287	2582581280	*Proteobacteria*	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*

Table S2.1 continued.

Strain Name	Strain ID	Phylum	Class	Order	Family
Caulobacter sp. OV484	2585428106	*Proteobacteria*	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*
Caulobacter henricii YR570	2582581293	*Proteobacteria*	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*
Sphingobium sp. YR768	2687453792	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Sphingobium sp. AP50	2687453790	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Novosphingobium sp. GV064	2738543033	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Caulobacter henricii OK261	2582581279	*Proteobacteria*	*Alphaproteobacteria*	*Caulobacterales*	*Caulobacteraceae*
Sphingobium sp. AP49	2512564014	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Novosphingobium sp. GV027	2738541275	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Novosphingobium sp. GV022	2757320419	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Novosphingobium sp. GV057	2757320435	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Novosphingobium sp. GV079	2738541301	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Novosphingobium sp. GV011	2757320424	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Novosphingobium sp. GV055	2738543022	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Novosphingobium sp. GV026	2757320425	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*

Table S2.1 continued.

Strain Name	Strain ID	Phylum	Class	Order	Family
Novosphingobium sp. GV078	2757320426	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Novosphingobium sp. GV070	2757320420	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Novosphingobium sp. AP12	2510917021	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Novosphingobium sp. GV061	2738541304	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*
Novosphingobium sp. GV076	2757320421	*Proteobacteria*	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*

Table S2.2: Genes and their neighboring sequences that might have been mutated for *Citrobacter* to have acquired the ability to grow on salicin.
Citrobacter sp. YR018 possesses nine GH1 and GH3 genes, but only Ga0215739_10977, Ga0215739_101543 and Ga0215739_101614 had either *bglF*/PTS or *salC* genes neighboring them. Out of these three, Ga0215739_10977 appears to be the gene that should be investigated first because it belongs to a gene cluster that is strikingly similar to *bgl* operon. This cluster was not identified due to criteria applied in our analysis. Mutations in the genes or the promoter regions of this probable operon might be responsible for the gain of function in *Citrobacter.*

Genes	GH1/GH3	Nearby genes	Percent identity
Ga0215739_10977	***bglB***	**PTS/*bglF***	**~ 54.89**
Ga0215739_10295	***bglB***	**None**	**~ 52.52**
Ga0215739_11458	***bglB***	**None**	**~ 50.73**
Ga0215739_103356	***bglB***	**None**	**~ 36.13**
Ga0215739_101543	***bglB***	**PTS**	**~ 38**
Ga0215739_109164	***bglB***	**None**	**~ 31.30**
Ga0215739_102346	***salA/salB/cbg1***	**None**	**~ 35.02**
Ga0215739_101614	***salA/salB/cbg1***	***salC***	**~ 36.66**
Ga0215739_10393	***cbg1***	**NA**	**~ 27**

Table S2.3: Three strains that could likely possess both Sal and Lac systems other than the ones already identified in the study.
In this analysis, each organism containing either systems were manually checked to determine whether at least one of the genes of the other system (i.e. Lac system for the ones with Sal system and vice versa) is present neighboring the reference genes (*sal*A or *lac*A). Two strains (excluding *Caulobacter* sp. AP07) could contain both systems other than the ones already discussed in the chapter.

Strain Name	**lacA**	**lacB**	**lacC**	**salA**	**salB**	**salC**	**salR**	**Predicted salicin degradation systems**	**At least one significant neighboring gene found**
Caulobacter sp. AP07	YES	YES	YES	YES	YES	YES	YES	LacABC	Yes
Sphingomonas sp. OK281	YES	YES	YES	YES	YES	YES	YES	LacABC	Yes
Sphingobium sp. AP50	YES	YES	YES	YES	YES	YES	YES	SalRCAB	Yes

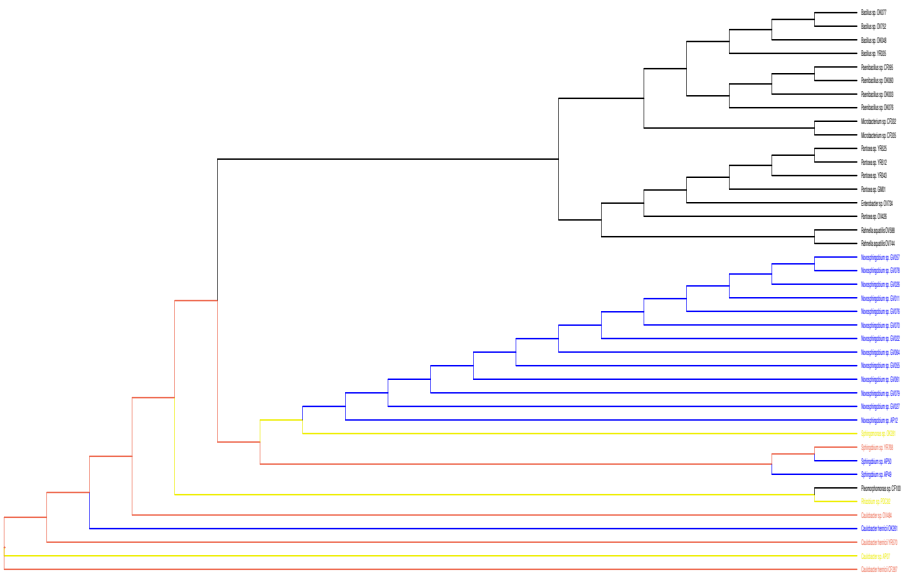


Figure S2.1: Phylogenetic distribution of salicin degradation pathways of 45 strains predicted to contain either Bgl, Sal or Lac system.
A distance tree was created using JGI/IMG distance tree tool, and then modified using ggTree package (v. 1.14.6) in R (v. 3.5.1). Color code: Bgl – Black, Sal – Blue, Yellow – Lac and Orange – Sal + Lac.

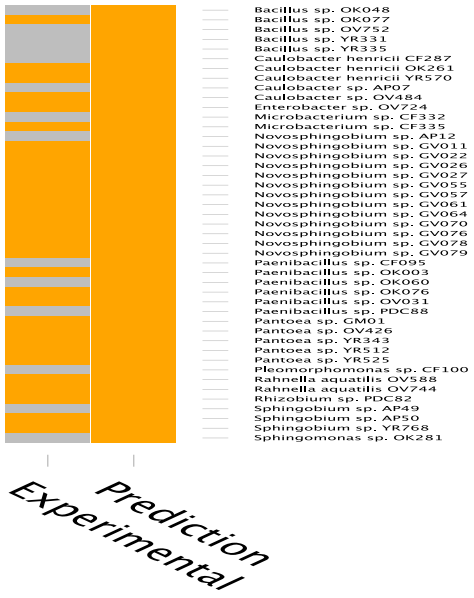
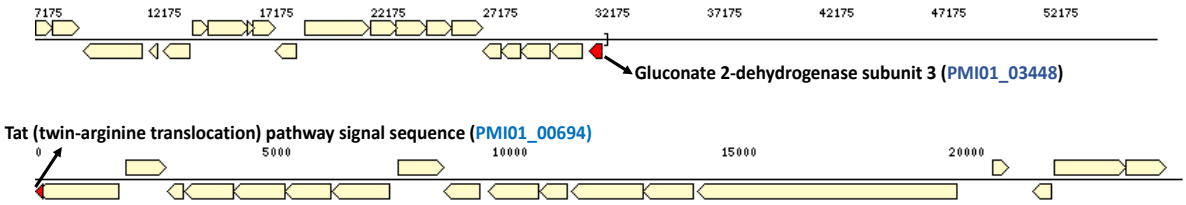


Figure S2.2: Comparison of prediction and experimental results between the strains expected to comprise either Bgl, Sal or Lac systems.
All 45 strains that possessed either of the three systems were compared with the experimental results. 31 matches were found, and out of 14 mismatches, at least one (*Caulobacter* sp. AP07) is expected to grow on salicin.



MLSRRDAIRGLTLTIGVAATGWGGRATGWAGRAAAQTLTWTPTALAPDQARLVDVVAELIIPATDTPGAREAGVPQFIDRALAN
YYDKAQKERLLGGLARMDADARAAHGDVFVALTPQQQVELLTGYDREASAVRNETPDDAHFFLVLENLVTVGYFTSEPGATVAL
RYDPIPGAYHGCVPLSELGRAWATR

Figure S2.3: *Caulobacter* possesses a complete *lacC* homolog.
It is not apparent that *Caulobacter* contains a complete *lacC* possibly because it is divided by two scaffolds due to sequencing and annotation limitations. The sequences were concatenated and translated using Expasy translation tool. The amino acid sequence of concatenated gene is provided here. The *lacC* protein sequences of *Caulobacter* strains were aligned using blastp and were found to be ~ 65% identical at 100% query coverage.

YR018 vs YR018-SAL sequence alignment

Alignment of trimmed nucleotide sequences (orange) of YR018 and YR018-SAL 16S genes shows that both sequences are > 99% identical suggesting both strains are same.

```
YR018_R_PREMI       1 --------------------------------------------------       0

YR018SAL_R_PR       1 NNNNNNNNNNNNNNNNNNNNNNNNNNNANNNTNCNNNNNNNNNNNNNNAANNN      50

YR018_R_PREMI       1 -------NNNNNNNNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNN      43

YR018SAL_R_PR      51 NNNNNNN---------------------------------------------      57

YR018_R_PREMI      44 AANNNNNNNNNNACTTCTTTTGCAACCCACTCCCATGGTGTGACGGGCGG      93
                                  ||||||||||||||||||||||||||||||||||||
YR018SAL_R_PR      58 -------------CTTCTTTTGCAACCCACTCCCATGGTGTGACGGGCGG      94

YR018_R_PREMI      94 TGTGTACAAGGCCCGGGAACGTATTCACCGTAGCATTCTGATCTACGATT     143
                      |||||||||||||||||||||||||||||||||||||||||||||||||||
YR018SAL_R_PR      95 TGTGTACAAGGCCCGGGAACGTATTCACCGTAGCATTCTGATCTACGATT     144

YR018_R_PREMI     144 ACTAGCGATTCCGACTTCATGGAGTCGAGTTGCAGACTCCAATCCGGACT     193
                      |||||||||||||||||||||||||||||||||||||||||||||||||||
YR018SAL_R_PR     145 ACTAGCGATTCCGACTTCATGGAGTCGAGTTGCAGACTCCAATCCGGACT     194

YR018_R_PREMI     194 ACGACGCACTTTATGAGGTCCGCTTGCTCTCGCGAGGTCGCTTCTCTTTG     243
                      |||||||||||||||||||||||||||||||||||||||||||||||||||
YR018SAL_R_PR     195 ACGACGCACTTTATGAGGTCCGCTTGCTCTCGCGAGGTCGCTTCTCTTTG     244

YR018_R_PREMI     244 TATGCGCCATTGTAGCACGTGTGTAGCCCTACTCGTAAGGGCCATGATGA     293
                      |||||||||||||||||||||||||||||||||||||||||||||||||||
YR018SAL_R_PR     245 TATGCGCCATTGTAGCACGTGTGTAGCCCTACTCGTAAGGGCCATGATGA     294
```

```
YR018_R_PREMI    294 CTTGACGTCATCCCCACCTTCCTCCAGTTTATCACTGGCAGTCTCCTTTG    343
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
YR018SAL_R_PR    295 CTTGACGTCATCCCCACCTTCCTCCAGTTTATCACTGGCAGTCTCCTTTG    344

YR018_R_PREMI    344 AGTTCCCGGCCGAACCGCTGGCAACAAAGGATAAGGGTTGCGCTCGTTGC    393
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
YR018SAL_R_PR    345 AGTTCCCGGCCGAACCGCTGGCAACAAAGGATAAGGGTTGCGCTCGTTGC    394

YR018_R_PREMI    394 GGGACTTAACCCAACATTTCACAACACGAGCTGACGACAGCCATGCAGCA    443
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
YR018SAL_R_PR    395 GGGACTTAACCCAACATTTCACAACACGAGCTGACGACAGCCATGCAGCA    444

YR018_R_PREMI    444 CCTGTCTCAGAGTTCCCGAAGGCACCAAAGCATCTCTGCTAAGTTCTCTG    493
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
YR018SAL_R_PR    445 CCTGTCTCAGAGTTCCCGAAGGCACCAAAGCATCTCTGCTAAGTTCTCTG    494

YR018_R_PREMI    494 GATGTCAAGAGTAGGTAAGGTTCTTCGCGTTGCATCGAATTAAACCACAT    543
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
YR018SAL_R_PR    495 GATGTCAAGAGTAGGTAAGGTTCTTCGCGTTGCATCGAATTAAACCACAT    544

YR018_R_PREMI    544 GCTCCACCGCTTGTGCGGGCCCCCGTCAATTCATTTGAGTTTTAACCTTG    593
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
YR018SAL_R_PR    545 GCTCCACCGCTTGTGCGGGCCCCCGTCAATTCATTTGAGTTTTAACCTTG    594

YR018_R_PREMI    594 CGGCCGTACTCCCCAGGCGGTCGACTTAACGCGTTAGCTCCNGAAGCCAC    643
                     |||||||||||||||||||||||||||||||||||||||| ||||||||
YR018SAL_R_PR    595 CGGCCGTACTCCCCAGGCGGTCGACTTAACGCGTTAGCTCCGGAAGCCAC    644

YR018_R_PREMI    644 GCCTCAAGGGCACAACCTCCAAGTCGACATCGTTTACGGCGTGGACTACC    693
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
YR018SAL_R_PR    645 GCCTCAAGGGCACAACCTCCAAGTCGACATCGTTTACGGCGTGGACTACC    694
```

```
YR018_R_PREMI    694 ACGGNNTCTAATCCTGTTTGCTNCCCACGCTNTNNNACCTGAGCGTCAGT    743
                     |.||  ||||||||||||||| |||||||| |   ||||||||||||||
YR018SAL_R_PR    695 AGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGCACCTGAGCGTCAGT    744

YR018_R_PREMI    744 CTTTGTNNNNNNNGGGNNGCCTTCNNNACCNGNANTCCTCCAGATNNATAC    793
                     ||||||      |||  ||||||   ||| | | ||||||||||  .|||
YR018SAL_R_PR    745 CTTTGT-CCAGGGGGCCGCCTTCGCCACCGGTATTCCTCCAGATCTCTAC    793

YR018_R_PREMI    794 GCANNTCATCNCTANNCCNNNNAGNNNNNCCACCTCNNCAAGANNNNANC    843
                     |||  |||.| |||  ||    |.     ||.||||  |||||    | |
YR018SAL_R_PR    794 GCATTTCACCGCTACACCTGGAATTCTACCCCCCTCTACAAGACTCTAGC    843

YR018_R_PREMI    844 CTGCCNGTNTCNNNTGCANNNCCCNNNCTGNGNNNNNNNNNNNNNATNTCC    893
                     |||||  || ||   ||||   |||   .|| |           || ||.
YR018SAL_R_PR    844 CTGCCAGTTTCGAATGCAGTTCCCAGGTTGAG-----CCCGGGGATGTCA    888

YR018_R_PREMI    894 CACCTGA-NTGANNNANNNCCTG-----GNNNNNNCTTCNNGCCAATTAA    937
                     ||.|.||  |||   |   ||||     |      |||.  |||.|.|||
YR018SAL_R_PR    889 CATCCGACTTGACAGACCGCCTGCGTGCG------CTTTACGCCCAGTAA    932

YR018_R_PREMI    938 -----ATTAAANNNAGNNNNNGCNCCATNNNNNNNCNCGNGGGNNNGGNN    982
                         |||||                          ||
YR018SAL_R_PR    933 TTCCGATTAA--------------------------CG-----------    944

YR018_R_PREMI    983 NNANNNNNNNGCNNNNNNNNNNNNNNNNNNNNAAANNNNNNNNGGTTNNNN   1032

YR018SAL_R_PR    945 ------------------------------------------------    944

YR018_R_PREMI   1033 NNNNNAANNNCNNNCCCCCNNCTCCNNNNAAANTAGANNNNNAANNNNCC   1082
                           |   |.||   ||||
YR018SAL_R_PR    945 ----------CTTGCACC---CTCC-------------------------    956
```

```
YR018_R_PREMI   1083 GNNNGGGTTTTTTTTNGCCGGGGGGNNGNNGG------GTTNNNGGGT--   1124
                        ||.||     .||  |.||  |  ||      |||   .|||
YR018SAL_R_PR    957 ------GTATT-----ACC--GCGGCTGCTGGCACGGAGTTAGCCGGTGC    993

YR018_R_PREMI   1125 --------------------------------------------------   1124

YR018SAL_R_PR    994 TTCTTCTGCGAGTAACGTCAATCGCTGCGGTTATTAACCACAACGCCTTC   1043

YR018_R_PREMI   1125 ------------------------------   1124

YR018SAL_R_PR   1044 CTCCTCGCTGAAAGTACTTTACANNNCGAAG   1074
```

Appendix B: Supplementary data for chapter 3

Table S3.1: Genes used as queries for blastp.
Genes representing proteins of either GH1 or GH3 family were used as queries against *Agrobacterium.* The genes belong to organisms *Escherichia coli* K-12 MG1655 (*bglB*), *Azospirillum irakense* DSM 11586 (*sal* genes) and *Rhizobium radiobacter* (*cbg1*).

Genes	Homolog
Ga0063390_02056	*bglB* (GH1)
P27034	*cbg1* (GH3)
G473DRAFT_2580	*salA* (GH3)
G473DRAFT_2579	*salB* (GH3)

Table S3.2: Genes of the *Agrobacterium rhizogenes* OK036 cluster were some of the top hits when protein sequences of *Streptomyces* salicylate pathway genes were used as queries in blastp.
Amino acid sequences of salicylate cluster genes were used as queries against *Agrobacterium rhizogenes* OK036 genome, and many of the genes of the proposed salicin cluster were matched with percent identity ranging from 24% – 68%. Many of them were best or only hits for specific genes. EW94DRAFT_05422 matched with two different genes at two different termini of its entire length indicating dual function which has been addressed in the chapter.

***Streptomyces* genes**	***Agrobacterium* genes**	**Best Hit**	**Percent Identity**	**>70% Query Coverage**
***sdg*A**	EW94DRAFT_05422	No	24%	No
***sdg*B**	EW94DRAFT_05927	Yes	28%	Yes
***sdg*C**	EW94DRAFT_05422	No	38%	No
***sdg*D**	EW94DRAFT_05425	Yes	62%	Yes
***sdg*F**	EW94DRAFT_05426	Yes	26%	Yes

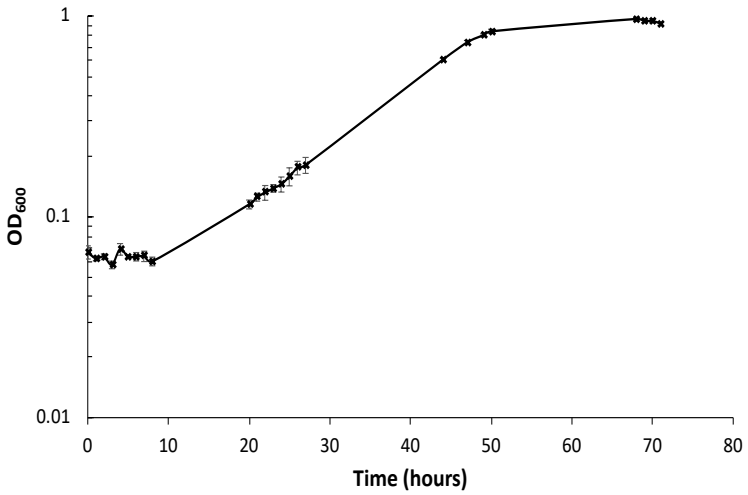


Figure S3.1: *Agrobacterium* grows on salicin to an OD_{600} close to 1.
Agrobacterium grew on MOPS minimal media containing salicin as sole carbon source and reached an OD close to 1. The OD is the highest ever observed OD_{600} in any organism that we have worked on so far for the substrate concentration of 3 mM.

Blastp alignment result of BH1999 (gentisyl-coenzyme A thioesterase) from *Bacillus halodurans* C-125 against EW94DRAFT_05421 of *Agrobacterium rhizogenes* OK036 (percent identity: 32%, query coverage: ~ 79%)

```
Query  7    HFRVKFGDTDAAGIVFYPNYYKWMDEACHHFLTELGFPTSELIDK--KIGFPIVEATCQF  64
            +V FGD D AGI FYPN ++W+D+A H +L   G   ++L  +   IG  ++EA  QF
Sbjct  3    QIQVSFGDCDPAGIAFYPNIFRWLDKAFHDWLRPRG-GHADLCARLDAIGIGLIEAKAQF  61

Query  65   KAPLLFADHVFIRTSIRELKDKSFILEHHFIKQGRVIASGHEKRVWANFSNG  116
            + P++  D + +  +++E   K+  LE+     G +     E R     +NG
Sbjct  62   RRPIVDGDTLVVDLTVKEWGRKTVTLEYEGRVDGVLAFQAQEIRGLFKRNNG  113
```

Calculation of ODs on catechol

Catechol minimal media (without any cell, 3 replicates) was incubated at 25°C with gentle shaking for 48 hours. For the time duration used, the growth plot is relatively linear. The OD_{600} was measured at different times points and plotted on a xy axis where x = time and y = OD_{600}. A best-fit line ($r^2 = 0.99$) was determined for the plot with linear equation –

$$y = 0.0067x + 0.0534$$

Using this formula, Catechol-OD_{600} was determined for each time point and this value could be subtracted from the *Agrobacterium* catechol culture (OK036-OD_{600}) for the specific time point in order to calculate the OD_{600} of OK036 strain for that time point. The difference OK036-OD_{600} - Catechol-OD_{600} is plotted on the graph shown in the figure 3.3.

Appendix C: Supplementary data for chapter 4

Table S4.1: Probable salicyl alcohol exporters in *Rahnella* sp. OV744.
Protein sequences of genes that are involved in the export of compounds related to salicyl alcohol such as toluene were used as query against *Rahnella* sp. OV744 genome. The gene hits were matched with proteome data to list the proteins that might be involved in salicyl alcohol secretion.

Locus tag	Description	Detected in glucose	Detected in salicin monoculture	Detected in salicin coculture
EX29DRAFT_00286	RND family efflux transporter, MFP subunit	Yes	No	No
EX29DRAFT_01426	RND family efflux transporter, MFP subunit	Yes	Yes	Yes
EX29DRAFT_00133	RND family efflux transporter, MFP subunit	No	Yes	Yes
EX29DRAFT_05030	RND family efflux transporter, MFP subunit	No	Yes	No
EX29DRAFT_03845	RND family efflux transporter, MFP subunit	Yes	Yes	No
EX29DRAFT_03845	RND family efflux transporter, MFP subunit	Yes	Yes	No
EX29DRAFT_01425	The (Largely Gram-negative Bacterial) Hydrophobe/Amphiphile Efflux-1 (HAE1) Family	Yes	Yes	Yes
EX29DRAFT_05031	heavy metal efflux pump, CzcA family	No	Yes	No
EX29DRAFT_00690	type I secretion outer membrane protein, TolC family	Yes	Yes	Yes
EX29DRAFT_05023	efflux transporter, outer membrane factor (OMF) lipoprotein, NodT family	Yes	Yes	Yes
EX29DRAFT_02794	The (Largely Gram-negative Bacterial) Hydrophobe/Amphiphile Efflux-1 (HAE1) Family	No	No	Yes
EX29DRAFT_05031	heavy metal efflux pump, CzcA family	Yes	No	No
EX29DRAFT_05023	efflux transporter, outer membrane factor (OMF) lipoprotein, NodT family	Yes	Yes	Yes
EX29DRAFT_03848	efflux transporter, outer membrane factor (OMF) lipoprotein, NodT family	Yes	No	No
EX29DRAFT_03777	efflux transporter, outer membrane factor (OMF) lipoprotein, NodT family	No	No	Yes

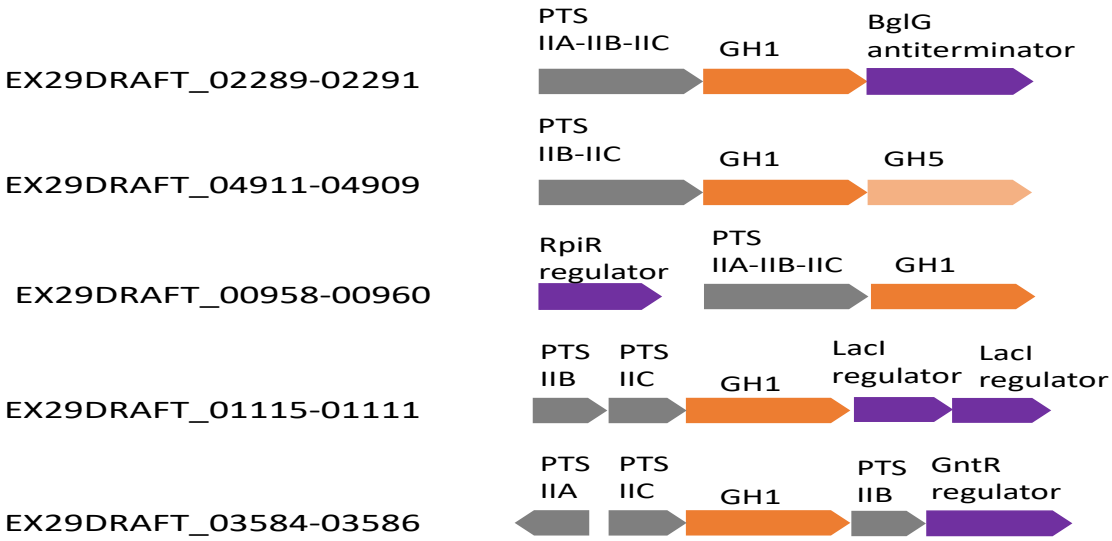


Figure S4.1: Possible GH1 enzyme encoding genes and their clusters in *Rahnella* sp. OV744.
Rahnella possesses five glycosyl hydrolase family 1 (GH1) homologs. Each gene is neighbored by phosphotransferase (PTS) based transporters and regulators of different types. Based on the cluster structure, EX29DRAFT_02289 - EX29DRAFT_02291 was proposed to be involved in salicin degradation because it resembled *bgl* operon which has been demonstrated to be involved in salicin degradation in multiple strains of *Enterobacteriaceae* family [1-4].

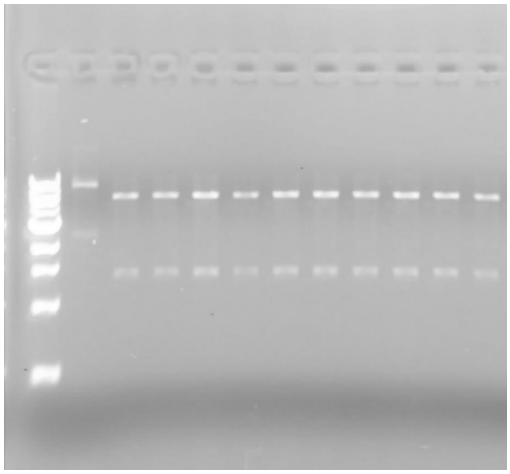


Figure S4.2: Gel electrophoresis of pSRK.km_OVCompYR digested with SacI and KpnI.
pSRK.km was ligated to create pSRK.km_OVCompYR which was transformed to NEB 10β competent cells. Following transformation and selection, colonies were picked from the LB plates containing kanamycin antibiotic and inoculated in LB liquid medium (with kanamycin). Cultures were grown overnight and then plasmid was extracted. Plasmid was then digested using SacI and KpnI enzymes followed by gel electrophoresis for visualization. In the picture, 1st band is for ladder, second band is uncut pSRK.km whereas others are double digested pSRK.km_OVCompYR. All transformants appear to contain the desired plasmid. One of them was chosen for complementation assay.

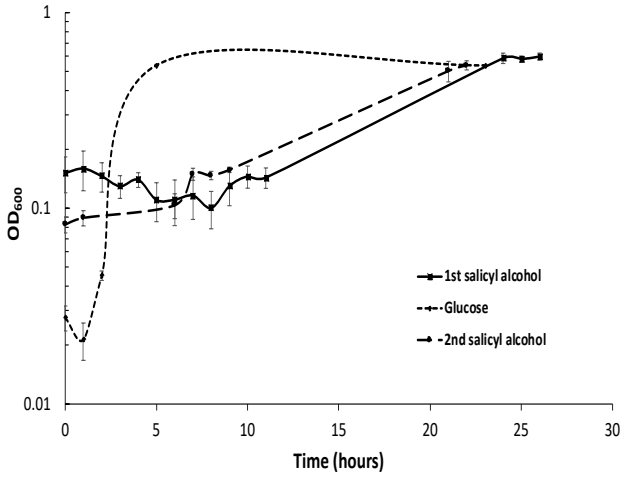


Figure S4.3: Growth delay of *Pseudomonas* sp. GM16 in salicyl alcohol is not due to emergence of mutants.
GM16 cells when grown on salicyl alcohol minimal medium demonstrate a long lag time before the growth picks up. To determine whether this is due to physiological adaptation or mutation, first cells were incubated in salicyl alcohol minimal medium followed by subculture in glucose minimal medium and subsequent subculture in salicyl alcohol minimal medium. OD_{600}s were read for each culture. The lag time observed in the first salicyl alcohol culture persisted even in the second culture in salicyl alcohol. Therefore, this lag in growth is most likely due to cells adapting to the substrate.

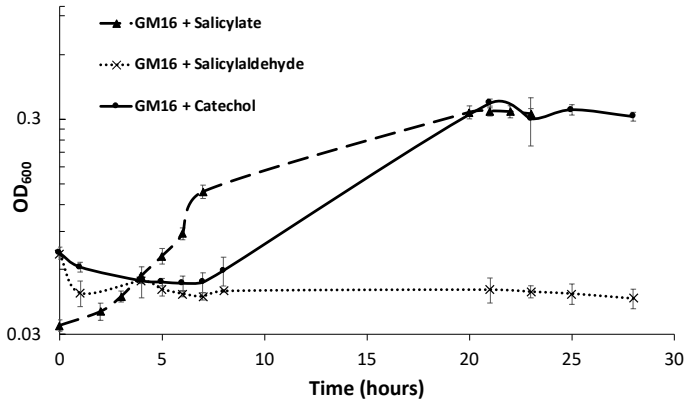


Figure S4.4: Growth curve analysis of *Pseudomonas* sp. GM16 on key intermediates of proposed pathway.
GM16 was subcultured on MOPS minimal media containing three intermediates – salicylaldehyde, salicylate and catechol of the proposed salicyl alcohol pathway. GM16 was able to grow on salicylate and catechol minimal media (see supplementary text for calculations required for catechol growth determination). Furthermore, for catechol, cell counting was also performed, and cell numbers were observed to increase over time (data not shown).

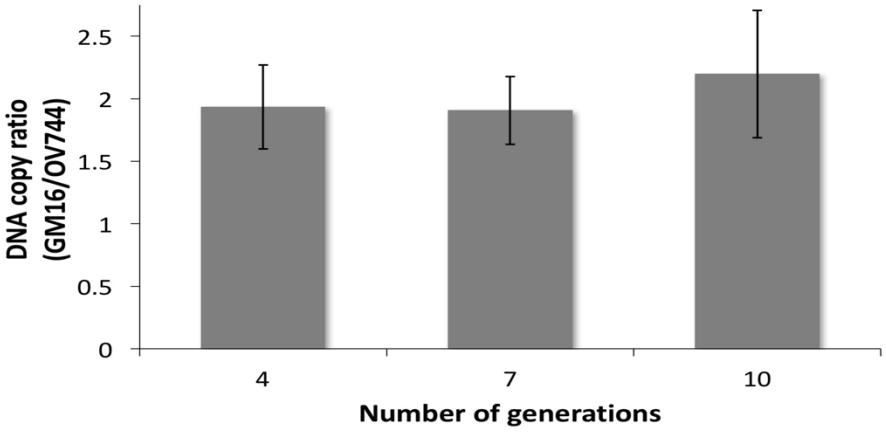


Figure S4.5: Stable state of the co-culture demonstrates that *Pseudomonas* sp. GM16 and *Rahnella* sp. OV744 grow in the ratio of ~ 2:1.
GM16 and OV744 were grown in salicin co-culture overnight. The overnight culture was diluted in the ratio 1:10 in fresh salicin minimal medium and grown overnight again. The dilution-growth process was repeated over several days. For each dilution, equal volumes (1 ml) of samples were harvested. Total DNA was isolated and then qPCR was performed using strain-specific primers. The qPCR data was converted to DNA copy number using in-house protocol (see supplementary text). The generation number was calculated using population doubling times for the co-culture to reach stationary phase. For ten generations, the ratio of DNA copy count of GM16 and OV744 was approximately 2:1 suggesting *Pseudomonas* takes over the growth of co-culture at certain time point.

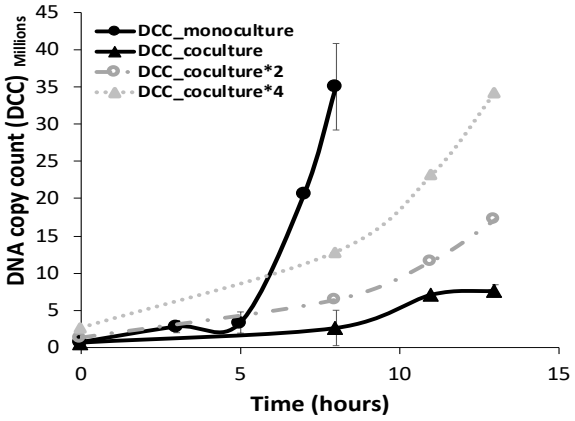


Figure S4.6: Growth of *Rahnella* sp. OV744 is likely affected by the growth of *Pseudomonas* sp. GM16 in co-culture.
A time-course qPCR was used to visualize the growth of *Rahnella* in both salicin co-culture and monoculture using the same set of primers. The DNA copy count (DCC) increased over time suggesting growth of the strain. As suggested in figure 4.3c and S4.5, the growth of OV744 appears to be inhibited by the presence of *Pseudomonas*. In the figure, the dark lines are the *Rahnella* DCC curves from the monoculture and co-culture samples. The grey lines represent hypothetical DCC curves if the original co-culture DCC numbers were multiplied by 2 and 4. As suggested in the figure, DCC of OV744 was significantly lower in co-culture compared to monoculture even when original DCC is multiplied by 2, but not significant if original DCC is multiplied by 4 (see supplementary text for calculation protocol).

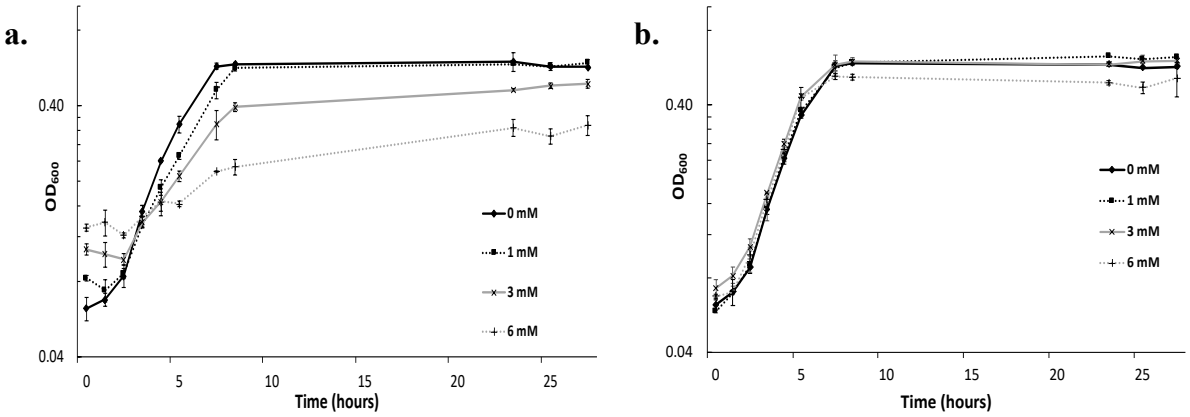


Figure S4.7: Salicyl alcohol (SalOH) does not significantly inhibit the growth of *Rahnella* sp. OV744.
a) OV744 was grown on glucose minimal media, and different concentrations (0, 1, 3 and 6 mM) of salicyl alcohol (SalOH) were added at the beginning of the culture. No significant differences between the salicyl alcohol treatments were found up to 3 mM. Minor growth inhibition at 6 mM salicyl alcohol was observed. b) Similarly, SalOH was added in the early exponential phase during growth in glucose minimal media. Same as before, no significant differences between the treatments were found.

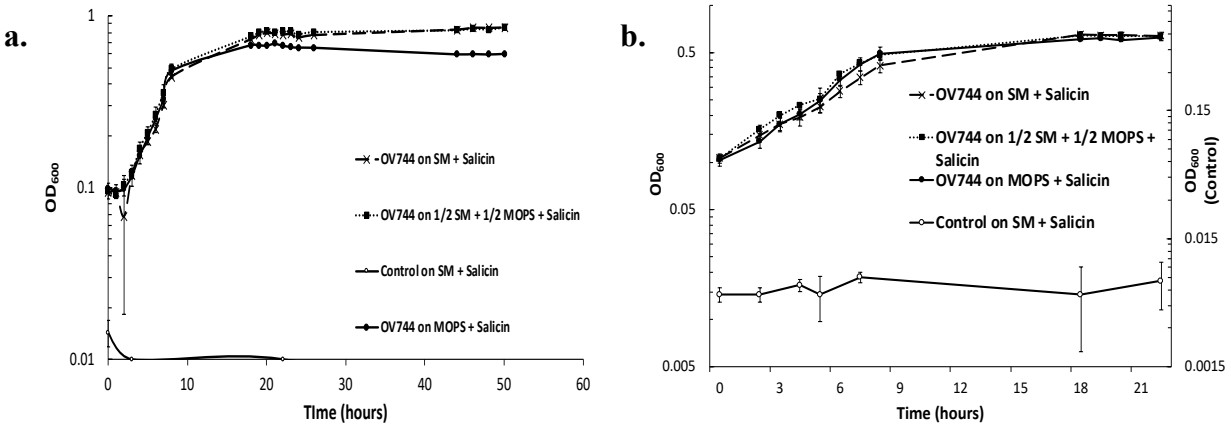


Figure S4.8: *Pseudomonas* sp. GM16 secretion does not affect the growth of *Rahnella* sp. OV744.
In both pictures, control refers to no cell negative control. a) GM16 grown salicyl alcohol spent media (SM) was mixed with MOPS minimal media or distilled water. Salicin was added as the carbon substrate. Then, these growth media were inoculated with OV744 cell pellets. No significant differences in growth could be observed between different growth conditions. In fact, OV744 cultures grew to higher OD in media containing SM than those in MOPS salicin media only (positive control). b) SM of OV744-GM16 salicin co-culture was prepared and mixed with MOPS minimal media or distilled water followed by addition of salicin. OV744 cells were subcultured in these growth media. In these experiments, no significant differences were observed between different growth conditions. In the figure, secondary axis was required to visualize the growth curve of negative control.

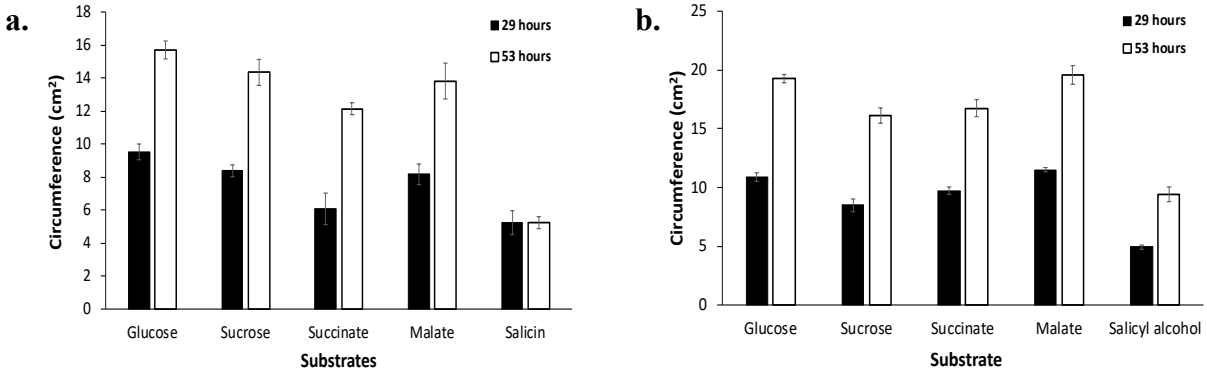


Figure S4.9: Chemotaxis assays demonstrate *Pseudomonas* sp. GM16 is attracted towards salicyl alcohol.
For this experiment, soft agar (0.3%) MOPS plates were prepared with different substrates at concentration of 1 mM. 2-3 µl of cell cultures of *Rahnella* and *Pseudomonas* were spotted at the center of the plates and the plates were incubated at 25°C. The circumference of the spot was calculated at 29 hours and 53 hours of incubation. a) OV744 appears to respond to all five substrates but seems to have low motility towards salicin. Salicin plates were cloudy and exact reason for this is unknown, and that could be the reason for low motility. b) *Pseudomonas* demonstrated attraction towards all substrates including salicyl alcohol as indicated by the increase in circumference over time.

Chemotaxis assay

First of all, soft agar (0.3%) 3-(N-morpholino)propanesulfonic acid (MOPS) [94] plates were prepared as described elsewhere [214]. Desired substrates (glucose, sucrose, succinate, malate, salicin, salicyl alcohol) at concentration 1 mM were added right before the plates were prepared. Then on fresh plates, cell cultures of *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16 were spotted with culture volume of 2-3 µl spotted at the center of the plates. The plates were incubated at 25ºC for more than two days (53 hours). The circumference of the spot was calculated at two different time points – 29 hours and 53 hours to elucidate the motility (or lack thereof) of strains towards the substrates.

Calculation of ODs on catechol

Catechol minimal media was incubated at 25ºC with gentle shaking and at different time points, OD_{600} was measured. The ODs were plotted, and a best fit line was determined. For the time duration used, the increase in OD_{600} occurs in a linear fashion. This line had $r^2 = 0.99$. Its linear equation is:

$y = 0.0067x + 0.0534$ where $y = OD_{600}$, x = time

Using this formula, Catechol-OD_{600} could be calculated for each time point. ODs for *Pseudomonas* catechol culture (GM16-OD_{600}) were noted and for every time point, the GM16-OD_{600} – Catechol-OD_{600} was calculated. This difference was then plotted on the graph shown in figure S4.4.

Calculation of gene copy count in qPCR method (see attachments S4.5, S4.6 and S4.7 for example)

First of all, starting concentration was calculated by multiplying starting quantity value by 10 to account for dilution factor. Let this number be $X_{initial}$. Molecular weight (MW) was calculated using the formula:

$$MW = N_{base} * (1 - GC) * (MW_{dATP} + MW_{dCTP}) + N_{base} * GC * (MW_{dGTP} + MW_{dTTP})$$

where, N_{base} is the size of the genome in base pairs

GC is the GC% of the genome

MW_{dATP}/ MW_{dTTP}/ MW_{dGTP}/ MW_{dCTP} is the molecular weight of the bases.

After this, following set of calculations were performed:

(concentration in g/µl) $X_{g/\mu l} = X_{initial} * 10^{-9}$

(concentration in mol/µl) $X_{mol/\mu l} = X_{g/\mu l}/MW$

(DNA copy count) $X_{DCC} = X_{mol/\mu l} * 6.022 * 10^{23} * 2$

Finally, means and standard deviation of each of the samples were calculated.

Calculation of hypothetical DNA copy count (DCC) for comparing *Rahnella* co-culture and monoculture copy counts

First, the starting DCC of *Rahnella* in co-culture was doubled or quadrupled. This number was then used for calculation of other DCC at different time points (N_t) using the formula:

$N_t = N_0 e^{rt}$ where N_0 is starting DCC, r = growth rate and t = time point

In this case, r = 0.196 hr^{-1} which is the growth rate of the co-culture.

Appendix D: Other works I have been part of during my PhD

Exploring complex cellular phenotypes and model-guided strain design with a novel genome-scale metabolic model of *Clostridium thermocellum* DSM 1313 implementing an adjustable cellulosome

Published in Biotechnology for Biofuels in September 2016

Authors: Thompson, R. Adam, ***Dahal, Sanjeev****, Garcia, Sergio, Nookaew, Intawat and Trinh Cong T.*

The genome-scale metabolic model of *Clostridium thermocellum* DSM 1313 was based and improved upon the model of *Clostridium thermocellum* ATCC 27405. This model consists of 872 reactions with 114 transport and exchange reactions. Furthermore, the model incorporates the cost of synthesis of cellulosome. The model was validated using multiple experimental growth data and then was used to predict strain engineering strategies to increase production of biofuels.

In this study, I was involved in refining the model by identifying the putative transporters which were later manually inspected. The list of transporters was created using three different methods of which InterProScan 5 [215] was the first. Then, genes encoding transporters from validated *Clostridium* models were extracted and subjected to reciprocal blast hit (RBH) and hmmscan [216] to identify respective orthologs in DSM 1313 strain. Finally, Transporter Substrate Database [217] was used to extract protein sequences of transporters for each substrate exchange reaction within the model. Those sequences were compared to the genome of DSM 1313 using RBH and hmmscan.

Next, I used the program Cofactory to analyze the cofactor specificity of reactions to

further refine the model [218]. This program identifies the cofactor of an enzyme by predicting Rossman folds from the amino acid sequence. Reactions that were predicted to contain different cofactors in the automated model were consolidated using the Cofactory results.

Genome-Scale Modeling of Thermophilic Microorganisms

Published in Advances in Biochemical Engineering/Biotechnology in December 2016

Authors: ***Dahal, Sanjeev****, Poudel, Suresh and Thompson, R. Adam.*

In the area of biofuel technology, thermophilic organisms are generating immense interest because their metabolism is optimized for higher temperatures. Even though this is the case, to engineer these strains to enhance biofuel production is a challenging task. The metabolic processes of these organisms can differ vastly from those of mesophilic counterparts. Therefore, it is imperative to identify those differences before the experiments can be performed on these organisms. One effective way is to create metabolic models to investigate the metabolic pathways. Then these models can be simulated to validate and to predict effective metabolic engineering strategies. Numerous metabolic models have been created, but the details about these models have not been compiled into one comprehensive article. Therefore, this chapter was written as a literature review describing all the known metabolic models of thermophilic organisms ranging from bacteria to archaea. The review elucidates the similarities and differences in the metabolism of these thermophiles as interpreted from the studies involving the respective organisms. This article can be used as a guide by researchers that want to work on thermophilic organisms.

Vita

Sanjeev Dahal was born in Lalitpur, Nepal to Ram Sharan Dahal (father) and Bhagawati Simkhada Dahal (mother). He went to Adarsha Vidya Mandir (Lalitpur) and United Academy (Lalitpur) for primary/middle and high school, respectively. For undergraduate degree, he applied to and got admitted in the department of Biological Sciences at the University of New Orleans (Louisiana, USA). During his undergraduate schooling, he conducted research under the tutelage of Dr. Mary J. Clancy. After successful completion of undergraduate education, he joined the PhD program in Genome Science and Technology at the University of Tennessee, Knoxville in August 2013.