October 2019

BACTERIAL BIOTRANSFORMATION OF LIGNIN IN ANOXIC ENVIRONMENTS

Gina M. Chaput
University of Massachusetts Amherst

BACTERIAL BIOTRANSFORMATION OF LIGNIN IN ANOXIC ENVIRONMENTS

A Dissertation Presented

by

GINA M. CHAPUT

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2019

Microbiology

BACTERIAL BIOTRANSFORMATION OF LIGNIN IN ANOXIC ENVIRONMENTS

A Dissertation Presented

by

GINA M. CHAPUT

Approved as to style and content by:

_____
Kristen DeAngelis, Chair

_____
James F. Holden, Member

_____
Klaus Nüsslein, Member

_____
Michael Henson, Member

_____
James F. Holden
Microbiology Department Head

DEDICATION

To the next generation of women in science and to the ones who stood before me that made this possible.

ACKNOWLEDGMENTS

"*Not chance of birth or place has made us friends, Being oftentimes of different tongues and nations, But the endeavor for the selfsame ends, With the same hopes, and fears, and aspirations*." *– Dedication by Henry Wadsworth Longfellow*.

I am truly humbled and grateful every day for each person who has touched my life over the years –you helped me make this journey.

To my family – Mom and Dad, I love you and thank you for always being there. I hope I get to be even an ounce of what fantastic, loving, and supportive parents you are to me to my children. My little brother, Christian, thank you for every laugh you have brought out of me even in my most stressful moments. Auntie Shell and Uncle Jerry – thank you for making an effort to remind me to have fun. I'll always be your peanut.

To Dr. Kristen DeAngelis – you motivated me to become the independent scientist that I am now, and I will always be grateful for the lessons I learned under your mentorship. To my lab – especially Grace, Andrew, Roberto, Priya, Alex, Luiz, and Melissa, I loved coming to work because of you. Thank you to my committee members who have guided me throughout my professional and scientific development. Thank you to the Microbiology Department, my science family. Special thank you to the three amazing women I first met at UMass: Jen, Ju, and Emily – you kept my sanity throughout all these years and for that, I am eternally grateful and love you girls. To the graduate students in and outside of my department, I have learned so much from our discussions, enjoyed all our get-togethers, and thankful I can call each of you a friend and colleague. To every student I have had the pleasure of training in lab, watching you succeed has been by far what I consider my greatest achievement during my Ph.D. program. You are what continues to inspire me to be here every day.

To Dr. Ihab Farag. You saw the researcher in me that I couldn't even dream to be. Your mentorship throughout undergraduate school has continued to shape me every year since. I hope you know your yearly calls on my birthday have always been a highlight for me. To my UNH Wildcats – Julie, Machayla, Colin, Becky, Maggie, Meghan, and Ben. Six years, long distance, and at different stages in our lives, but you always made the time to check in on me and cheered me on. Thank you.

To the Merrimack BioDiesel Crew, led by my science teachers, Mr. Tray Sleeper and Mr. Sean Muller. This afterschool program was the foundation for my love of alternative energies and my drive to inspire the next high school student in science. To Kristen Trimble, ten years and counting, and we still dream big together. Thanks for always being my best friend and laughing uncontrollably with me when life brings its challenges.

Finally, to Derrick Alcott. Your love and support have been immeasurable- you inspire me to be the best partner I can be, and I am so grateful to call you mine.

ABSTRACT

BACTERIAL BIOTRANSFORMATION OF LIGNIN IN ANOXIC ENVIRONMENTS

SEPTEMBER 2019

GINA M. CHAPUT

B.S., UNIVERSITY OF NEW HAMPSHIRE DURHAM

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Dr. Kristen DeAngelis

There is a growing need to reduce reliance on non-renewable fuels, especially fossil fuels that contribute to the climate crisis. Plant lignocellulose is an abundant and undervalued source of energy, but its use is hindered due to the recalcitrance of the lignin-component. Current methods to remove lignin have sustainability concerns and are costly for industrial applications such as paper mill pulping. An alternative and greener approach is biopulping, which uses microbes and their enzymes to break down lignin. However, there are limitations to biopulping that prevent it from outcompeting other pulping processes, such as requiring constant aeration and mixing.

The work presented in this dissertation investigates anaerobic bacteria as a promising alternative source for consolidated depolymerization of lignin and its conversion to valuable byproducts. We first ask if anaerobic aromatic metabolism is vertically inherited or horizontally transferred across bacteria. We analyzed seven out of the nine known central intermediate pathways. Of the seven, benzoyl-CoA metabolism had the strongest phylogenetic signal, suggesting vertical inheritance is

the driver of its phylogenetic distribution. This information can be used in future studies to test if predictions can be made for uncharacterized taxa and anaerobic benzoyl-CoA related metabolism.

We also investigated the mechanisms of two uncharacterized isolates, *Sodalis* sp. strain 159R and *Tolumonas lignolytica* BRL6-1. Strain 159R contains many genes related to both aerobic and anaerobic aromatic metabolism but lacks extracellular enzymes for anaerobic lignin depolymerization. Conversely, strain BRL6-1 did not demonstrate lignin metabolism but instead relies on iron redox and organic radicals to potentially modify lignin structure under anoxic conditions. The electron exchange between iron, lignin, and BRL6-1 suggests a protein that acts as a chelator and redox molecule is the intermediate between the bacteria and substrate. The two isolates demonstrate the importance that lignin depolymerization and metabolism may be found separately in organisms and should be considered in future designs for anaerobic biopulping and lignin valorization to be a competitive process on the market.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

1.1 Lignocellulose and the Paper Industry

The United Sates consumes 69 million tons of paper and paperboard materials per year (1). Under the US Environmental Protection Agency (US EPA) Solid Waste Disposal Act, successful efforts have been made to recycle post consumption of paper products; however, the waste aspect of the pulping process in paper production still needs to be addressed, including solid sludge and wastewater (2,3). Pulping is the separation of lignin from the hemicellulose and cellulose components of woody biomass. Lignin is a recalcitrant, aromatic biopolymer, making it very difficult to remove from the cellulose fibers that are used for paper (4). Chemical, thermal, or mechanical processes are used to remove lignin from wood pulp for paper (2). Currently, over 80% of the United States paper industry relies on the chemical separation process, Kraft pulping, which involves NaOH and $NaSO_2$ (5). Despite the high-quality product from this process, Kraft pulping is costly in its chemical recovery and has many environmental challenges. This includes gas emissions of HCl, NH_3, CO, methanol, NO_x, and SO_2 as well as water effluent containing high levels of lignin, organochlorines (from bleaching), organic acids, phosphorus and sulfur compounds and metals (2,3,6). Kraft pulping also contaminates the lignin by-product with sulfur, making it unusable for other applications such as a phenol source for carbon fiber, activated carbon, and other aromatic added-value chemicals (4). Instead, 98% of lignin from the pulping

industry is burned (7). Due to the human and environmental health concerns of the Kraft process, alternative methods need to be developed that are less harmful.

1.2 Biopulping and Lignin Valorization

Bio-pulping, the use of microbes to remove lignin from woody biomass, can be a sustainable replacement for processes such as Kraft pulping due to its low energy consumption, effective delignification, and freedom from chemicals such as chlorine (8–10). Even now, biological processes are used more regularly in treatment of pulp effluent in an effort to reduce lignin and tannins, with an increased interest in bacteria over fungi due to the environmental conditions being too harsh to maintain fungal cultures (3,11–13). Bacteria not only have the ability to depolymerize lignin but also uptake aromatics as a carbon source (14–17). This idea of consolidated bioprocessing, the depolymerization of a material and conversion to a desired product in one process, has been achieved before with polysaccharide break down and conversion to ethanol (18). Therefore, lignin can be converted into flavors, fragrances, dyes, and other valuable secondary metabolites (19).

Microbial-mediated removal of lignin from lignocellulose has predominantly focused on aerobic mechanisms whether cultivating one to a few isolates or using an enzyme cocktail (20–24). Under oxic conditions, microbially-produced iron chelators from both fungi and bacteria can modify and depolymerize lignin. Brown rot fungi, such as *Coriolus versicolor*, and bacteria, such as *Pantoea ananatis* Sd-1, rely on chelator-mediated Fenton chemistry, which is the production hydroxyl radicals from hydrogen peroxide, to disrupt the lignin structure (25,26). Additionally, laccase and peroxidase enzymes can catalyze the depolymerization of

lignin (27). However, employing aerobic mechanisms at an industrial scale is energy intensive and costly to maintain (23,28,29). A solution to this issue is to use anaerobic bacteria, which would be amenable to industrial engineering, can be cultured to high cell density, and would not require aeration. However, few bacteria degrade lignin anaerobically, including *Klebsiella* sp. strain BRL6-2 (30), *Enterobacter lignolytica* SCF-1 (31), and *Agrobacterium s*p. (14), and the mechanisms of anaerobic lignin depolymerization still remain largely not well understood.

1.3 Research Approach and Significance

This dissertation investigates anaerobic bacteria as a promising alternative source of enzymes and microbes that are applicable to consolidated depolymerization of lignin and its conversion to valuable byproducts. In Chapter 2, phylogenetic analysis of anaerobic aromatic metabolism across bacteria demonstrates that benzoyl-CoA catabolism under anoxic conditions has a strong phylogenetic signal and a moderate clade depth that is significantly non-random. These results support that if horizontal gene transfer did occur, it is not recent and instead vertical inheritance has had a stronger role in its phylogeny. This notion was suggested before based on the sequence organization and GC content of the aromatic catabolic island found in *Geobacter metallireducens* (32,33). This information can be used in phylogeny-based prediction algorithms (34) in order to characterize potential lignin degraders and uncover novel mechanisms for biotechnological applications.

In Chapters 3 and 4, bacteria that were originally isolated on lignin as a sole carbon source under anoxic conditions were investigated for mechanisms and proteins related to anaerobic depolymerization and catabolism of lignin. Chapter 3 focuses on temperate forest isolate, *Sodalis* sp. strain 159R, which displays no lignolytic activity when screened for clearing zones on lignin-mimicking dye and lacked genes annotating for extracellular anaerobic lignolytic enzymes. However, strain 159R did contain many genes relating to intracellular anaerobic and aerobic aromatic metabolism. It is likely that strain 159R relies on other bacteria to depolymerize lignin in the soil and then consumes the aromatic monomers as a carbon source. Conversely in Chapter 4, tropical forest isolate, *Tolumonas lignolytica* BRL6-1, does not show evidence of lignin catabolism but instead biotransforms lignin with the aid of a protein that acts as a chelator and redox molecule. Metals, specifically iron, bound to the lignin are removed by the redox protein and as a result organic radicals form and the lignin polymer becomes unstable. *Sodalis* sp. strain 159R and *Tolumonas lignolytica* BRL6-1 are two novel examples to the growing literature that emphasizes that bacteria are part of a larger microbial community that is comprised of either members depolymerizing lignin to access cellulose or hemicellulose components of plant litter, members up-taking and degrading the lignin-derived monomers, or some members that can do both functions (35).

The work presented here advances the effort in identifying isolates that can perform anaerobic lignin depolymerization and catabolism. However, once the foundation is laid, there should also be an emphasis of how to construct microcosms

from such isolates that are efficient for consolidated processing. Using methodological approaches such as bioinformatics, protein expression, metabolite production, and chemical structural analysis of lignin can give a comprehensive outlook of how microbes interact with lignin and each other. This information to then can be applied to design new biotechnology for the pulping industry.

CHAPTER 2

PHYLOGENETIC STRUCTURE OF ANAEROBIC AROMATIC METABOLISM IN BACTERIA

2.1 Abstract

With anthropogenic climate change and industrial input of aromatic compounds, such as PAHs and xenobiotics that are a human and environmental health concern, it is important to understand the mobility and persistence of aromatic compounds in the C cycle. This includes illuminating the metabolic mechanisms of bacteria and their distribution in anoxic environments that act as vital carbon sinks. Characterization of anaerobic aromatic catabolic pathways has increased in the last thirty years, however, the phylogenetic conservation, i.e. the extent to which a phenotype is shared amongst closely related organisms, of each mechanism has yet to be quantified. Elucidating whether or not vertical inheritance or horizontal gene transfer drives the phylogenetic conservation of aromatic anaerobic pathways advances future prediction capabilities of not just taxa that have that function but also what downstream products will be produced. Such knowledge can be implemented for managing bioremediation efforts, biotechnological applications, as well as predicting overall residence time of carbon in soil or marine sediments in C cycle models. In this study we ask (1) which bacteria have the capability for anaerobic aromatic metabolism, and (2) is vertical inheritance or horizontal transfer driving the phylogeny of anaerobic aromatic metabolic pathways?

Seven of the nine known central intermediate pathways were analyzed to determine if anaerobic aromatic metabolism is phylogenetically conserved. Benzoyl-CoA catabolism under anoxic conditions had a strong phylogenetic signal (Fritz and Purvis D) and a moderate clade depth (consenTRAIT $\boldsymbol{\tau}_D$) that was significantly non-random, supporting that vertical inheritance has had a stronger role in its phylogeny. Benzoyl-CoA acyl-CoA hydratase and hydroxyacyl-CoA dehydrogenase both had similar values for clade depth and phylogenetic signal compared to when all three enzymes of the modified β-oxidation pathway were used. Therefore, either benzoyl-CoA acyl-CoA hydratase or hydroxyacyl-CoA dehydrogenase would be suitable markers for phylogenic predictions of benzoyl-CoA metabolism.

2.2 Introduction

Aromatic compounds are the second most abundant form of carbon on Earth, playing an integral role in the carbon cycle (17,36–38). Today, aromatics account for 10-43% of carbon in aquatic systems (39), 17% of the total atmospheric non-methane organic carbon (40), and 20% of plant litter input into soils (41). These compounds can include lignin polymers and monomers, humic substances, fulvic acids, polycyclic aromatic hydrocarbons (PAHs), xenobiotics, and petroleum (17). Understanding the mobility and fate of aromatics between terrestrial, aquatic, and atmospheric systems is critical for modeling C cycling and predicting its ecological effects, especially with influences such as climate change and anthropogenic aromatic pollutants shifting the natural balance of these compounds (42–47).

The flux of aromatics between terrestrial, aquatic, and atmospheric systems is largely controlled by microbial communities and their environments, including abiotic factors such as the presence or absence of oxygen (44,48). Under anoxic conditions, aromatics can act as extracellular electron acceptors as well as serve as a carbon source for bacteria (17,35,38,49,50). Bacterial anaerobic metabolism of aromatics involves the channeling of monomers to a group of central intermediates via peripheral pathways (17). These intermediates undergo dearomatization by reductive reactions before being funneled into the central metabolism (32) (**Fig. 2.1**). For benzoate and its analogs, this involves being first activated to arylcarboxyl-coenzyme A (CoA) esters, followed by dearomatization, and then further alteration in structure by a modified β-oxidation reaction. Other aromatic compounds, such as resorcinol, hydroxyhydroquinone (HHQ), and phloroglucinol, rely on dehydrogenases and reductases for the benzene ring to be broken without becoming a CoA ester first (17,32) (**Fig. 2.1**). These mechanisms of anaerobic aromatic metabolism are distributed differently across bacterial taxa. For instance, the trait of anaerobic benzoyl-CoA metabolism is phylogenetically split between two groups, one comprising of *Thauera*, *Magnetospirillum*, and *Geobacter* strains, and the other with *Azoarcus*, *Aromatoleum*, and *Syntrophus* strains; however, within the two groups are obligate and facultative anaerobes with a vast range of electron acceptors, suggesting that horizontal gene transfer may be a driver in its distribution (32,50,51). Conversely, for the trait of anaerobic HHQ metabolism, the oxidative HHQ pathway has only been found in nitrate-reducing bacteria, therefore potentially being more phylogenetically conserved then benzoyl-CoA (17).

Quantifying the phylogenetic conservatism of anaerobic aromatic metabolic pathways informs us whether or not we can use phylogenetic relatedness to infer capabilities of uncharacterized taxa via phylogeny-based prediction algorithms. Having this ability would enhance C cycle modeling as well as bioremediation efforts for PAH and xenobiotic contamination (49,52–56). Additionally, if these traits can be predicted for certain bacteria, culturing techniques could be adapted to isolate more members and further elucidate the regulation of these mechanisms (15). This knowledge could then be adapted for biotechnological applications such as lignin valorization for bio-energy and bio-material production (57).

Phylogeny, the evolutionary history of species, is a measure of the shared genetics and functional capabilities (traits) of related taxa, where traits are assumed to be shared through vertical inheritance (34,58–62). Vertical gene inheritance, i.e. the passing of genes from ancestor to descendants, is the predominant driver of a functional trait being present in a bacterial taxa despite events of convergent evolution, gene gain/loss, and horizontal gene transfer (34,59,61). For example, bacterial aerobic metabolism of the aromatic compound, p−hydroxyphenylacetic acid, is significantly non-random in phylogenetic distribution, which may be evidence of its vertical transmission; however, this trait is more phylogenetically dispersed than methanogenesis or nitrogen fixation (59), traits considered to be highly conserved in bacteria. The larger phylogenetic dispersion of p−hydroxyphenylacetic acid metabolism was attributed to the complexity (i.e. number of genes) required for a trait. Martiny and colleagues demonstrated that as the number of genes for a trait increases, so does the clade depth and phylogenetic

clustering of that trait (59). In comparison to p−hydroxyphenylacetic acid, aerobic metabolism of benzoyl-CoA contains more genes required for the trait. As expected, this trait had a more clustered phylogenetic signal and deeper clade depth than p−hydroxyphenylacetic acid (34). Since aerobic aromatic metabolism can be predicted based on phylogeny, we hypothesize that anaerobic aromatic metabolism can be as well since the mechanisms require a similar quantity of genes (22,32).

In this study, we analyzed seven of the nine known central intermediate pathways of anaerobic aromatic metabolism to determine what genes could act as indicators for the phylogenetic structure of each pathway and to what extent traits are phylogenetically conserved. This research addresses the following questions: (1) which bacteria have the capability for anaerobic aromatic metabolism, and (2) is vertical inheritance or horizontal transfer driving the phylogeny of anaerobic aromatic metabolic pathways?

If vertical inheritance is the predominant factor shaping the phylogenetic distribution of anaerobic aromatic catabolism, we expect to see a strong phylogenetic signal and a significantly deeper clade depth where the bacteria share the trait than if assumed randomly distributed (59,63). Alternatively, if horizontal gene transfer is the main driver but is occurring only amongst phylogenetically close bacteria, then we would expect to still have a strong phylogenetic signal, a significantly shallower clade depth than expected if the trait was randomly distributed. Finally, if there is no phylogenetic relationship with anaerobic aromatic metabolism, then we expect a low phylogenetic signal for pathways and a shallow or non-significant clade depth. We measure phylogenetic signal and clade depth with

Fritz and Purvis phylogenetic dispersion (D) and consenTRAIT (consensus analysis of phylogenetic trait distribution) algorithm's τ_D, respectively. We also expect that if vertically inherited, enzymes of the same pathway will co-occur with each other and have similar phylogenetic signal as individual units or as a group. To test co-occurrence, association rules via Apriori algorithm as well as Pearson's r were calculated.

The aim of this study was to quantify the phylogenetic conservation of anaerobic aromatic catabolic pathways in order to provide information for future phylogeny-prediction algorithms that could aid in finding taxa for managing bioremediation efforts, biotechnological applications, and C cycle models (50,51,64).

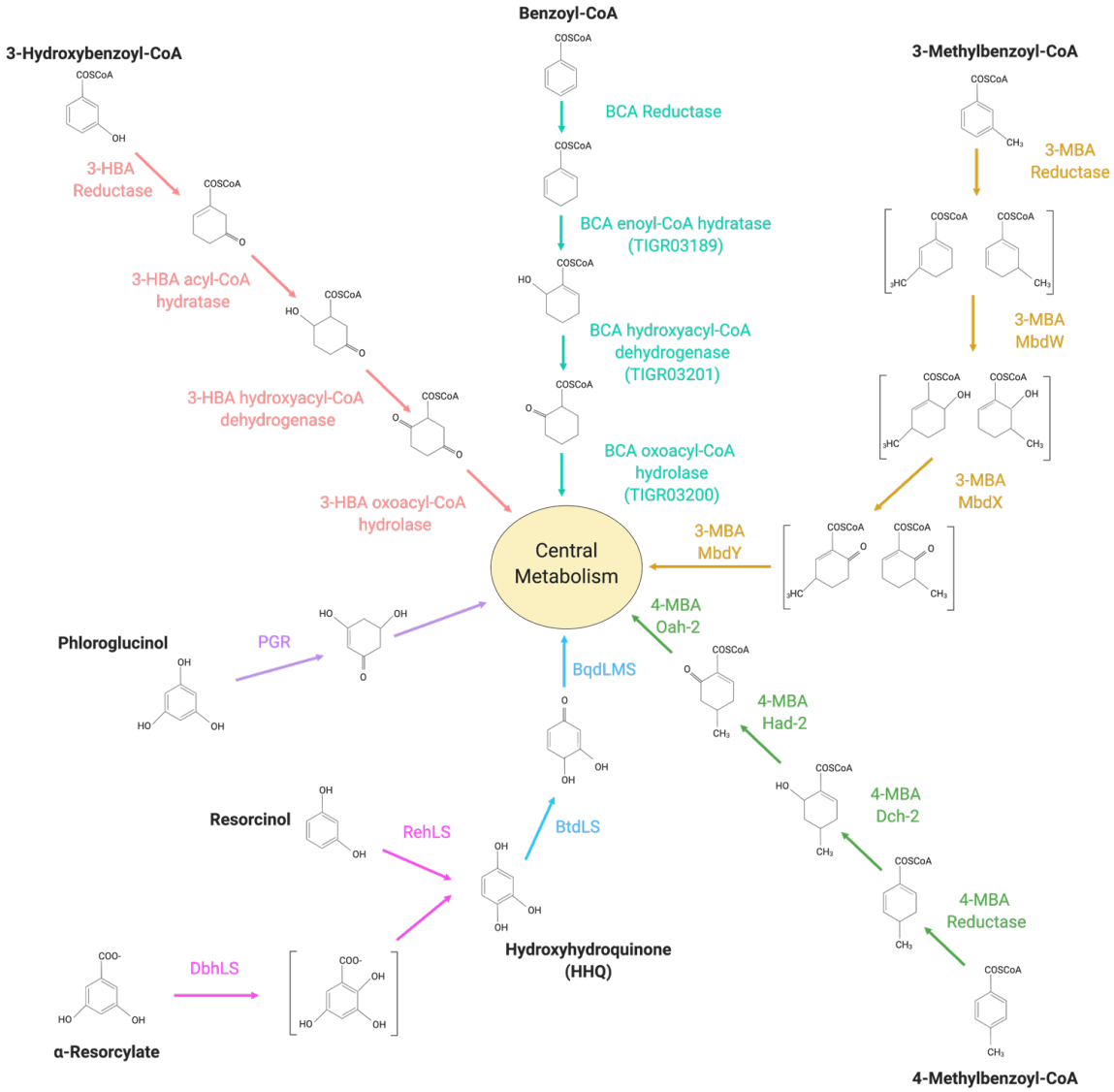


Figure 2.1. Enzyme mechanisms for each of the seven anaerobic aromatic pathways. Ligase reaction and lower pathway of the benzoyl-CoA and analogs are not included. Abbreviations are defined as follows: BCA, benzoyl-CoA; 3-MBA, 3-methylbenzoyl-CoA; MbdW, 3-methylbenzoyl-CoA acyl-hydratase; MbdX, 3-methylbenzoyl-CoA hydroxyacyl-CoA dehydrogenase; MbdY, 3-methylbenzoyl-CoA oxoacyl-hydrolase; 4-MBA, 3-methylbenzoyl-CoA; Dch-2, 3-methylbenzoyl-CoA acyl-hydratase; Had-2, 3-methylbenzoyl-CoA hydroxyacyl-CoA dehydrogenase; Oah-2, 3-methylbenzoyl-CoA oxoacyl-hydrolase; DbhLS, 3,5-dihydroxybenzoate hydroxylase; RehLS, resorcinol hydroxylase; BtdLS, HHQ dehydrogenase; BqdLMS, HHQ ring cleavage enzyme;PGR, Phloroglucinol reductase; 3-HBA, 3-hydroxybenzoyl-CoA. Created with BioRender.com.

2.3 Materials and Methods

2.3.1 Enzyme Selection

Enzymes were selected from 7 of the 9 known anaerobic aromatic central intermediate metabolic pathways that were previously identified and experimentally tested (17,32). To avoid false detection of a pathway, enzymes chosen had to be substrate-specific for compounds of interest and were not involved in dual aerobic/anaerobic reactions, as seen with benzoate-CoA ligase (32). Additionally, enzymes that are assigned two different classes that distinguish between strict and facultative anaerobes, such as benzoyl-CoA reductase, are not informative in capturing the primary phylogenetic signal for a pathway and therefore, were not considered for this study. Instead, the modified β-oxidation reaction, which consists of three enzymes that convert dienoyl-CoA to 3-hydroxypimelyl-CoA, was chosen as it is conserved across bacteria with a wide variety of redox conditions (**Appendix A, Table 1**; **Fig. 2.1**). This is especially true for last enzyme of the reaction, oxoacyl-CoA hydrolase, which has acted as a functional marker for anaerobic aromatic metabolism in environmental samples (50,51,64). Additionally, these three enzymes of the modified β-oxidation reaction are known to be substrate specific between benzoyl-CoA and its analogs. For the phloroglucinol pathway, only one enzyme has been identified, the phloroglucinol reductase (65). Central intermediate resorinol and analog, α-resorcylate each have enzymes identified (RehLS, resorcinol hydroxylase and DbhLS, 3,5-dihydroxybenzoate hydroxylase, respectively) that convert these compounds to

hydroxyhydroquinone (HHQ)(17). HHQ is another central intermediate that is further converted by two enzymes in the oxidative pathway, BtdLS, a HHQ dehydrogenase, and BqdLMS, a HBQ dehydrogenase. DbhLS and RehLS are considered the "Resorcinol pathway" for this analysis and BtdLS and BqdLMS are grouped for the "HHQ pathway" (**Appendix A, Table 1; Fig. 2.1**).

2.3.2 Identification of genomes and genes encoding anaerobic aromatic metabolism

From the Joint Genome Institute Integrated Microbial Genomes/Microbiomes (JGI IMG/M) system, a total of 51,422 isolate genome IDs with corresponding IDs from the National Center for Biotechnology Information (NCBI) were exported on March 28, 2018. After duplicates (organisms having the same NCBI taxon ID) were removed, 34,471 genomes remained. Aligned 16S rRNA gene sequences were exported from the SILVA aligned SSU Ref NR database (675,833 sequences after removing duplicates based on NCBI ID) and matched to the corresponding organisms in the IMG dataset using the NCBI IDs as reference, resulting in 14,980 organisms remaining. Of this set, 12,888 genomes were available to be downloaded from IMG JGI or NCBI as protein FASTA files for HMMER analysis (HMMER 3.2.1; hmmer.org). Proteins of interest that have been identified previously in more than one bacterium were aligned using MUSCLE (66) (**Appendix A, Table 1**). For the benzoyl-CoA pathway, TIGRfam IDs TIGR03189, TIGR03200, and TIGR03201 sequence alignments were used (67). For the other six pathways, HMMER analysis was completed with either a profile search with the multiple sequence alignments or a jackhmmer search for those enzymes that had only one sequence available

(**Appendix A, Table 1**). To determine an appropriate E-value cut-off for each enzyme, the strict aerobe genus *Acinetobacter* was used as an indicator that should not contain any of the enzymes of interest (68). Any organism whose protein had an E-value equal to or greater than any *Acinetobacter* species was removed. The dataset was then screened for the presence or absence of genes related to anaerobic respiration using IMG JGI Function Profile and selected KEGG IDs modified from the list of Llorens-Marès *et al.* 2015 (**Appendix A, Table 2**). If an organism had at least one set of selected KEGG IDs, it was retained. Finally, isolates designated as strains (>97% 16S rRNA gene sequence similarity) were removed in an effort to not have multiple strains of the same species, reducing the final dataset of positive genotypes to 1,874 (**Table 2.1**). Four outgroup organisms were added: *Arabidopsis lyrata lyrata* MN47, *Aspergillus niger* (ATCC 1015), *Candida albicans* SC5314, and *Saccharomyces cerevisiae* S288C. Negative genotypes (those lacking all genes for any of the selected pathways) are needed to run the consenTRAIT analysis. A proof of concept using benzoyl-CoA enoyl dehydratase was completed to confirm that repeated random choices gives the same trait depth across 100 different negative species sets (average τ_D was 0.0227 with a variance of $1.5e^{-6}$). Therefore, from Step #5 (**Table 2.1**) 20% negative genotypes (total of 1,102 organisms) from the 12,888 OTUs (with strain level organisms removed) were selected at random and added back into the dataset, making the total 2,985 organisms considered in our analyses.

Table 2.1. Parameters for Dataset Construction

Step	Description	Database Size (number of genomes)
1	March 28, 2018 IMG JGI Isolate Genome Database of IMG and NCBI IDs	51,422
2	NCBI ID duplicates removed	34,471
3	Genomes that had SILVA 16S rRNA gene available with matching NCBI ID	14,980
4	Genomes available for export from IMG JGI	12,876
5	Added 11 known positive genomes(Durante-Rodríguez et al 2018) from NCBI not available on IMG JGI	12,888
6	Genomes containing at least one enzyme of the 17 target pathways ("Positive Genotypes") based on HMMER analysis	6,304
7	Removed genomes that are same species but strain level duplicates	3,961
8	Kept genomes that were positive for KEGG facultative of anaerobic respiration	1,874
9	Adding 20% of genomes with no target enzymes ("Negative Genotypes") back into dataset and 4 outgroup species	2,985

2.3.3 Trait Correlation, Association Rules, and Gain/Loss Events

Pairwise enzyme combinations were compared using Pearson product-moment correlation coefficient. Additionally, association rules to predict enzyme and pathway co-occurrence were determined by Apriori algorithm with the arules package in R (69,70). Only positive genotypes with at least one trait were considered for this analysis with a support threshold of 0.01, a confidence of 0.8, and a minimum item count of 3. Finally, an estimation of trait gain and loss events was determined based on a Wagner parsimony with Count and the gain penalty set to 1 (71).

2.3.4 Phylogenetic Analysis

The 16S rRNA gene sequences of the 2,985 genome dataset were bootstrapped using the SEQBOOT program from the PHYLIP software package to

generate 100 data sets (72). Pairwise genetic distances between 16S rRNA gene sequences in each of the 100 data sets were measured with the PHYLIP DNADIST program utilizing the F84 model of nucleotide substitution (72). The distance matrices generated from DNADIST were then run through the PHYLIP NEIGHBOR program to construct neighbor-joining trees. One tree (not bootstrapped) was created in NEIGHBOR and visualized using iTol (**Fig. 2.2**) (73).

To determine if anaerobic aromatic metabolism is significantly phylogenetically conserved, two measurements were used: trait depth ($\boldsymbol{\tau}_D$) and phylogenetic dispersion (D) (Martiny et al 2013; Fritz and Purvis 2010). Trait depth is calculated using the consenTRAIT (consensus analysis of phylogenetic trait distribution) algorithm, which calculates the genetic distance (clade depth) from the tips of the tree to the last common ancestor of at least 90% of the relatives sharing the trait of interest. To determine if a trait was phylogenetically conserved (i.e. significantly non-random), the binary values of a trait were randomized 1000 times and the respective $\boldsymbol{\tau}_D$ values were calculated to make a null distribution. The observed $\boldsymbol{\tau}_D$ was then compared bidirectionally to the null distribution to determine if the trait was either significantly shallower than expected if randomly distributed (horizontal gene transfer) or significantly deeper than expect if randomly distributed (vertically inherited). This was calculated as the fraction of randomized $\boldsymbol{\tau}_D$ less than or equal to the observed $\boldsymbol{\tau}_D$ and the fraction of randomized $\boldsymbol{\tau}_D$ greater than or equal to ($P < 0.05$). Phylogenetic dispersion (D) from the Fritz and Purvis test was calculated with the *caper* package in R (74). 1000 permutations are run for a trait through two null models, one assuming phylogenetic randomness and the

second assuming a “Brownian-motion”- like evolutionary distribution. If D value is less than 0, then the trait is highly conserved; if D ≈ 0 then the trait follows the Brownian model of conservation; if D is greater than 1 then the trait is phylogenetically over-dispersed. Using both $\boldsymbol{\tau}_D$ and D values concurrently allow us to evaluate if there is a strong correlation between phylogeny and a trait, if there is a moderate correlation, or if the trait is randomly associated to phylogeny.

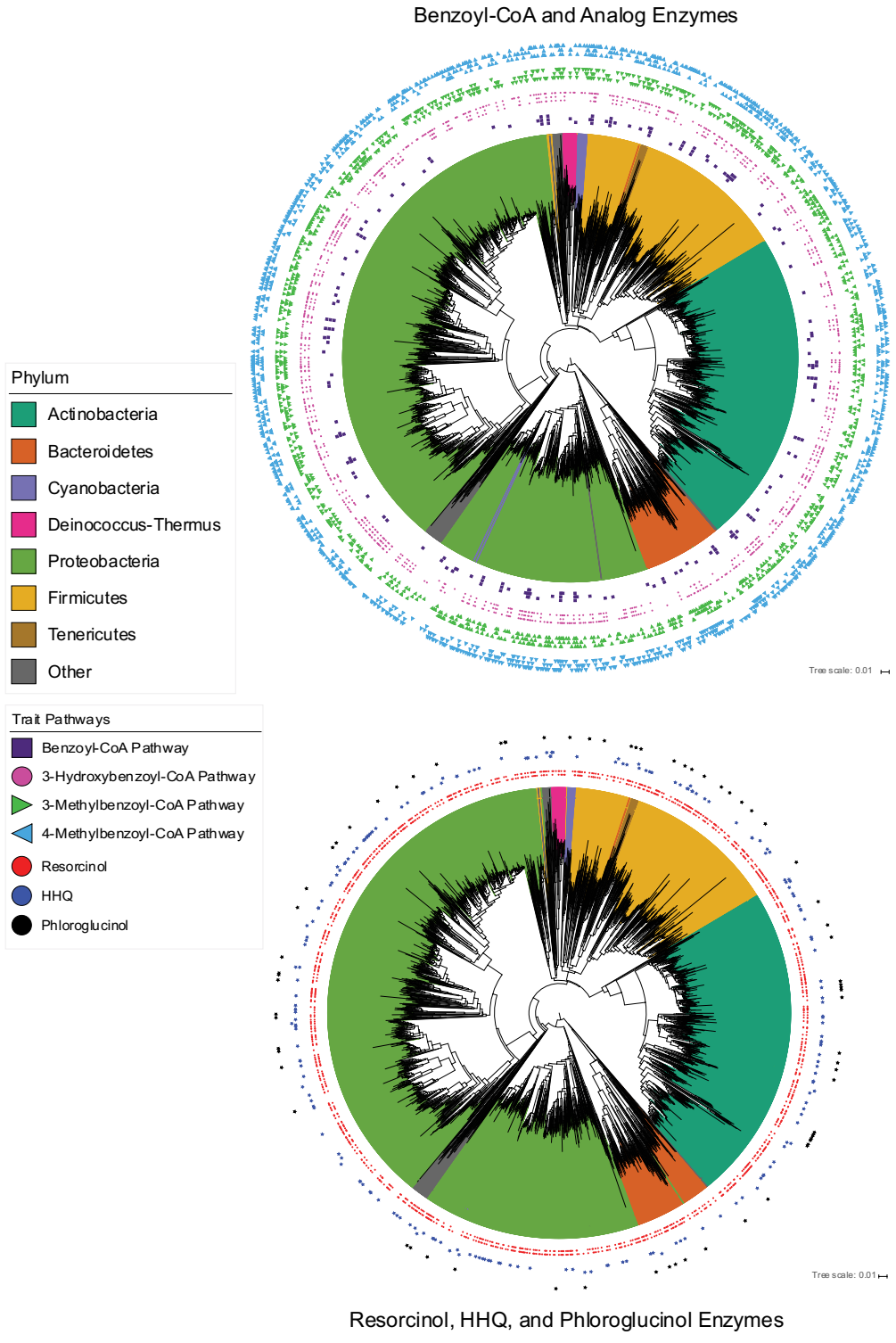


Figure 2.2. Phylogenetic tree of the dataset with the distribution of each enzyme and its respective pathway of interest. Trees were visualized with iTol (73) and contains 2,985 tips for both (A) benzoyl-CoA and analogs and (B) resorcinol, hydroxyhydroquinone, and phloroglucinol.

2.4 Results and Discussion

2.4.1 Frequency of Enzymes and Pathways Across Phyla

To assess if our parameters for selecting genomes were giving an over- or underestimation of positive genotypes, we compared the number of positive genotypes for benzoyl-CoA (BCA) enoyl dehydratase, hydroxyacyl-CoA dehydrogenase, and oxoacyl-CoA hydrolase to the AnnoTree database, a functionally annotated bacterial tree of life (75). When unclassified bacterial species were removed, Annotree genome results were similar in count to our dataset (**Table 2.2**). In addition, based on analysis of the IMG JGI Function Profile across the 51,422 isolate genomes that were originally exported, positive genomes for the three enzymes were also similar in count to our positive genotypes (**Table 2.2**) (76).

Table 2.2. Comparison of positive genotypes of our dataset to Annotree and Function Profile/IMG JGI.

Enzyme	Our Database	Annotree	Function Profile/IMG JGI
BCA acyl-CoA hydratase	41	88	76
BCA hydroxyacyl-CoA dehydrogenase	84	61	52
BCA oxoacyl-CoA hydrolase	54	67	52

A total of 2,985 organisms across 23 phyla were analyzed in this study, with Proteobacteria being the most representative phyla (52.5% of genomes analyzed), followed by Actinobacteria (22.6%), Firmicutes (15%), and Bacteroidetes (5%). The remaining phyla in the data set all contained <50 genomes each. Of the positive genotypes (1,874 organisms), α-, β-, δ-, and γ-Proteobacteria, Firmicutes, and

Actinobacteria had the highest frequency of individual enzymes across pathways and β- and γ-Proteobacteria had the highest frequency of genomes with complete sets of enzymes for BCA and BCA analogs' modified β-oxidation reaction (**Fig. 2.3A**). These results agree with previous observations of phyla frequency in microbial communities that are commonly found in soil where natural aromatic compounds are high in concentration (77) or exposed to PAH contamination (78). Our analysis also supports findings of which classes of the Proteobacteria phylum associate with anaerobic aromatic metabolism, specifically α-, γ-, and β-Proteobacteria have large roles in PAH degradation (79).

For aromatic compounds with *meta*-positioned hydroxyl groups, our data suggests that phloroglucinol pathway predominantly consisted of Firmicutes and Bacteroidetes whereas genotypes containing the resorcinol pathway were found in Proteobacteria, Firmicutes, and Actinobacteria (**Fig. 2.3B**). There was no overlap between HHQ dehydrogenase and HBQ dehydrogenase enzymes across genomes and therefore these two enzymes could not be analyzed as a group to represent the HHQ pathway. This may be due to lack of sequences since only two organisms have been studied to contain the complete pathway of HHQ metabolism, *Azoarcus aromatica* and *Thaera aromatica* (80), neither of which were included in the dataset due to availability on the IMG database.

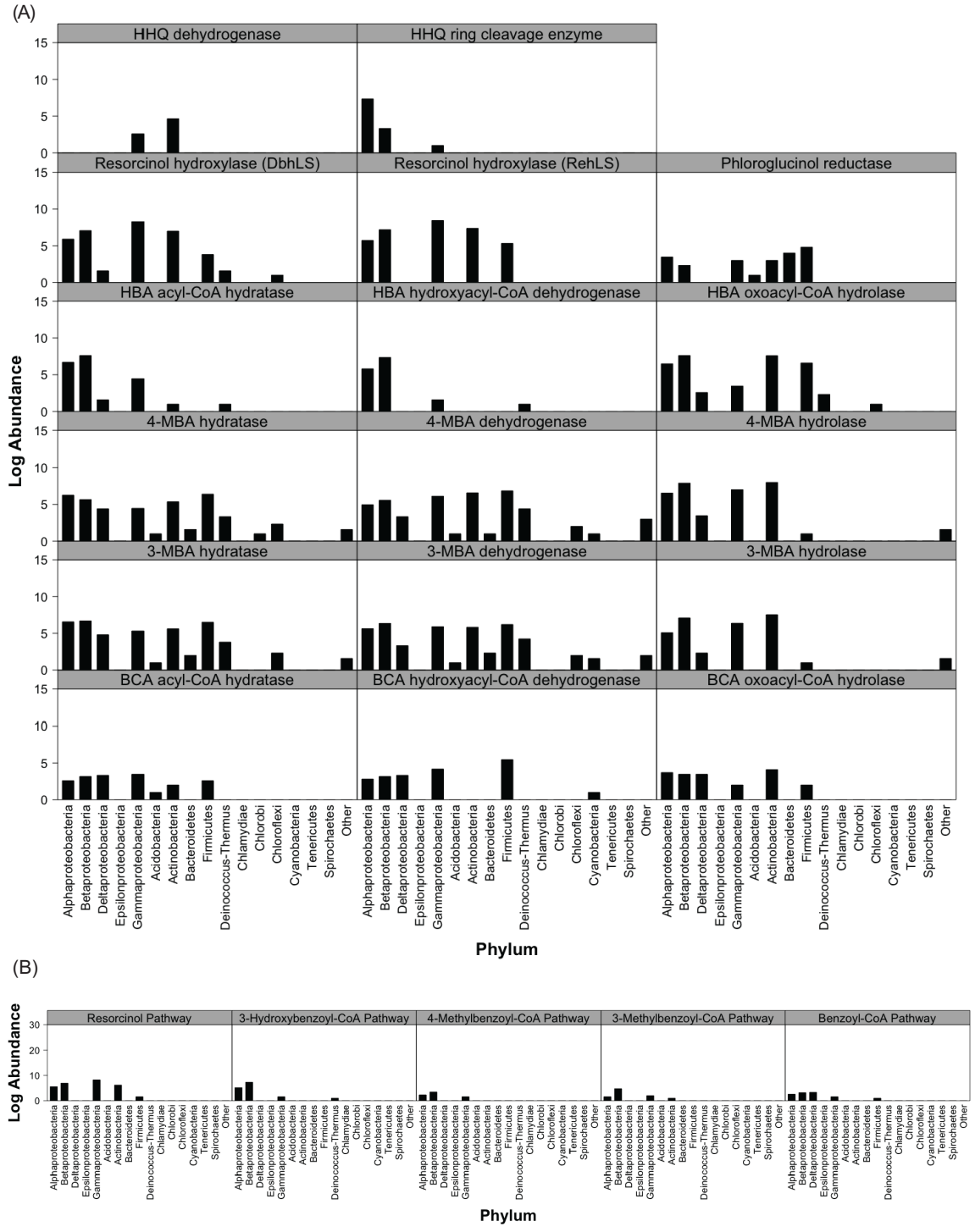


Figure 2.3. Log abundance of positive genotypes grouped by phylum containing traits for (**A**) individual enzymes and (**B**) groups of enzymes that were chosen to represent a pathway.

2.4.2 Pathway Correlation and Association Rules

Gene amplification has been thought to be a driver in the evolution and diversity of many catabolic functions, including aromatic metabolism (81). Evidence for this concept includes *Azoarcus, Thauera*, and *Geobacter* strains that have been identified as capable of metabolizing benzoyl-CoA as well as many of its analogs (32). Therefore, we expected to see benzoyl-CoA analog pathways to co-occur in an organism that had the ability to anaerobically metabolize benzoyl-CoA. To test this, genomes were considered for analysis only if they contained the full set of enzymes assigned to represent the pathway (**Appendix A, Table 1**). When comparing pairwise between positive genotypes, there was a significant but weak positive correlation between the 3-hydroxybenzoyl-CoA and both the benzoyl-CoA (BCA) pathway and the 3-methylbenzoyl-CoA pathway ($r = 0.1088$ and 0.0378, respectively; $P < 0.05$; **Fig. 2.4A**). A weak positive correlation was found between 3-methylbenzoyl-CoA and 4-methylbenzoyl-CoA ($r = 0.2597$; $P < 0.05$). Resorcinol showed to have a weak positive correlation to benzoyl-CoA analog pathways ($r = 0.0775$, 0.0606, and 0.060 for 3-MBA, 4-MBA, and HBA, respectively) but no significant correlation to benzoyl-CoA. Results suggest that pathways for the anaerobic metabolism of benzoyl-CoA and its analogs occurred independently and not due to gene amplification.

To determine if individual enzymes assigned to a pathway are present together in a genome, pairwise comparison was calculated for positive genotypes and their gene content. There was a strong positive correlation between the enzymes within the BCA pathway, the HBA pathway, and the resorcinol pathway

(**Fig. 2.4B**). Weak to no correlation was seen for enzymes within 3-MBA and 4-MBA pathways; however, enzymes of the same function between 3-MBA and 4-MBA had strong positive correlation. This suggests 3-MBA and 4-MBA enzymes of the same function may be too close in sequence similarity and more sequences are needed to distinguish them between genomes. Another possibility is that the3-MBA and 4-MBA enzymes are functionally interchangeable; however, Lahme *et al*. (2012) identified 3-methylglutarate as a metabolic intermediate in the 4-MBA pathway of *Magnetospirillum* sp. strain pMbN1, suggesting that the 4-MBA is conserved (82).

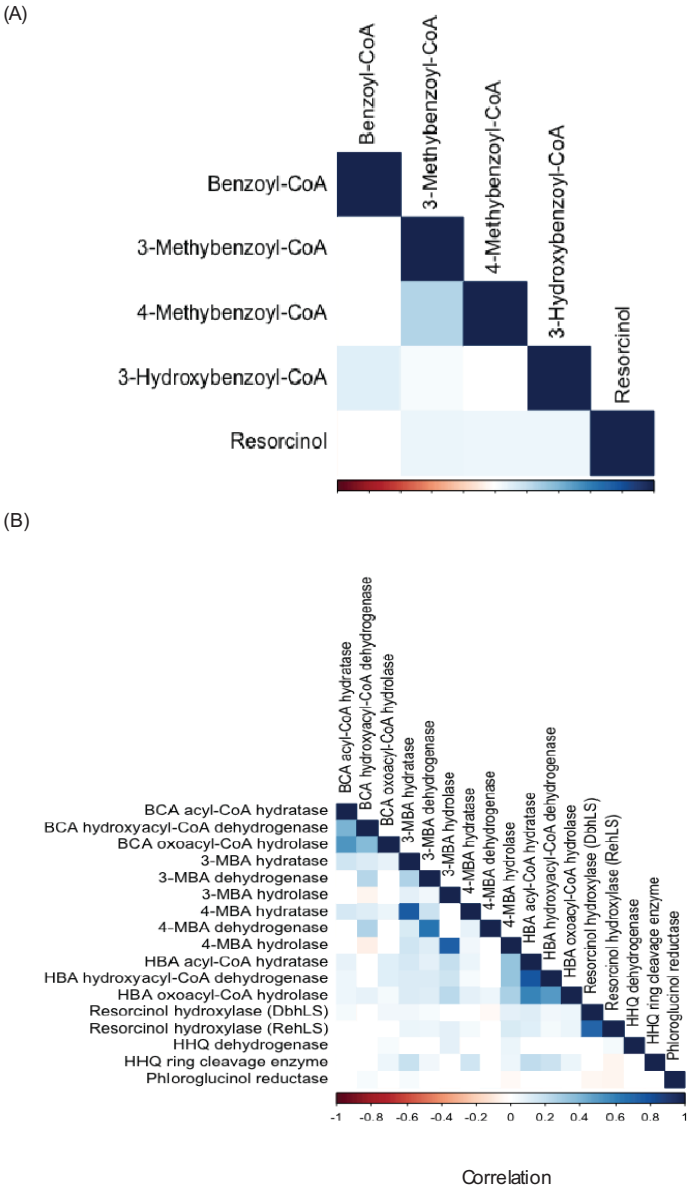


Figure 2.4. Pearson product-moment correlation coefficient for (**A**) individual enzymes pairwise and (**B**) groups of enzymes representing a pathway pairwise. Abbreviations are defined as follows: BCA, benzoyl-CoA; 3-MBA, 3-methylbenzoyl-CoA; 4-MBA, 3-methylbenzoyl-CoA; HHQ, hydroxyquinone.

To further look at the co-occurrence between two or more individual enzymes or pathways occurring in a bacterial genome, we data-mined our positive genotypes for association rules using the Apriori algorithm (83,84). An association

rule implies the co-existence of a number of enzymes in a portion of a positive genotype dataset and that the frequent existence of two or more enzymes in the same genome implies a relationship among them (83). For example, we would expect to see an association rule for the three enzymes of the modified β-oxidation reaction for anaerobic BCA and its analog pathways. A total of 473 association rules were determined from the Apriori algorithm. For the BCA pathway, if a genome had both hydroxyacyl-CoA dehydrogenase and oxoacyl-CoA hydrolase, then it was 42 times more likely to also have acyl-CoA hydratase than if the algorithm assumed that the presence of the enzymes were unrelated to each other (confidence = 0.93). However, if a genome had acyl-CoA hydratase and hydroxyacyl-CoA dehydrogenase, it was 34 times more likely to have the oxoacyl-CoA hydrolase (confidence = 1.0). This difference between enzymes within the same pathway may be due to gene duplication, where gene amplification and gene rearrangements could make it easier for an enzyme such as BCA oxoacyl-CoA hydrolase to be horizontally transferred and therefore appear to be less conserved than that of the complete set of enzymes for the modified β-oxidation reaction of the BCA pathway (81). For example, in *Geobacter metallireducens*, the BCA oxoacyl-CoA hydrolase is duplicated in its genome and the paralog is located in a different region from the 300 kb genomic island of aromatic catabolic genes (32,33). Conversely, the gene clusters present on genomic islands, like in *G. metallireducens* and *G. daltonii*, are speculated to be no longer mobile and are vertically transferred, making the co-occurrence of these enzymes together higher (33,85).

Other than the enzymes within the BCA pathway, rules of co-occurrence did not display any clear biological relevance. Majority of association rules grouped 4-5 enzymes from BCA analog pathways and enzymes from resorcinol and HHQ pathways with the likelihood of occurrence between 2-10 times. No associations with phloroglucinol reductase were seen which was surprising since phloroglucinol originates from lignin derivative gallate, and therefore would be present with many other aromatics in nature (17,65). In addition, no association rules were seen when comparing pathways as groups of enzymes. Results support the Pearson's r analysis that BCA enzymes seem to cluster together on genomes within ancient genomic islands that are vertically transferred whereas the other pathways may be more susceptible to gene amplification and horizontal gene transfer (33,86).

2.4.3 Phylogenetic Conservation of Traits

Enzymes in our analysis that displayed moderate to strong phylogenetic signal and shallow but significantly non-random clade depth suggest that horizontal gene transfer may be occurring for these traits but between phylogenetically close microbes (**Fig. 2.5 and 2.6**) (34). An explanation for this observation would be anthropogenic inputs of aromatics such as resorcinol, hydroquinone, and catechol (which is funneled into the HBA pathway) increasing environmental pressure for horizontal gene transfer in contaminated areas (32,81,87–89). As such, our analysis reveals resorcinol, HHQ, and HBA are the three pathways where enzymes have the strongest evidence for horizontal gene transfer. Therefore, microdiversity, where phylogenetically related groups are distinct in physiology due to location, may be a

driver for the phylogenetic conservation for these traits (87,90,91). Microdiversity has also been suggested before for other enzymes such as chitinases and b-N-acetyl-glucosaminidases (87).

To see if phylogenetic conservation patterns observed for individual enzymes are consistent to those observed for pathways, $\boldsymbol{\tau}_D$ and Purvis D were calculated again but with genomes only considered "positive" for a trait if they contained the complete set of marker enzymes for a pathway (**Appendix A, Table 1**). Benzoyl-CoA pathway had a τ_D of 0.015 which was significantly non-random (**Table 2.3**, **Fig. 7**) and a Purvis D of 0.387 suggesting a stronger phylogenetic conservatism than what has been previously seen for simple carbon substrate utilization and extracellular enzyme production ($\tau_D < 0.010$)(59,87). In addition, the clade depth of anaerobic BCA catabolism was closer to that seen for aerobic pathway ($\tau_D = 0.24$)(34). The benzoyl-CoA analogs (HBA, 3MBA, 4MBA) and resorcinol had very shallow clade depths, however, with HBA being randomly distributed. This observation of the BCA pathway having a deeper trait depth compared to its analog compounds supports previous findings that the pathway's genes are clustered together on a genomic island and are no longer mobile and phylogenetically conserved (33). Similarly, the results of the HBA and resorcinol pathways support the results of their individual enzymes where horizontal gene transfer and microdiversity are likely recent drivers for the phylogenetic distribution of these traits.

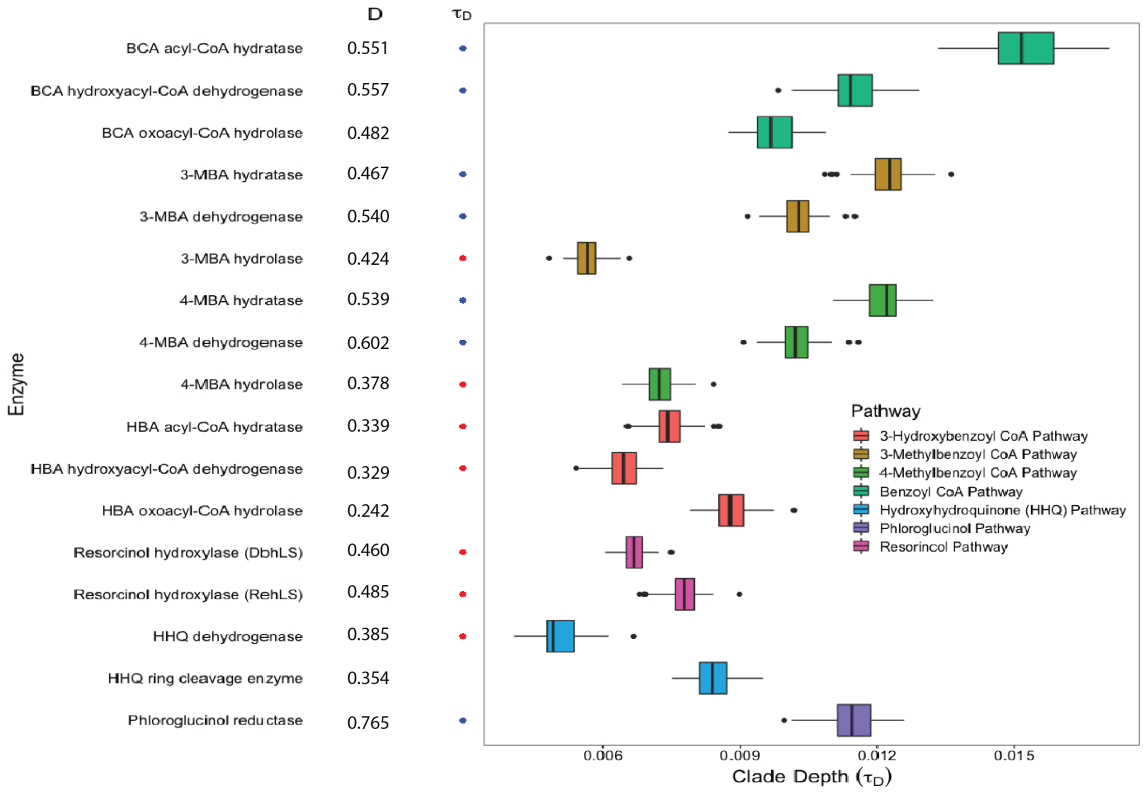


Figure 2.5. τ_D of anaerobic aromatic metabolic enzymes that was calculated with consenTRAIT with the boxplot represents the τ_D values from the 100 bootstrap trees. Dots represent if the observed τ_D was non-randomly distributed ($P < 005$), being significantly shallower compared to the randomized null distribution of τ_D (red) or being significantly deeper compared to the randomized null distribution of τ_D (blue). Fritz and Purvis D values ($P < 0.05$) are listed for each enzyme.

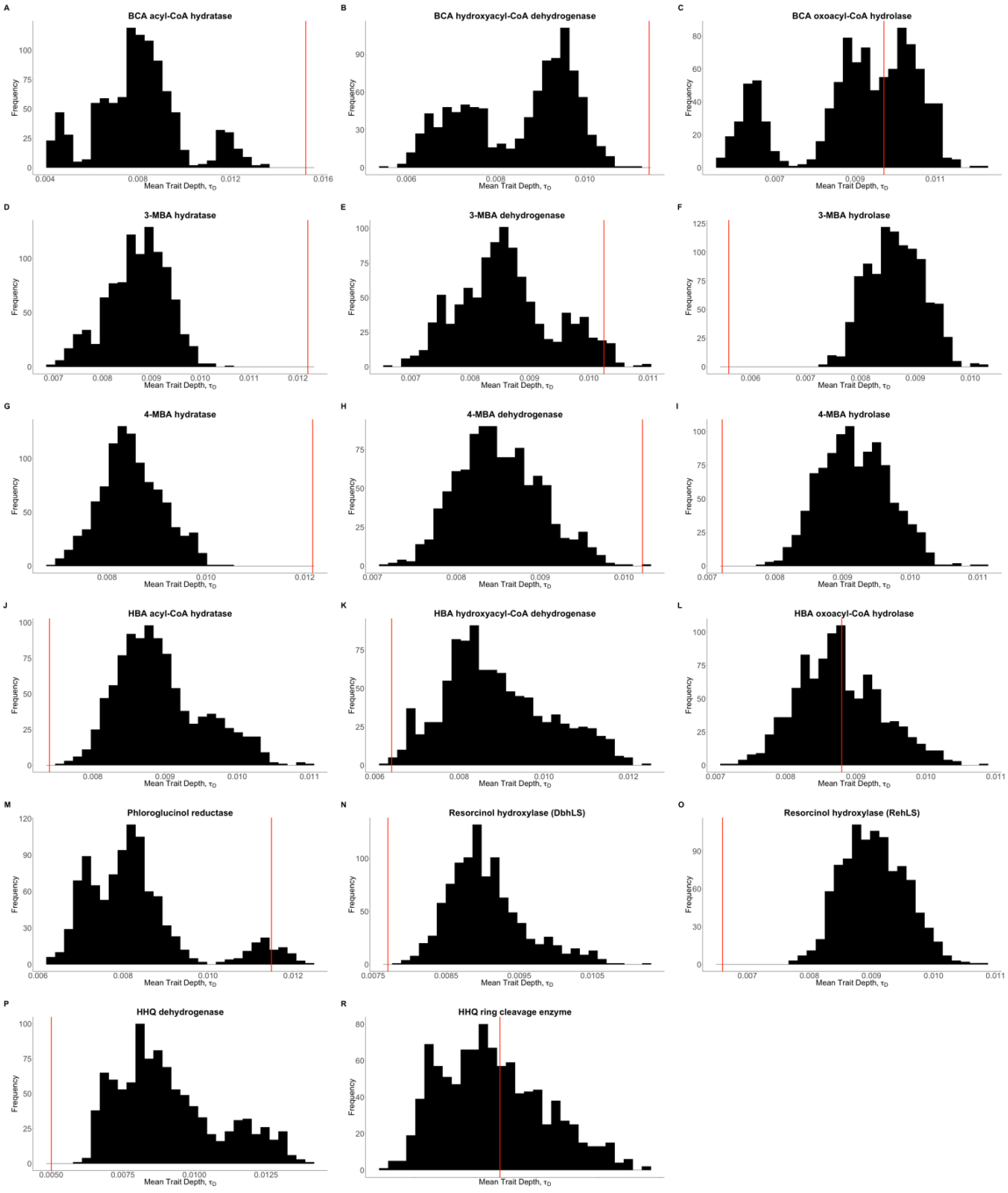


Figure 2.6. Comparing observed values of τ_D (highlighted as red line) to the randomized null distribution of individual enzymes (black bars). (**A**) Benzoyl-CoA acyl-CoA hydratase, (**B**) Benzoyl-CoA hydroxyacyl-CoA dehydrogenase, (**C**) Benzoyl-CoA oxoacyl-CoA hydrolase, (**D**) 3-Methylbenzoyl-CoA acyl-CoA hydratase, (**E**) 3-Methylbenzoyl-CoA hydroxyacyl-CoA dehydrogenase, (**F**) 3-Methylbenzoyl-CoA oxoacyl-CoA hydrolase, (**G**) 4-Methylbenzoyl-CoA acyl-CoA hydratase, (**H**) 4-Methylbenzoyl-CoA hydroxyacyl-CoA dehydrogenase, (**I**) 4-Methylbenzoyl-CoA oxoacyl-CoA hydrolase, (**J**) 3-Hydroxybenzoyl-CoA acyl-CoA hydratase, (**K**) 3-

Hydroxybenzoyl-CoA hydroxyacyl-CoA dehydrogenase, (**L**) 3-Hydroxybenzoyl-CoA oxoacyl-CoA hydrolase, (**M**) Phloroglucinol reductase, (**N**) Resorcinol hydroxylase, (**O**) Resorcinol hydroxylase, (**P**) Hydroxyquinone dehydrogenase, (**R**) Hydroxyquinone reing cleavage enzyme.

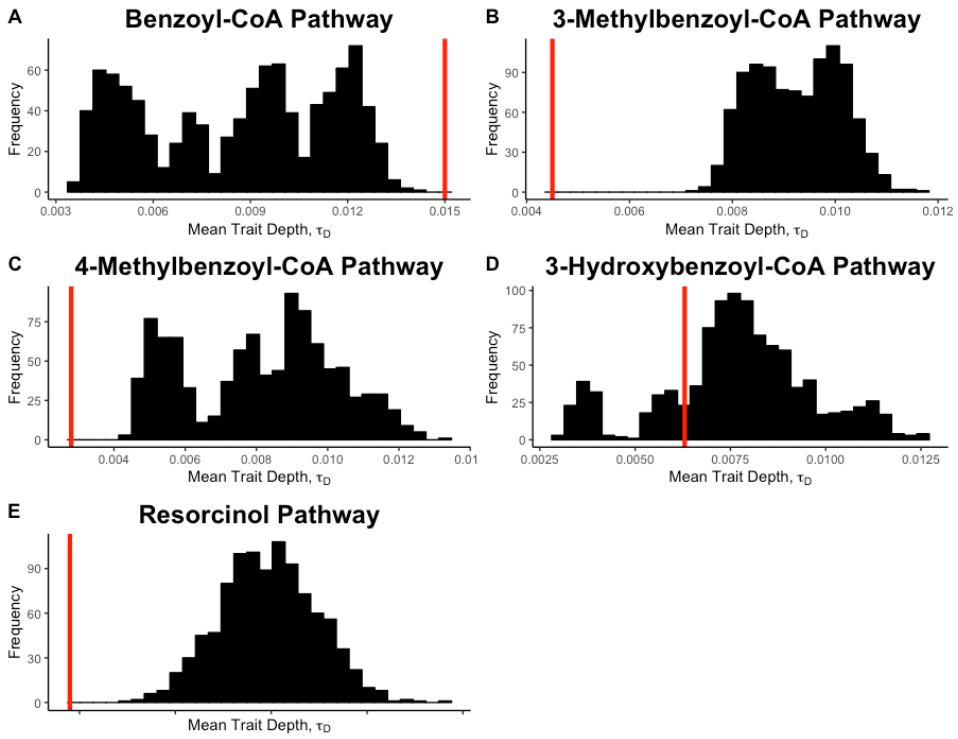


Figure 2.7. Comparing observed values of τ_D (highlighted as red line) to the randomized null distribution of groups of enzymes that represent pathways (black bars). (**A**) Benzoyl-CoA pwathway, (**B**) 3-Methylbenzoyl-CoA pathway, (**C**) 4-Methylbenzoyl-CoA pathway, (**D**) 3-Hydroxybenzoyl-CoA pathway, (**E**) Resorcinol Pathway.

Table 2.3 Phylogenetic conservation parameters calculated with HMMER (enzyme positive genotypes), consenTRAIT (mean clade size and trait depth for both observed and randomized values), and Count (trait gain and loss). Abbreviations are as follows: BCA, benzoyl-CoA; 3-MBA, 3-methylbenzoyl-CoA; 4-MBA, 4-methylbenzoyl-CoA; HBA, 3-hydroxybenzoyl-CoA; HHQ, hydroxyhydroquinone.

Data set	Enzyme-positive genotypes	Mean clade size Observed	Mean trait depth (τ_d) observed	Mean Clade size Randomized	Mean trait depth (τ_d) randomized	Trait gains	Train losses
BCA acyl-CoA hydratase	41	1.28	0.0152	1.00	0.0081	28	1
BCA hydroxyacyl-CoA dehydrogenase	84	1.35	0.0114	1.00	0.0085	37	1
BCA oxoacyl-CoA hydrolase	54	1.31	0.0097	1.00	0.009	57	1
3-MBA hydratase	423	1.74	0.0122	1.04	0.0087	185	32
3-MBA dehydrogenase	358	1.63	0.0103	1.03	0.0086	190	20
3-MBA hydrolase	443	1.71	0.0057	1.05	0.0086	187	35
4-MBA hydratase	307	1.39	0.0122	1.03	0.0085	167	26
4-MBA dehydrogenase	394	1.50	0.0102	1.04	0.0085	239	15
4-MBA hydrolase	721	2.21	0.0072	1.08	0.0092	224	71
HBA acyl-CoA hydratase	322	1.87	0.0075	1.03	0.0089	164	60
HBA hydroxyacyl-CoA dehydrogenase	220	2.15	0.0064	1.02	0.0089	109	29
HBA oxoacyl-CoA hydrolase	591	2.06	0.0088	1.06	0.0088	57	23
Resorcinol hydroxylase (DbhLS)	645	1.97	0.0078	1.07	0.0091	236	60
Resorcinol hydroxylase (RehLS)	744	1.85	0.0067	1.09	0.0090	270	85
HHQ dehydrogenase	29	1.58	0.0051	1.00	0.0091	17	0

HHQ ring cleavage enzyme	172	1.54	0.0084	1.01	0.0083	66	25
Phloroglucinol reductase	71	1.18	0.0115	1.00	0.0083	57	1
Benzoyl-CoA Pathway	25		0.015		0.0086	14	1
3-Methybenzoyl-CoA Pathway	32		0.0045		0.0081	17	3
4-Methybenzoyl-CoA Pathway	16		0.0028		0.0076	10	1
3-Hydroxybenzoyl-CoA Pathway	197		0.0063		0.0076	58	17
Resorcinol Pathway	542		0.0069		0.0089	184	52

2.5 Conclusion

In this study we asked (1) which bacteria have the capability for anaerobic aromatic metabolism, and (2) is vertical inheritance or horizontal transfer driving the phylogeny of anaerobic aromatic metabolic pathways? Based on Purvis D and τ_D results, the trait of benzoyl-CoA metabolism under anoxic conditions was the most phylogenetically conserved and suggests that vertical inheritance is the predominant driver in its distribution. Conversely, resorcinol, HHQ, and HBA pathways have strong evidence for horizontal gene transfer and microdiversity having roles in their evolution, likely due to increased anthropogenic inputs of aromatic contaminants (88).

With new evidence that anaerobic benzoyl-CoA catabolism has strong phylogenetic conservatism, it may be possible to apply this trait towards phylogeny-prediction algorithms to characterize novel taxa of unknown function (34). Benzoyl-CoA is the most common intermediate for anaerobic bacteria, with many lignin derivatives being funneled into benzoyl-CoA via peripheral pathways (17). Therefore, identification of novel isolates could lead to new mechanisms for biotechnological applications such as bio-pulping and lignin valorization.

CHAPTER 3

SODALIS SP. STRAIN 159R, ISOLATED FROM AN ANAEROBIC LIGNIN DEGRADING CONSORTIA

3.1 Abstract

A novel strain, designated 159R, was isolated from a consortia originated from temperate forest soil and enriched on organosolv lignin as the sole carbon source under anoxic conditions. Phylogenetic analysis based on 16S rRNA gene sequencing placed the strain within the genus *Sodalis*. Genome sequencing revealed a genome size of 6.38 Mbp and a G+C content of 54.9 mol%. The genome contains genes for host-symbiont associations with insects as seen with other *Sodalis* members as well as genes for lignin derived monomer catabolism. Pairwise whole genome average nucleotide identity (ANI) values suggest that strain 159R represented a new species. To resolve the phylogenetic position of the new strain, its phylogeny was reconstructed from sequences of 400 conserved genes in PhyloPhlan as well as from 49 Clusters of Orthologous Groups (COG) domains via KBase. Our phylogenetic analysis revealed that 159R is more distantly related to the *Sodalis* clade than close-relative, *Biostraticola tofi*. However, percentage of conserved proteins (POCP) supported that strain 159R was part of the *Sodalis* genus. Based on these results, strain 159R could represent a more ancient precursor of the *Sodalis* clade than the only other free-living member, *Sodalis praecaptivus* HS.

3.2 Introduction

The genus *Sodalis* is in the family Enterobacteriaceae and was established by Dale and Maudlin with the description of species *S. glossinidius* strain M1. M1 was isolated from the tsetse fly, *Glossina morsitans morsitans*, using an agar-based medium and was the first isolated insect secondary endosymbiont (92). Insect secondary endosymbionts are recently established symbiotic associations that can be horizontally or vertically transmitted as well as introduced to the host by the environment (93). *Sodalis* species are generally endosymbionts to a range of insect hosts such as long-horned beetles, louse flies, and bees (94–96). There is evidence that certain insect hosts may also serve as vectors for the transmission of *Sodalis* to alternative hosts, such as plants (97). Due to the *Sodalis*-allied clade predominantly consisting of facultative and obligate mutualistic symbionts, their genomes are degenerated, having lost majority of the gene inventory and becoming smaller in size (98,99). Since many of the *Sodalis* species are speculated to be more recently acquired symbionts, there is a unique opportunity to understand how these associations evolve (99). However, resolving the evolutionary relationships of *Sodalis* is difficult due to *Sodalis* genomes containing pseudogenes, mobile DNA, gene rearrangements, duplications, and deletions (99,100).

The only free living *Sodalis* species identified to date is *S. praecaptivus* HS (101). HS was isolated from an infected human wound that had been impaled by a dead crab apple tree branch, suggesting that HS was either a pathogen or saprophyte residing on the bark or woody tissue of the plant (102). Metabolic capabilities of HS include catabolism of insect, plant, and animal derived sugars, the

latter two being unique when compared to the *S. glossinidius* and close relative *Biostraticola tofi* (98). Genome analysis between HS and *Sodalis* relatives, *S. glossinidius* and *Candidatus* 'S. pierantonius SOPE', reported that the genomes of two recent endosymbionts were actually subsets of the HS genome (98,99,102). Therefore, HS is believed to be the clade's evolutionary precursor whose diverse metabolism facilitated independent descent of *Sodalis* endosymbionts for both animal and plant-feeding insects (98,99,102). Identifying free-living relatives such as HS advances our understanding of how endosymbiotic associations evolve between plant, animal, and insect hosts.

Free-living opportunistic pathogens such as *S. praecaptivus* HS can be found in the soil (103,104) and to survive, they can consume carbon that originates from plant litter, including lignin and lignin-derived aromatics (105,106). These bacteria are part of a larger microbial community that include members that depolymerize lignin to access cellulose or hemicellulose components of plant litter, that consume the lignin-derived monomers, and some members that can do both functions (35). Efforts to elucidate the identity of these groups' members and their mechanisms are of interest to industries that utilize lignocellulose as a raw material, such as paper and biofuel manufacturers (18,19).

By enriching soil consortia on organosolv lignin as a sole carbon source, we aimed to identify novel bacterial isolates with capabilities of lignin depolymerization, catabolism, or both. Organosolv is an ethanol-based separation process of lignin, hemicellulose, and cellulose from woody biomass. This chemical treatment produces a form of lignin closer to its original properties (107), making it

a suitable substrate to determine if bacteria can break down and utilize raw material directly for pulping. To address the economic challenges that are faced with microbial mediated biotechnologies, such as the need for aeration and mixing, we chose to identify anaerobic bacteria that could be applied towards bioreactors (18,23,29).

Here, we describe a novel, free living *Sodalis* species isolated from temperate forest soils (Petersham, MA, USA), and propose the species name *Sodalis* sp. strain 159R. Having the largest genome to date within the *Sodalis* clade, this strain is more diverse in its metabolic capabilities compared to HS and its endosymbiotic relatives, including the genetic potential to catabolize plant derived aromatics such as vanillate and catechol. Genome size and phylogenetic evidence suggest that strain 159R may be an evolutionary precursor to *Sodalis* endosymbionts as well as free-living *S. praecaptivus* HS, consistent with the genomic streamlining observed in the evolutionary adaptation of other organisms to obligate endosymbiosis (99,102,108).

3.3 Isolation and Ecology

Strain 159R was isolated from temperate forest soil (Harvard Forest, Petersham, MA; 42.54N, 72.18W). Soil samples were taken from plots that had been warmed 5°C above ambient temperature for 23 years at the time of collection, along with consortia derived from control plots that were not heated. Three independently adapted consortia each from the heated plots (H16, H15, H8) and control plots (DC13, DC5, DC3) were inoculated anaerobically into Rhizosphere Isolation Media (RIM) (109) containing organosolv lignin as the sole carbon source

instead of acetate. Every four to nine weeks, consortia were diluted 10^{-3} onto fresh RIM for 465 days. To confirm that consortia were viable, headspace gas composition was measured for CO_2 respiration before and after each passage of the community with a Quantek 906 infrared gas analyszer (IRGA; Quantek Instraments, Grafton, MA, USA) (**Fig. 3.1**). At the end of 465 days, DC13, DC5, and H16 were chosen for further analysis based on respiration activity. Direct cell counts using DAPI staining determined that the microbial biomass was 10^5 cells/mL for all three consortia. To obtain isolates, DC13, DC5, and H16 were diluted to 1-5 cells/mL onto a 0.001% five carbon mixture (110) incubated in the dark at 25°C for 6 weeks anaerobically, then streaked onto R2A for colony picking. Purified isolated strains were maintained and routinely grown on the same medium and preserved at -80°C in 20% tryptic soy broth supplemented with glycerol (30 %, v/v).

To screen for lignin depolymerization capabilities, isolates were grown anoxically on R2A plates containing lignin mimicking dyes, malachite green and Congo red (111). An isolate from DC13, strain 159R, displayed no activity for both dyes, indicating that strain 159R may play the role of lignin catabolism rather than depolymerization within the organosolv enriched microbial community (111).

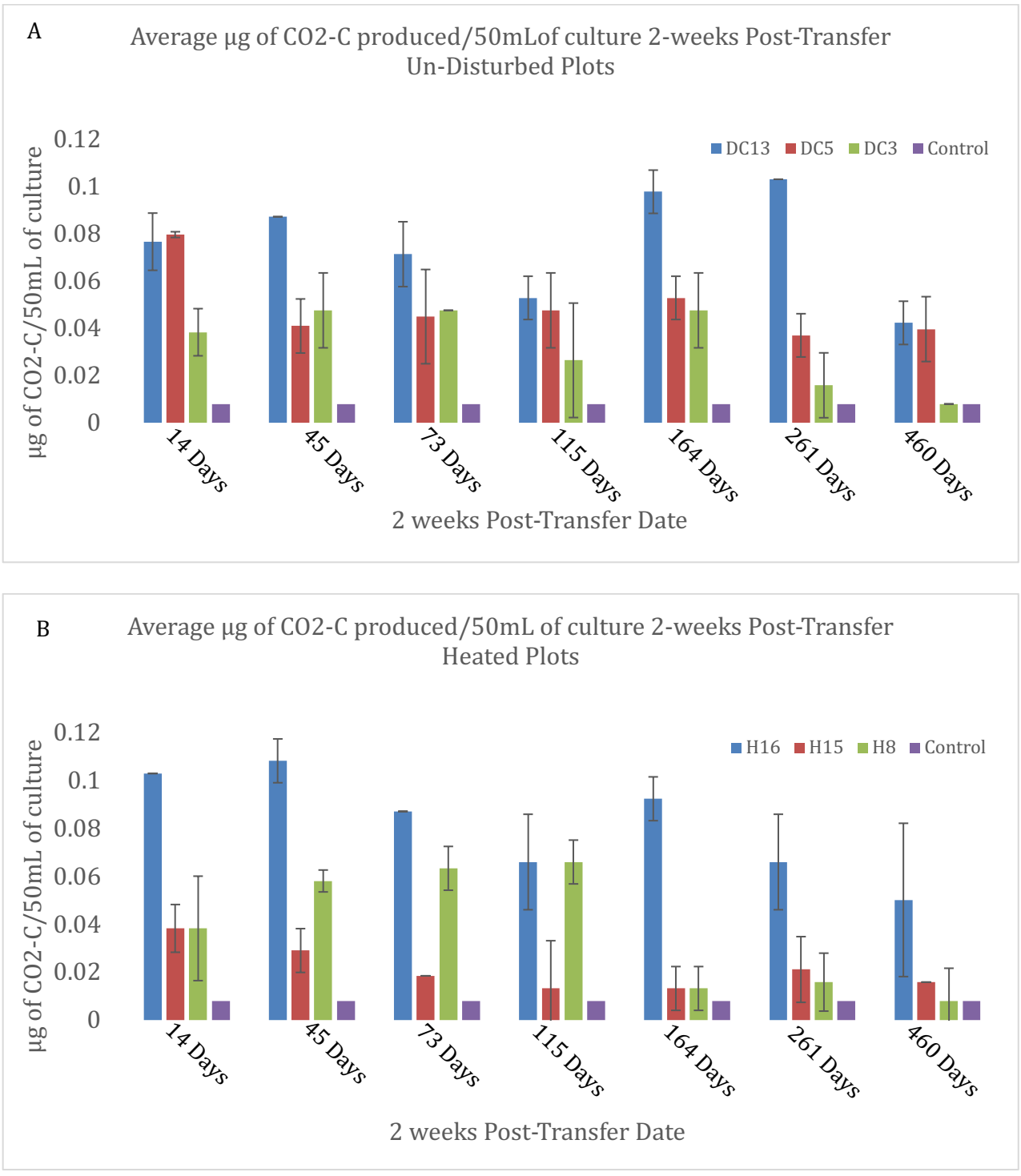


Figure 3.1. CO_2 Respiration measurements (µg of CO2-C/50mL of culture) of microbial consortia from (**A**) control plots, with consortia labelled as DC13, DC5, and DC3, and (B) heated plots, with consortia labelled as H16, H15, and H8. Abiotic controls are in purple. Measurements were taken after 2 weeks of being transferred to new media. Control plot consortia DC13 and D5 as well as heated plot consortia H16 were selected for dilution to extinction culturing experiments.

3.4 Physiology and Chemotaxonomy

Strain 159R formed non-pigmented, opaque circular colonies with shiny surfaces after 24 hours incubating at 30°C on R2A plates. Cells were Gram negative and had rod-shaped cells. Growth at different temperatures (15-42°C) and pH (4-10) under oxic conditions were examined in liquid R2B (**Table 3.1**). Strain 159R could grow between 25-37°C and at 30°C, had an optimal pH of 7. Substrate utilization tests were performed with Biolog micro plates under oxic conditions (Biolog GN2). Carbon substrate utilization of strain 159R was compared to available descriptions of *S. praecaptivus* HS, *S. glossinidius*, and *B. tofi* in literature (92,98,101)(**Table 3.1**).

Table 3.1 Traits characterizing strain 159R (1), *S. praecaptivus* HS (2), *S. glossinidius* (3), and *B. tofi* (4). Data not available: ND.

	1	2	3	4
Catalase	+	+	–	+
α-D-Glucose-1 Phosphate	+	ND	ND	+
α-D-Glucose	+	+	+	+
α-D-Lactose	+	+	–	+
Cellobiose	–	+	ND	+
D-Glucose-6-Phosphate	+	ND	ND	+
D-Fructose	+	+	–	+
D-Galactonic Acid Lactone	+	ND	ND	+
D-Galactose	+	+	–	+

D-Gluconic Acid	+	ND	ND	+
D-Glucuronic Acid	+	ND	ND	–
D-Mannitol	+	+	+	+
D-Mannose	+	+	–	+
D-Serine	+	ND	ND	–
D-Sorbitol	+	+	+	–
D-Trehalose	+	+	–	+
D,L-α–Glycerol Phosphate	+	ND	ND	–
D,L-Lactic Acid	+	ND	ND	–
Glycerol	+	+	-	+
L-Aspartic Acid	+	ND	ND	–
Maltose	+	–	–	–
N-Acetyl D Galactosamine	+	ND	ND	–
N-Acetyl D Glucosamine	+	+	+	+
Pyruvic Acid Methyl Ester	+	ND	+	–
Succinic Acid	+	ND	–	–
Mono-Methyl Succinate	+	ND	ND	–
Sucrose	–	ND	–	–

3.5 Phylogeny and Genomic Features

To confirm purity and to genotype strain 159R, the 16S ribosomal RNA (rRNA) gene was PCR amplified and sequenced using the primer pair 27F (5'-

AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-GGTTACCTTGTTACGACTT-3'). The raw sequence data were checked for accuracy, assembled, and edited using 4Peaks software version 1.8 (Nucleobytes, Aalsmeer, Netherlands). The 16S rRNA gene of strain 159R was then compared using the EZbiocloud service (http://ezbiocloud.net) and the GenBank database to identify its closest relative species. Strain 159R was 96.79% identical to *Sodalis praecaptivus* HS, 96.38% identical to *Sodalis glossinidius*, and 95.97% identical to *Biostraticola tofi*, which is close relative to the *Sodalis*-clade. Because 159R was less than 97% identical in rRNA gene sequence to its closest known relatives, we considered this evidence that strain 159R may be a novel *Sodalis* species (112). Because it also has the potential of anaerobic aromatic metabolism due to its provenance, its genome was chosen for sequencing.

To sequence and annotate the genome of strain 159R, cells were grown on an R2A plate incubated for 3 days under aerobic conditions at room temperature; genomic DNA (gDNA) was extracted using the Qiagen Genomic-tip protocol for bacteria. The draft genome of strain 159R was generated at the DOE Joint Genome Institute (JGI) using the Pacific Biosciences (PacBio) sequencing technology (113). A >10kpb Pacbio SMRTbellTM library was constructed and sequenced on the PacBio RS2 platform, which generated 160,466 filtered subreads totaling 450,943,085 bp. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov. The raw reads were assembled using HGAP (smrt analysis/2.3.0 p5, HGAP 3) (114). The final draft assembly contained 1 contig in 1 scaffold, totaling 6,384,591 bp in size. The input read coverage was 45.4X with a

G+C content of 54.98 mol%. One chromosomal origin of replication, located at 3384 kb, was identified with the oriloc function from the seqnir R package (**Figure 3.1**) (115).

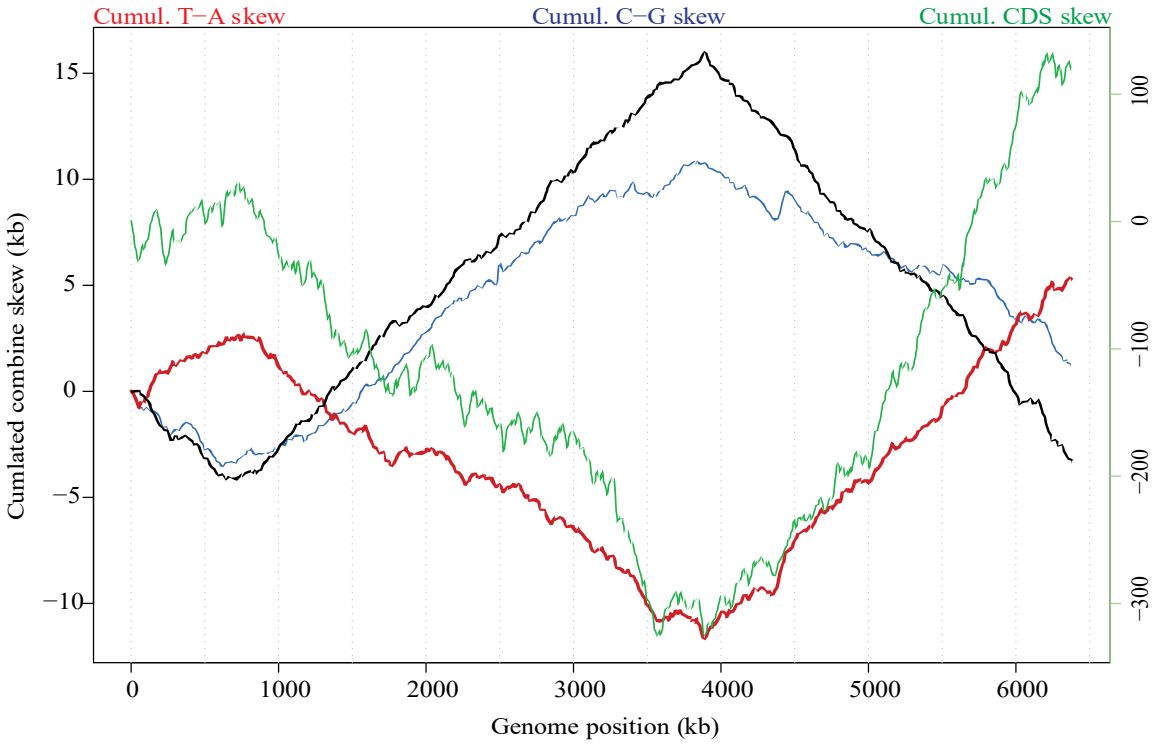


Figure 3.2. Cumulative GC(TA)-skew analysis of *Sodalis* strain 159R using oriloc analysis. The cumulated combine skew is in black, the cumulative GC skew is in light blue, the cumulative TA skew is in red, and the cumulative coding sequences (CDS) skew is in green. The minimum and maximum of GC skew is used to predict the origin of replication at 3384 kb.

To confirm that strain 159R was a novel species, average nucleotide identity (ANI) was calculated using the Pairwise ANI tool from DOE JGI IMG/M (116). Strain 159R was 76-77% similar to the available *Sodalis* genomes (**Table 3.2**), well below the accepted 95-96% species threshold (117). In addition, the estimated genome-sequence based digital DNA–DNA hybridization (dDDH) values were calculated with

the Genome-to-Genome Calculator (GGC) software version 2.1, developed by DSMZ, using the formula 2 option as recommended (http://ggdc.dsmz.de/ggdc.php). Strain 159R was less than the 70% species boundary compared to all available genomes of the *Sodalis*-allied clade (**Table 3.2**) (117). Therefore, both ANI and dDDH values supported that strain 159R is a novel *Sodalis* species.

To test whether or not strain 159R is a novel genus or a member of the genus *Sodalis*, the percentage of conserved proteins (POCP) was calculated comparing strain 159R to *Sodalis* members as well as *B. tofi* (118,119). POCP estimates genus demarcation between two organisms based on proteins that are shared (118). If the POCP is greater than or equal to 50%, then the organisms are considered to be within the same genus. Strain 159R had a POCP of 51.27% to *Candidatus* 'Sodalis pierantonius SOPE', a POCP of 61.87% to *B. tofi*, and a POCP of 63.68% to *S. praecaptivus* HS. All other POCP values were less than 50% when comparing strain 159R to other *Sodalis* members (**Table 3.2**); however these low POCP values can be explained by genome degeneration of the *Sodalis* endosymbionts (118). POCP results supported that strain 159R is within the same genus as *S. praecaptivus* HS as well as *Candidatus* 'Sodalis pierantonius SOPE', which is one of *S. praecaptivus*' closest relatives and similar in genome size.

Table 3.2. Genome size, average nucleotide identity (ANI), average amino acid identity (AAI), digital DNA–DNA hybridization (dDDH), and percentage of conserved proteins (POCP) estimates comparing Sodalis sp. strain 159R (6.38Mbp) to the *Sodalis*-allied clade and closest relatives based on 16S rRNA genes.

Organism	Genome Size (Mb)	ANI%	AAI%	dDDH estimate % (GLM-Based)	POCP %
Sodalis praecaptivus HS	5.15	78.97	72.55	21.50	63.68

Biostraticola tofi DSM 19580	4.29	78.71	73.30	20.80	61.87
Candidatus *Sodalis pierantonius* SOPE	4.51	79.10	73.23	21.40	51.27
Candidatus *Sodalis* sp. SoCistrobi 3249	3.06	79.06	74.60	20.90	48.82
Sodalis sp. TME1	3.41	79.35	73.70	22.00	42.65
Sodalis glossinidius mositans	4.29	79.22	71.37	22.00	41.95
Sodalis-like endosymbiont of *Proechinophthirus fluctus*	2.17	78.94	68.41	22.30	29.77
Sodalis-like symbiont of *Philaenus spumarius* PSPU	1.38	78.96	74.97	21.80	29.13

To further elucidate the phylogenetic position of strain 159R, a phylogenetic tree was constructed using maximum likelihood algorithm via the KBase app Insert Genome Into Species Tree 2.1.10 (120). This KBase program combines genomes provided by the user with a set of closely related genomes selected from all public KBase genomes. Based on alignment similarity to a select subset of 49 COG (Clusters of Orthologous Groups) domains, the phylogenetic tree is then reconstructed using FastTree (version 2.1.10; **Figure 3.2a**) (121). The evolutionary relationship of *Sodalis* sp. strain 159R was also calculated with PhyloPhlan (122). This analysis uses 400 conserved proteins across the bacterial domain to produce a phylogeny using the maximum likelihood inference approach. Visualization and editing of both trees were completed with iTol software version 3 (**Figure 3.2b**) (73). With strong bootstrap support for both trees, *Sodalis* sp. strain 159R was positioned as a basal member to the *Sodalis*-allied clade, further supporting the ANI and DHH values that strain 159 is a novel *Sodalis* species. However, strain 159R was also more distantly

related to the *Sodalis*-allied clade than *Biostraticola tofi*, which suggested that strain 159R could be a novel genus.

To elucidate the evolutionary relationship between strain 159R and *B. tofi*, we investigated the previous characterization of *B. tofi.* It was noted by Verbarg and colleagues that the decision to make *B. tofi* a new genus was due to the "distant phylogenetic position as compared to any other representative of the [*Enterobacteriaceae*] family and the significant phenotypic differences to its nearest phylogenetic neighbor, *Sodalis glossinidius*" (101). The phenotypic differences that were observed between *B. tofi* and *S. glossinidius morisitans* were most likely due to the smaller genome size of the latter compared to that of its evolutionary precursor, *S. praecaptivus* HS. Therefore, we calculated the POCP between *B. tofi* and HS to see if *B. tofi* may be a *Sodalis* member. The POCP was 68.44%, supporting that *B. tofi* is very similar to the *Sodalis*-allied clade and potentially is not a separate genus; however additional comparisons of genomic features and physiology between *B. tofi* and HS would be needed to confirm this. Overall, the POCP results confirmed that strain 159R was a member of the *Sodalis* genus and likely an ancient precursor to the *Sodalis*-allied clade and potentially, *B. tofi*.

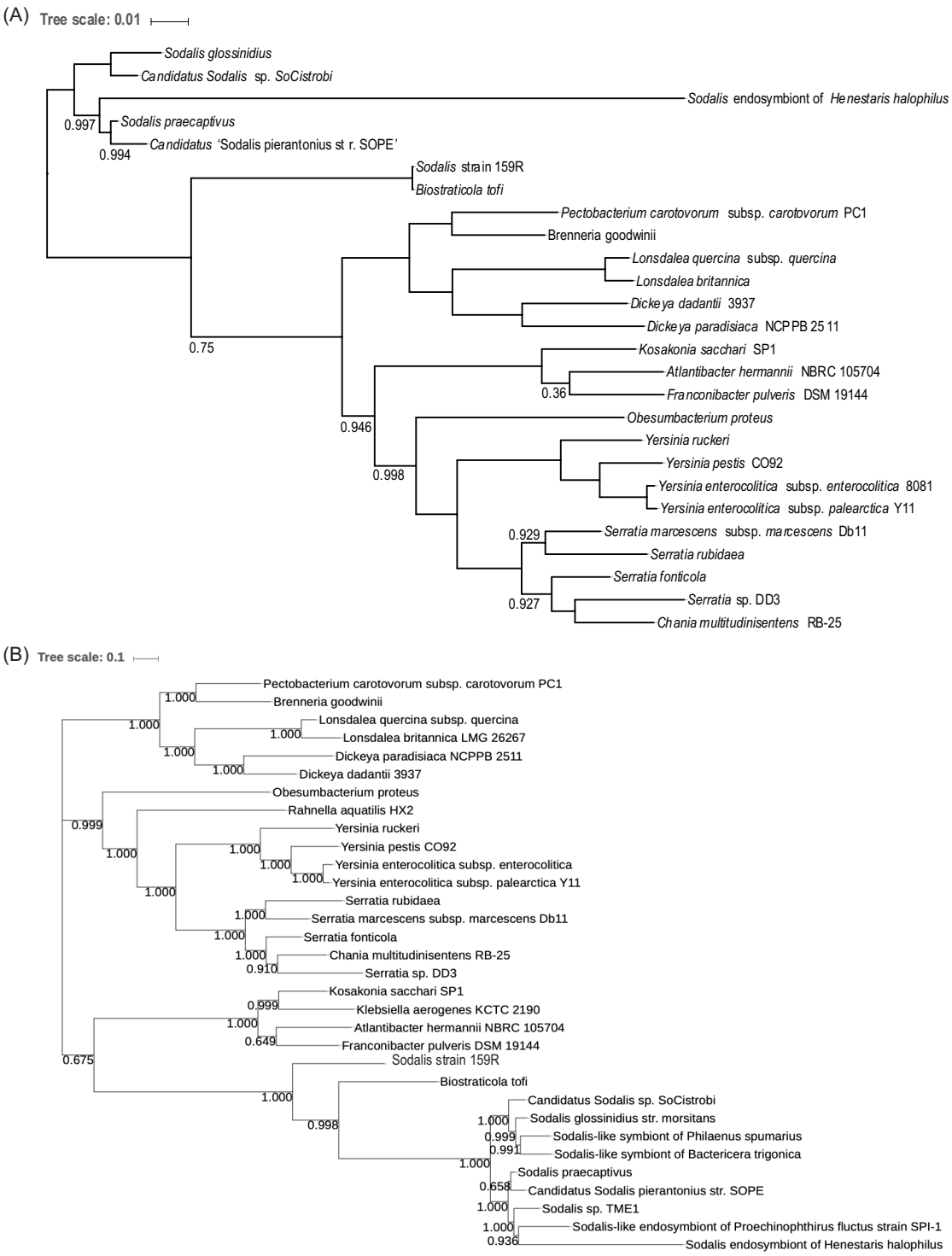


Figure 3.3. Reconstruction of the phylogenetic position of strain 159R based on (A) COG similarity using KBase's Insert Genome Into Species Tree 2.1.10 and (B) 400 conserved protein sequences using PhyloPhlan. Both trees are presented as maximum-likelihood trees with bootstrap values.

To further determine if strain 159R may be an evolutionary precursor to both the free-living and endosymbiont members of *Sodalis*, synteny analysis was completed with SyMap (123) between the chromosomes of 159R and *S. praecaptivus* HS; 159R and *S. glossinidius morisitans*; 159R and *Candidatus* 'Sodalis pierantonius SOPE'; and 159R and *B. tofi* (plasmids were excluded from the analysis; **Figure 3.3**). As expected, the endosymbiont *Sodalis*-clade members as well as free-living species, *S. praecaptivus* and *B. tofi*, shared high synteny with the strain 159R genome, with 92% of *S. praecaptivus'* genome, 91% of *S. glossinidius morisitans*' genome, 71% of *B. tofi*'s genome, and 54% of *Candidatus* 'Sodalis pierantonius SOPE''s genome being syntenic. It has been previously seen that SOPE has had many rearrangements compared to HS, and therefore would have lower synteny to strain 159R as it does with HS (99). Synteny block coverage was greater in genomes of *Sodalis* members and *B. tofi* compared to strain 159R, suggesting that the former are subsets of the strain 159R genome. This evidence supports the notion that 159R is an evolutionary precursor to the *Sodalis*-clade, including free-living HS.

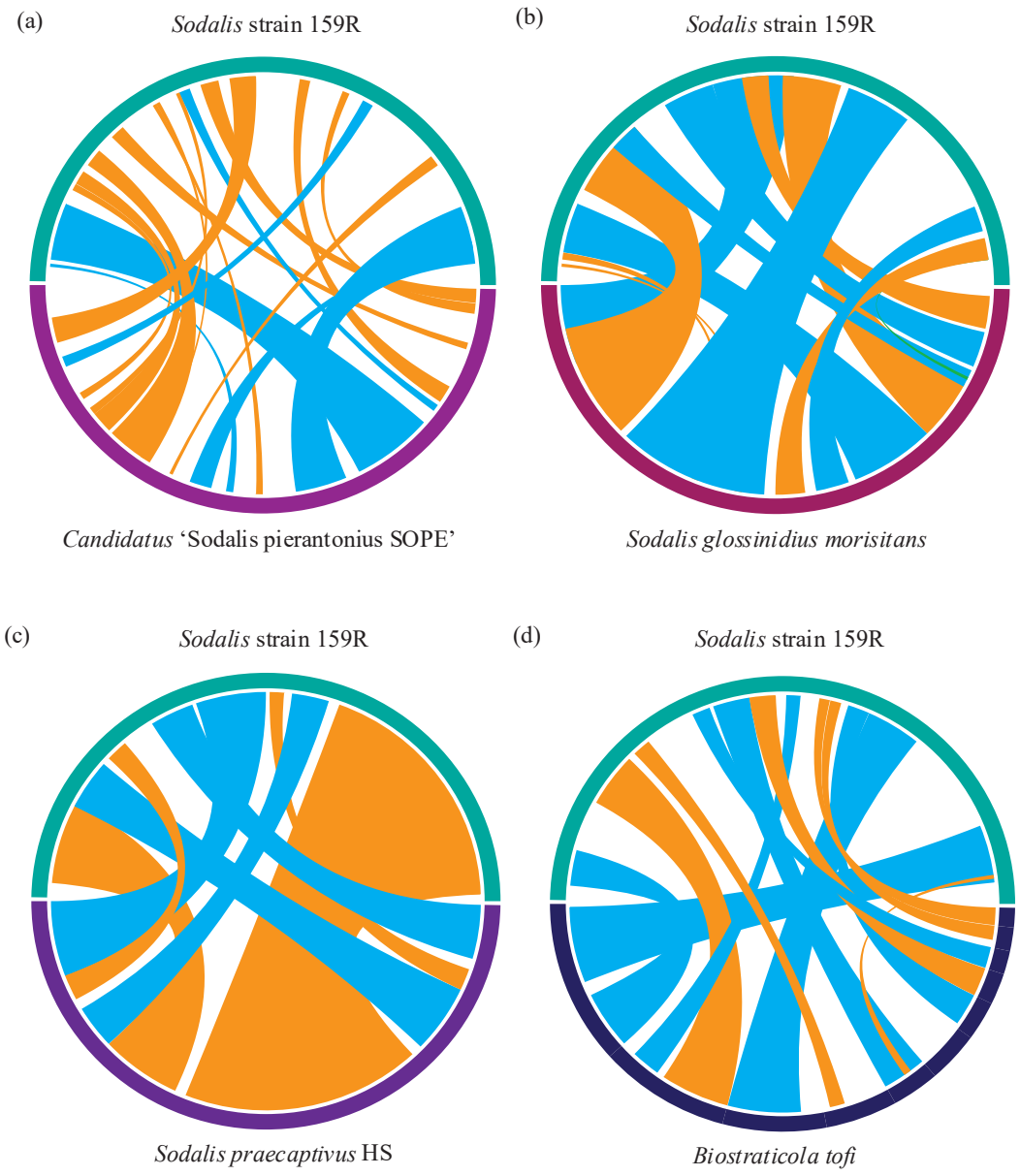


Figure 3.4. Synteny analysis comparing the chromosome of strain 159R (in teal) to chromosomes of (a) *Candidatus* 'Sodalis pierantonius SOPE', (b) *Sodalis glossinidius morisitans*, (c) *Sodlais praecaptivus* HS, and (d) *Biostraticola tofi*. Direct blocks of synteny are represented in orange and inverted blocks are represented in light blue. Blocks of synteny account for a larger portion of the *Sodalis*-clade members and *B.* tofi than that of strain 159R, suggesting that genomes are a subsets of strain 159R.

The genome of strain 159R consists of 5,684 predicted coding sequences. For energy production, strain 159R has genes encoding for aerobic respiration as well

as NarGHI for nitrate reduction as seen in the core genomes of *Sodalis praecaptivus* HS, *Candidatus Sodalis pierantonius* SOPE, and *Sodalis* TME1 (124). Strain 159R also contains a sulfide dehydrogenase, suggesting that strain 159R is a sulfate-reducing bacterium unlike the other *Sodalis*-clade members. When compared to the available 6 genomes of the *Sodalis*-allied clade as well as *B. tofi* in IMG JGI Phylogenetic Profiler (**Table 3.2**), 2,012 genes are unique to *Sodalis* sp. strain 159R, with 1,179 genes assigned COG IDs (**Figure 3.4**).

When compared to the *Sodalis*-clade and *B. tofi* genomes, the largest group of unique genes present in strain 159R are those relating to transcription (222 genes). This corroborates with previous evidence that free living organisms tend to be enriched for transcription regulators in comparison to many endosymbionts due to the need to adapt to ever changing environmental conditions (125). Similarly, the second largest group of unique genes present in strain 159R were those relating to carbohydrate transport and metabolism (193 genes), likely required to adapt to the varying availability of metabolites found in the soil environment compared to the limited nutrient availability in a host (126). This group of unique genes included those associated with lignocellulose degradation, such as a GH43 family β-xylosidase and a feruloyl esterase, as well as genes for cell uptake and utilization of aromatic monomers. Enzymes included a 4-hydroxybenzoate transporter-like MFS transporter, nine glutathione S-transferases, four catechol 2,3-dioxygenase enzymes, a salicylate hydroxylase, a vanillate O-demethylase monooxygenase (*vanA*), a vanillate O-demethylase ferredoxin subunit (*vanB*), and a 4-carboxymuconolactone decarboxylase.

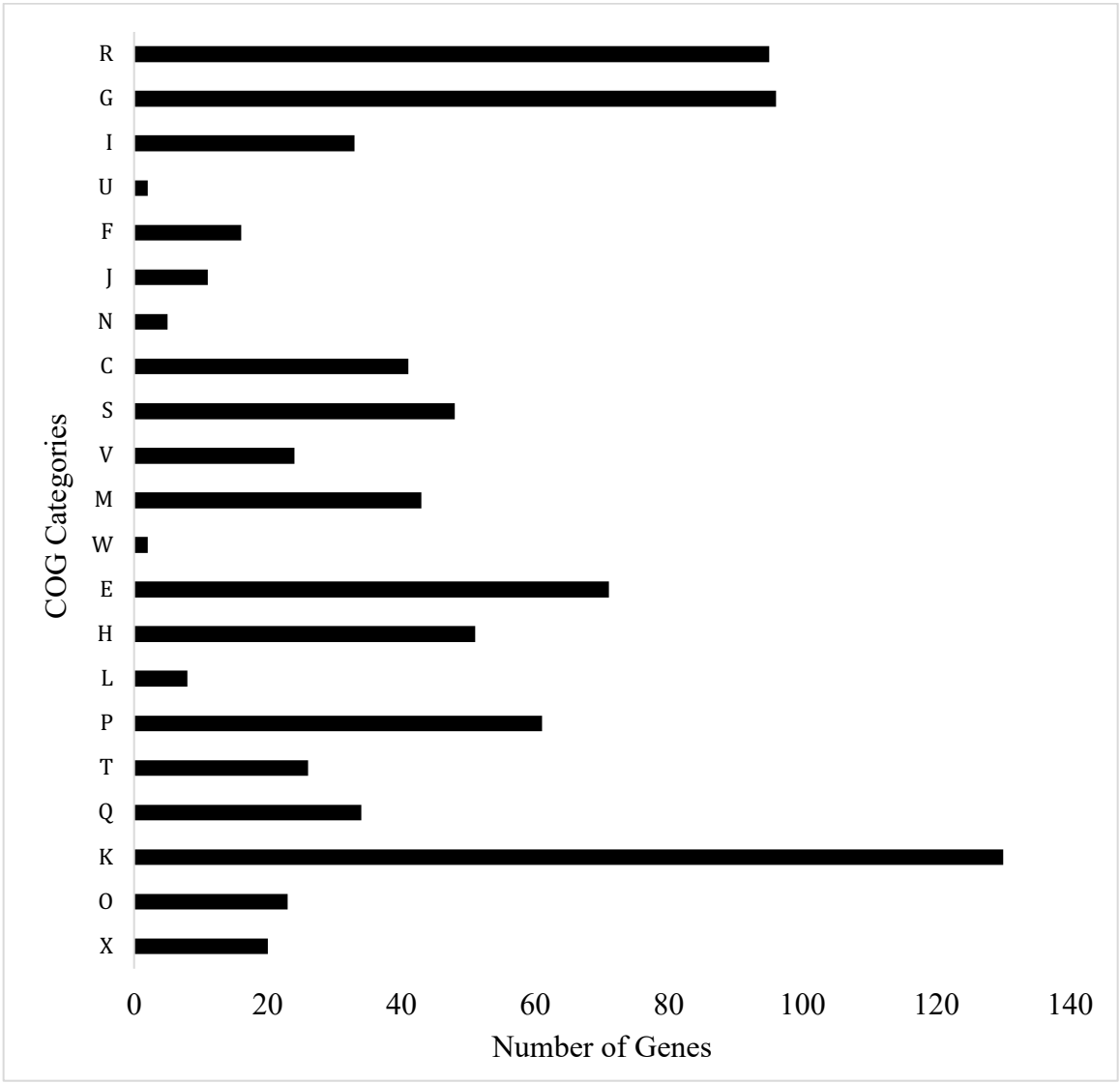


Figure 3.5. Unique gene abundance for strain 159R compared to all organisms listed in Table 3.2 based on COG category. Abbreviations are as follows: (X) Mobilome: prophages, transposons; (O) Posttranslational modification, protein turnover, chaperones; (K) Transcription; (Q) Secondary metabolites biosynthesis, transport and catabolism; (T) Signal transduction mechanisms; (P) Inorganic ion transport and metabolism; (L) Replication recombination and repair; (H) Coenzyme transport and metabolism; (E) Amino acid transport and metabolism; (W) Extracellular structures; (M) Cell wall/membrane/envelope biogenesis; (V) Defense mechanisms; (S) Function unknown; (C) Energy production and conversion; (N) Cell motility; (J) Translation, ribosomal structure and biogenesis; (F) Nucleotide transport and metabolism; (U) Intracellular trafficking, secretion, and vesicular transport; (I) Lipid transport and metabolism; (G) Carbohydrate transport and metabolism; (R) General function prediction only.

3.6 Lignin Metabolic Potential

We investigated the genetic potential of strain 159R for anaerobic lignin degradation as well as aromatic catabolism, which could be applied towards secondary chemical and biofuel production from lignocellulosic material (127). Using HMMER, genes selected as markers for anaerobic aromatic metabolism were compared to 159R genome (**Table 3.3**). Substrates included benzoyl-CoA, 3-hydroxybenzoyl-CoA, 3-methylbenzoyl-CoA, 4-methylbenzoyl-CoA, hydroxyhydroquinone, resorincol/α-resorcylate, and phloroglucinol. Between benzoyl-CoA and its analogs, enzymes in strain 159R were most homologous to anaerobic 3-hydroxybenzoyl-CoA pathway, having a homolog to a 3-hydroxybenzoyl-CoA enoly-CoA hydratase (Gene ID 2788604060; E-value = 9.4e-55) and a homolog to 3-hydroxybenzoyl-CoA hydroxyacyl-CoA dehydrogenase (Gene ID 2788603217; E-value = 1.9e-55). There were no homologs for hydroxybenzoyl-CoA acyl-hydrolase, which could suggest the presence of an alternative, and possibly novel, enzyme that funnels hydroxybenzoyl-CoA into the central metabolism. Additionally, strain 159R contained a homolog to phloroglucinol reductase (phloroglucinol pathway; Gene ID 2788602949, E-value = 6.7e-53) as well as homologs to α-resorcylate hydroxylase large (Gene ID 2788606053) and small subunits (Gene ID 2788606054) with E-values 3.3e-48 and 4.5e-44, respectively. Results suggest that 159R is capable of metabolizing aromatics under anaerobic conditions and should be further studied to determine other enzymes and pathways that may be present, including peripheral pathways

that were not investigated here, such as 4-hydroxybenzoyl-CoA that had substrate specific transporters present in the genome.

In addition to anaerobic aromatic metabolism, enzymes annotated for aerobic aromatic metabolism in the strain 159R genome included 4,5-DOPA dioxygenase extradiol (LigB) as well as homologs with >40% sequence identity to *ligF*, *ligJ*, *ligK*, *ligR*, and *ligV* genes that are also found in aerobic lignin degrader, *Sphingomonas paucimobilis* SYK-6 (128) (**Table 3.4**). Strain 159R also contains genes for the catechol degradation pathway. The genomic potential of lignin degradation and aromatic catabolism under both anaerobic and aerobic conditions are listed in **Table 3.3-3.5**.

Table 3.3. HMMER Marker Enzymes for Anaerobic Aromatic Metabolism. Enzyme name is on the far let column followed by either the sequences used to build the profile Hidden Markov Model (HMM) with HMMER hmmbuild program or NCBI GenBank protein ID used for HMMER jackhmmer program. When applicable, subunits are listed separately.

Benzoyl-CoA Pathway	
Enoyl-CoA Hydratase	>SP|O87873|DCH_THAAR/7-257 Cyclohexa-1,5-dienecarbonyl-CoA hydratase [Thauera aromatica] LKVWLERDGSLLRLRLARPKANIVDAAMIAAMRQALGEHLQAPALRAVLLDAEGPHFSFGASVDEHMPDQCAQM LKSLHGLVREMLDSPVPILVALRGQCLGGGLEVAAAGNLLFAAPDAKFGQPEIRLGVFAPAASCLLPPRVGQACAE DLLWSGRSIDGAEGHRIGLIDVLAEDPEAAALRWFDEHIARLSASSLRFAVRAARCDSVPRIKQKLDTVEALYLEEL MASHDAVEGLKAFLEKRSANWENR >RF|YP_385104.1/6-256 enoyl-CoA hydratase/isomerase [Geobacter metallireducens GS-15] LKVWLEKDGALLRLRLARPKANIVDAAMIAALQAALTEHLPSAKLRAVLLDAEGPHFSFGASVEEHMPESCAAML QSLHALVIQMLESPVPVLVAVRGQCLGGGLEVVAAGNLIFAAPGAMLGQPEIKIGVFAPAASCLLPERIGKTASEDL LFSGRSITAEEGFRIGLVTAVAEDPEQAAVAYFDEHLAGLSASSLRFAVRAARIGVLERTKTKIAAVEKLYLEELMA THDAVEGLNAFLGKRPAAWQDR >RF|YP_421505.1/9-259 Cyclohexa-1,5-dienecarbonyl-CoA hydratase [Magnetospirillum magneticum AMB-1] LKVWKDREGKLLRLRLSRPKANIVDAEMIAALSAALGDAHEDSALRAVLIDHEGPHFSFGASVAEHMPDQCAAML ASLHKLVIAMVDFPLPILVAVRGQCLGGGLEVALAGHMMFVSPDAKLGQPEIVLGVFAPAASCLLPERMPRVAAE DLLYSGRSIDGAEAARLGIANAVVDDPENAALAWFDNGPAKHSAASLRFAVKAARLGMNERVKAKIAEVEALYL NGLMATHDAVEGLNAFLEKRPALWEDR
Hydroxyacyl-CoA Dehydrogenases	>OMNI|NTL01AE3009/8-353 TWQMTEPGK-LQKTRVPMPELGSGDVVVKIAGCGVCHTDLSYFYMGVPTVQKPPLSLGHEISGTII--- GGEASMIGKEVIVPAVIPCGECELCKTGRGNRCLAQKMPGNSMGIYGGYSSHIVAQSKYLCVVEN---- RGDTPLEHLAVVADAVTTPYQAAVRADLKKDDLVIVVGAAGGVGSFMVQTAKGMGAKAVIGIDINEEKLEMMK GFGADFIINPKDK-SAKEVKELFKGFCKE RGLPSNYGWKIFEVTGSKPGQELALSLLSFTGKLVIVGYGTAETNYMLSKLMAFDAEIIGTWGCPPDRYAAVRDMC LDGRIQLGPFVETRPMSQIEHVFDEAHHGKLKRRVILTP >gi|19571180/20-368 6-hydroxycylohex-1-en-1-carbonyl-CoA dehydrogenase [Thauera aromatica] RWMMTSPGAPMVRAEFEIGELSADQVVVAVAGCGVCHTDLGYYYDSVRTNHALPLALGHEISGRVVQAGANAA QWLGRAVIVPAVMPCGTCELCTSGHGTICRDQVMPGNDIQ-- GGFASHVVVPARGLCPVDEARLAAAGLQLADVSVVADAVTTPYQA VLQAGVEPGDVAVVIGV-GGVGGYAVQIANAFGA-SVVAIDVDPAKLEMMSKHGAALTLNAREI- SGRDLKKAIEAHAKANGLRLT- RWKIFECSGTGAGQTSAYGLLTHGATLAVVGFTMDKVEVRLSNLMAFHARALGNWGCLPEYYPAALDLVLDKKI DLASFIERHPLDQIGEVFAAAHAHKLTRRAILTP >OMNI|NTL06MM2144/25-374

	RWMMTGVGQPMVKEAMEIAAPGAGEVLVEVAGCGVCHTDLDYYYNGVRTNHALPLALGHEISGRVIQAGAGAE SWVGKAVIISAVIPCGQCDLCKRGKGTICRSQKMPGNDLQ-- GGFATHITVPANGLCAVDEARLKAAGLELSEVSVVADALTTPYQAAVQAGIGQGDLVIVIGC- GGVGGYSVQVASAMGA-TVVALDIDPVKLEAVKAAGAKLTLNPKDFPSTREIKKEIGAFAKAQGLRST- EWIIMECSGSVPGQQSAFDLMVHGCTICVVGYTMNKAEFRLSNLMAFHARALGNWGCPPDLYPGALDLVLSGKIN VKNFVERRPLDSINDTFAAVHDHKLSRRAVLCP
Oxoacyl-CoA hydrolase	>OMNI|NTL01AE3010/12-371 IKDHALMGEEHFGTEAPSVL- FEKRPVTDPQGNVVPGLYAAWIILNNPKQYNSYTTEMVKAIIAGFQRASSDRTIVAAVFTAVGDKAFCTGGNTAEYA SYYAQRPNEYGEYMDLFNAMVDGILNCKKPTICRVNGMRVGGGQEIGMATDLTITSDMAIFGQAGPKHGSAPDGGS TDFLPWMLNMEDAMYNCISCEPWSAYKMKSKNLITKVVPVLKKDGEWVRNPLVRTDAYVDD-GELV YGEPVAADKAKAAKELIAQCTTDFAKLDEAVDALVWKFANLFPQCLIKSIDGIRGKKKFFWDQMKLANRHWLAAN MNHEAYLGFTAFNN-KKATGKDVIDFIKFRQLVAEGHAFDDAFAEQVL >OMNI|NTL01GM2088/16-376 LNDHNLIDRE- VESLCDGMVKYEKRPAKRHDGSVAEGIYNAWIILDNPKQYNSYTTDMVKAIILAFRRASVDRSVNAVVF TGVGDKAFCTGGNTKEYAEYYAGNPQEYRQYMRLFNDMVSAILGCDKAVISRVNGMRIGGGQEIGMACDFSIAQD LANFGQAGPKHGSAAIGGATDFLPLMVGCEQAMVSGTLCEPFSAHKAARLGIICDVVPALKVGGKFVANPTVVTDR YLDEYGRVVHGEFKAGAAFKEGQGQIKEGEIDLSLLDEKVESLCTKLLETFPECMTKSLEELRKPKLHAWNLNKENS RAWLALNMMNEA RTGFRAFNEGTKETGRE-IDFVKLRQGLAKGTPWTEELIESLM >OMNI|NTL06MM2143/17-372 LNDHNLV----PTTVVPGVL- YEKRPAKRADGTVAEGLYNAWITLDNQKQYNSYTTDMVKGVIMAFRDASNARDVSSVVF TGAGDKAFCTGGNTKEYAEYYAGNPQEYRQYMRLFNDMVSAILGCDKPVICRVNGMRIGGGQEIGMAADFSVAQD LAKFGQAGPKHGSAPIGGATDFLPVMIGCEQAMVSGSLCEPWSAHKAYRTGIIMDLVPALKVDGKFVANPLVITDR YLDEFGKIVHGESKTGAELAAGKELLKKGTIDLSLLDAKVEEICAKILHTFPDCFTKTIQELRKPKLNAWNANKENSR DWLGLNMMTEARTGFRAFNEGPKE-DRE-IDFVALRQALAKGAPWTPELIESLI >gi|3724166/17-373 LVDHNLV----PETVCPGVL- YEKRPARNLKGEVVPGLYNVWISLDNPKQYNSYTTDMVKGLILAFRAASCARDVASVVF TAVGDKAFCTGGNTKEYAEYYAGNPQEYRQYMRLFNDMVSAILGCDKPVICRVNGMRIGGGQEIGMAADFTVAQ DLANFGQAGPKHGSAAIGGATDFLPLMIGCEQAMVSGTLCEPFSAHKANRLGICMQIVPALKVDGKFIANPLVVTDR YLDEFGRIIHGEFKTGDELAAGKELMKRGEIDLSLLDEAVEKLCAKLISTFPECLTKSFEELRKPKLDAWNRNKENSR AWLALNMMNEARTGFRAFNEGNKETGRE-IEFTDLRQALAKGMPWTPELIESLM
3-Hydroxybenzoyl-CoA Pathway	

Enoyl-CoA Hydratase	>WP_050418522.1 enoyl-CoA hydratase/isomerase family protein [Azoarcus sp. CIB] MISLRIEDS----------VATVTLCRAPV-NAINEEWIAAFDRILAELEHTPRVNVL WIRSAERVFCAGADL-DVIGSLFATEAGRVQMIAITRRMQQLYARLERLPQVTVAEIGGA AMGGGFELALACDLRVVADSAKVGLPEARLGLLPAA-GGTQRMTRICGEAVARRLILGAE VVGGVDAVKLGCAHWVAPAAELEEFTRGVVTRIAALPALALSECKRCITVAVEGD-EDGY QVELAGSAALLADGETQQRVRAFLNR-------- >WP_011236223.1 enoyl-CoA hydratase/isomerase family protein [Aromatoleum aromaticum] MISLTIEAS----------VATVTLCRSPV-NAINEEWIEQLDRILAEIERTPRVNVL WIRSGERVFCAGADL-ELIRSLFDSETGRRQMIAMTRRMQEVYARLERLPQVSVVEIGGA AMGGGFELALACDLRVVADSARIGLPEARLGLLPAA-GGTQRMTRICGEAVARRLILGAE VIGGAEAVALGCAHWVAPAAELESVARAVVERIAALPGTALAECKRCIDVAVAAE-ENGF EVELSGSAALLADAETQRRVQRFLDKQRQ----- >CAC28159.1 putative hydrolase [Thauera aromatica] MSVVLVEQPTPD---------VAVVRLNRPDARNALNQEVRSALAEHFDRLGQAAEVRCI VLTGGERCFAAAPDIRAM------ADAG--AIEIMLRQTQRLWQAIAACPKPVIAAVNGY AWGGGCELAMHADIIIAGEGASFCQPEVKVGIMPGA-GGTQRLTRAVGKFQAMKMVLTGL PVSARERLAMGLASEVVADDAVQARALELARHIATLPPLAIAQIKEVLLAGQDASLDTAL MLERKAFQLLFASADQKEGMRAFLEKRPPVFRGG >CAC28155.1 unnamed protein product [Thauera aromatica] MYKLKAADWHPEHFKLEVANRVATITLNRPDRKNPLTFESYAELRDTFHKFQYVDDVRSI VITGAGGNFCSGGDVHDIIGPLTKMDMN--GLLTFTRMTGNLVKEMRTCPQPIISAIDGI CAGAGAIVSMASDMRYATPDAKTAFLFVRVGLAGCDMGACAILPRIIGHGRASELLYTGR VMSAQEGQAWGYFNDLVAPDQVLAKAQEMALSLANGPAFAHAMTKKCLHQEWDMSIEQAL ETEAEAQAICMQTQDFTRAYNAFVAKQKPVFEGN >WP_050418021.1 enoyl-CoA hydratase family protein [Azoarcus sp. CIB] MYKLKAAEWRPEHFKLEVADRVATITLNRPERKNPLTFESYAELRDTFIKLQYAEDVRAV VMTGAGGNFCSGGDVHDIIGPLTKMDMT--GLLAFTRMTGNLVKEMRNCPQPIISAVDGV CAGAGAIITMASDLRYATPEAKTAFLFVRVGLAGCDMGACSILPRIIGQGRASELLYTGR SMSAEEGRAWGYFNDVVPAEKVLAKAQEMALSLANGPAFAHSVTKKCLHQEWNQTIEQAL ETEAEAQAICMQTEDFTRAYNAFVNKQVPKFEGN >WP_011236224.1 enoyl-CoA hydratase family protein [Aromatoleum aromaticum] MYKLKAAEWRPEHFKLEVADRVATITLNRPERKNPLTFESYAELRDTFHKLQYVDDVRTV VITGAGGNFCSGGDVHDIIGPLTKMDMN--GLLTFTRMTGNLVKEMRNCPQPIISAVDGI CAGAGAIVSMASDLRYATPEAKTAFLFVRVGLAGCDMGACSILPRIIGHGRASELLYTGR SMSAEEGRAWGYFNDIVPAEKVLGRAQEMALSLANGPAFAHSMTKKCLHQEWNQTIEQAL ETEAEAQAICMQTQDFTRAYNAFVNKQVPKFEGN

Hydroxyacyl-CoA Dehydrogenases	>WP_050418028.1 SDR family oxidoreductase [Azoarcus sp. CIB] MTADSGRALAGKHVVITGGGRGIGAAIAAALSAQGARLTLMGRNRGQLEER--AAVLRTL GGESCEVHCEAVDVADEASVVSAFAAAAKRLGPVAVLVNNAGQAGSAPFLRTESALWQQM LAVNLTGTYLATRAALPDMLAAG-WGRIINVASTAGEKGYPYVTAYCAAKHGVIGLTRSL ALELAHKHVTVNAVCPGYTDTDIVRDAVTNIREKTGRSEAEALAELAKHNPQGRLVRPEE VANAVLWLCLPGSDAITGQAISVSGGEVM-- >CAC28156.1 putative alcohol dehydrogenase [Thauera aromatica] --MTHSRALSGKHAVITGGGRGIGAAIAHSLAEQGAAVTLMGRTLPRLEQQ--AEELRAF SQ----VHCEAVDVAQADSVAAAFAAAQARLGPVDILVNNAGQALSAPFVKTDPALWQQM LDVNLTGVFLGTRAVLPGMLAAG-WGRVINITSTAGQKGYPYVSAYCAAKHGVIGLTRAL ALETARKNVTVNAVCPGYTDTDIVRDSVSNIQTKTGRSEAEALAELTRFNPQGRLVRPQE VANAVLWLCLPGSEAITGQSISVAGGEMM-- >WP_041646819.1 SDR family oxidoreductase [Aromatoleum aromaticum] -----MRELSGKHAVVTGGGRGIGAAIAQRLAEQGACVTLMGRRREPLEER--ADALRAL IGVHCDMHCEAVDVADPASVAAAFDAAARRFGPVSILVNNAGQASSAPFVKTDLALWQRM LDVNLTGTYLGTKAVLSGMLAAG-WGRIVNVASTAGQKGYPYVSAYCAAKHGVIGMTRAL ALELAQKNITVNAVCPGYTDTDIVREAITNIRAKTGRSEAEAQGELAKHNPQGRLVRPDE VANAVLWLCLPGAEAITGQAISVSGGEVM-- >CAC28154.1 putative alcohol dehydrogenase [Thauera aromatica] ------MRLEGKTAVVTGGASGIGRATAETLAAAGAHVVI-----GDLDQEKGAAVAAAI RESGRKADYFPLDVTSLDSVGVFAKAVEENGLEVDIVVNVAGWGKIQPFMENSPDFWRKV IDLNLLGPVAVTHAFLGGMIARGRGGKVITVASDAGRVGSTGETVYSGAKGGAIAFGKAL AREMARYKINVNSVCPGPTDTPLLAAVPEKHQE----------AFVKATPMRRLGKPSE IADAVLFFASSDSDFITGQVLSVSGGMTMVG >WP_011236225.1 SDR family oxidoreductase [Aromatoleum aromaticum] ------MRLDGKTAVVTGGASGIGLATAETLARAGAYVLI-----GDIDEQKGAAVAGAL CEQQLGVDFIRLDVTDLDSIAAFKDEAYRRRPQIDIVANVAGWGKIQPFMENTPDFWRKV IDLNLLGPVAVSHAFLPQMIERG-AGKIVTVASDAGRVGSLGETVYSGAKGGAIAFTKSL AREVARYNINVNCVCPGPTDTPLLQAVPEKHRE----------AFVKATPMRRLAKPSE LADAVLFFASDRASFITGQVISVSGGLTLAG >WP_050418022.1 SDR family oxidoreductase [Azoarcus sp. CIB] ------MNLQGKTAVVTGGASGIGYATAETLARAGAKVVI-----GDIDAAKGAAAAGML AEQHLDVDFVRLDVTDIDSIHAFRDETYRRHPQVDIVANVAGWGKIQPFMENTPDFWRKV IDLNLLGPVAVSHAFLQQMIERG-SGKIVTVSSDAGRVGSLGETVYSGAKGGAIAFTKSL AREVARYNINVNCVCPGPTDTPLLQAVPEKHRE----------AFVKATPMRRLAKPSE LADAVLFFASDRASFITGQVISVSGGLTLAG

Oxoacyl-CoA hydrolase	>CAC28157.1 putative acyl-CoA dehydrogenase [Thauera aromatica] MSEKSYLEWPFFEDRHRKLEAELDSWATNNISEHH-GELDSACRELVAKLGAAGWLRYCV GGTSYGGEHETIDTRSICLLRETLARHSGLADFAFGMQGLGSGAITLHGSDAQKREYLPR VASGQALAAFALSEPGSGSDVAAMACSARLDGEYYVLDGEKSWISNGGIADFYVVFARTG EAPGARGLSAFIVDADTPGLEIAERIEVIAPHPLARLRFTDCRVHKSAMLGTPGLGFKVA MQTLDIFRTSVAAAALGFSRRALDEALRRATTREMFQQKLADFQITQVKLAQMATSVDIS ALLTYRAAWRRDQGHKVTREAAMAKMTATESAQQVIDSAVQIWGGCGVVSNHPVELLYRE IRALRIYEGATEVQQLIIARQTLTAYEDS--- >WP_050418027.1 acyl-CoA dehydrogenase [Azoarcus sp. CIB] MSDRSYLEWPFFEERHRGMQVELEAWAAAHIDGHPHGDLDDACRELVRKLGADGWLRYMV GGTAYGGRHDTIDTRAVCLLRETLARHSGLADFALGMQGLGSGAITLHGTDAQKRKYLSE VAAGRAIPAFALSEPDSGSDVAAMACSARRDGNDYVLDGEKTWISNGGIADFYVVFARTG EAPGARGLSAFIVEANLPGFEIAERIDVIAPHPLARLRFTGCRVPAANLLGAPGQGFKVA MQTLDIFRTSVAAAALGFARRALDEGLRRATTRDMFGKKLADFQITQAKLAQMATHVDTA ALLTYRAAWMRDQGKNITGAAAMAKMTSTETAQQVIDAAVQLWGGCGVVSEHPVERLYRE IRALRIYEGATEVQQLIIARQTLSAWEQEQAV >WP_011236231.1 acyl-CoA dehydrogenase [Aromatoleum aromaticum] MSDQTYLEWPFFDEPHRQLQIELEAWASANVTEHHGSDLDTACRELVAKFGAAGWLRYVV GGTAYGGCHDVIDTRAVCLLRETLGRHSGLADFAFGMQGLGSGAITLHGTDAQKRDYLPR VASGRAIAAFALSEPGSGSDVAAMACSARQDGDEYVIDGEKTWISNGGIADFYVVFARTG EAAGSRGLSAFIVDADRPGLEIAERIDVIAPHPLARLRFRECRVPKSCLLGVPGQGFKVA MQTLDIFRTSVAAAALGFARRALDEALKRATTRDMFGQKLADFQITQAKLAQMATAVDTS ALLTYRAAWLRDQGQTITGAAAMAKMTSTETAQQVIDAAVQMWGGCGVVSDHPVERLYRE IRSLRIYEGATEVQQLIIARQTLSAYERQQEH
3-Methylbenzoyl-CoA Pathway	
Enoyl-CoA Hydratase	CCH23021.1
Hydroxyacyl-CoA Dehydrogenases	CCH23023.1
Oxoacyl-CoA hydrolase	CCH23022.1
4-Methylbenzoyl-CoA Pathway	
Enoyl-CoA Hydratase	AIW63094.1
Hydroxyacyl-CoA Dehydrogenases	AIW63095.1

Oxoacyl-CoA hydrolase	AIW63096.1
Resorcinol Pathway	
3,5-dihydroxybenzoate hydroxylase large subunit (DbhL)	AIO06084.1
3,5-dihydroxybenzoate hydroxylase small subunit (DbhS)	AIO06085.1
Resorcinol hydroxylase large subunit (RehL)	ABK58620.1
Resorcinol hydroxylase large subunit (RehS)	ABK58619.1
Hydroxyhydroquinone Pathway	
Benzoquinone Dehydrogenase BqdL	>AIO06095.1 benzoquinone dehydrogenase alpha subunit [Thauera aromatica] MPKTIDLHYHAPWQEVVATADDWDHLGSATVLRMLHHLHLVRAFEETVLELDGEGLVHGP AHSSIGQDGGAVGAVSLLRSSDLITGSHRGHHQFLAKCLAHLDRGEADPRRTPLSEGVRT MLYRALAEILGLADGYCRGRGGSMHLRWAEAGALGTNAIVGGGVPLATGAAWACKRRGAG DVAFTFLGDGAVNIGAVPESMNLAALWSLPVCFFIENNGYAVSTKLSEETRETRLSSRGG AYGIPALRVDGMDPVAVRVATQMALDAMRAGQGPYIIEAEVYRYFHHGGGLPGSAFGYRS KDEEAAWRARDPLACLARGMIERDWLSADEDATLRAGARACMVEIAARLTEKDGSKRRIV PALWPQATFRDEGVRGDLAELAGVRCEELETASGKVGEVKFISAVAGVMARRMESDERIF CLGEDIHKLNGGTNGATRGLAARFPDRIVPTPIAEQGFVGLAGGVAMEGHYRPVVELMYA DFALVAADPLFNQIGKARHMFGGDMAVPLVLRSKCAIGTGYGSQHSMDPAGLYAMWPGWR IVAPSTPFDYVGLMNSALQCDDPVLVIEHVGLYNTTAPGPLEDFDYYIPLGKAKVVRPGT ALTVLTYLAMTPLAVKVADELGVDAEVIDLRSLDRAGIDWETIGDSVRKTNNVVVLEQGS QTASYGAMLADEVQRRLFDHLDQPVKRIHGGEAAPNVSKVLERAAFVGAEEVRAGFIEVL ADAGRPLAQTAPALG----- >ABK58621.1 dehydrogenase [Azoarcus anaerobius]

	MPRITNLDYAEPWIELASTPQDWKKLGKTELLRVLYYHHLVRAFEEAVLNLEKLGLVHGP AHSSIGQEGGAVGSVMLLNSSDMITGAHRGHHQFLVKGMQHIDSPSYDPRAAPLPEEVQT FLYRTLAEILGLSDGFCKGRGGSMHLRWVEAGAMGTNAIVGGGVPIANGLAWAQKRRNKG EVTFTFFGDGGMNIGAVPESMNLAALWNLPICFFIENNGYAVSTTLEEETRETRLSSRGG AYAIPAWRVDGMDPVAVRLASEAAIERMRAGKGPTIIEAVLYRYFHHGGSVAGSAFGYRK KDEESSWIAKDPLDRTVREMINLQWLTADENTAIRRHCESAMQGIVERLVEGEGSKRRIR AELWPKPEFRDQGLRGDLSEFKDARFEELETASGPVGDVKFVDAVARVMGRRMETDERVF CMGEDIHRLKGGTNGATKGLAERFPDRIIPAPIAEQGFVGLAGGVAQDGQYRPVVELMYS DFALVAADQLFNQIGKARHMFGGDSAVPLVLRTKCAIGTGYGSQHSMDPAGMYAMWPGWR IVAPSTPFDYVGLMNSALKCEDPVLVIEHTDLYNTTDQGPLEDLDYCIELGKAKVVRKGS AFTVLTYLAMTPLALKVADEMGLDVEIIDLRSLDRAGIDWATIGESIRKTNNVVVLEQGP LTVSYGAMLTDEIQRRFFDYLDQPVQRIHGGESSPSVSKVLERAAFVGAEEIRAGFTRMM ADMGQPLPATPSPAGNSITA
Benzoquinone Dehydrogenase BqdS	>AIO06106.1 benzoquinone dehydrogenase small subunit [Thauera aromatica] MPVEILMPSTGASMSEGNILRWLKQEGEAVERGEALLEIETDKAVVEAVTPARGILGKIL AAGGSEGVKVDSVVGLIAVDGEDPVALAGAVLAGATPAGSAPAGAATVATA--------- AGEASPAEVQRRIPASPLARRLARETGVDLAAVRGRGPHGRVLRADVESVARQAAAAAAP GGAAPLLAATVAAAGTAVPSAAGAAFEDIPHSAMRRVIAQRLGEAKRTVPHFYLSLDCAV DALLALRAQINAQLDAQVGAQVGAQVGAHPDGGKLSVNDFIVKAVALALRRVPGCNAAWT EAAVRRFAEVDIAVAVATPGGLITPIVRHADDKSLGSLSAEIRALAGRAREGRLKPEEYQ GGGFTLSNLGMYGIREFAAIINPPQACILAVGACEQRPVVRDGSLAVATLMSCTLSVDHR VVDGAQAAEFLAEFRRLIENPLAILV >ABK58622.1 dihydrolipoamide acetyltransferase [Azoarcus anaerobius] ------MPSVSTSMTEGTLARWLKKDGETVAKGEVIAEIETDKAILEVEAEAEGIFKAFV ADGAT--VKVGEPMGALLAPGETLGGTISAAQSAAAPTAAAVGGETAVAVAVAAPAAAPS TGHAPAAHDGTRIFASPLARSLALLHGLDLVNISGSGPQGRIVKRDIEA-AMSAQRPASG AVAAPVAEAPVKAPQPAAPQAAGAGYELIPHSSMRRVIAQRLSESKQQVPHFYLTVDCRL DKLLALRQQVN----------GSLPD-VKVSVNDFIVKAVAAAMKRVPATNASWS DEGVRRYRDIDISVAVATPNGLITPVVRQADAKSVGTISAEVKDLAERARQGKLKPDEYQ GGGFTISNLGMYGVRDFAAIINPPQACILAVGTAEKRPVIEDGAIVPATVMTCTLSVDHR VVDGAVGAEFLAAFKALLETPLGLLV
Benzoquinone Dehydrogenase BqdM	>ABK58623.1 putative dehydrogenase E3 component [Azoarcus anaerobius] -MAQEKFDLTVIGGGPGGYVAAIRAAQLGLRTALIEREHLGGICLNWGCIPTKALLRSAE IFDHFKHAGDFGLEVQGASFDLQKIVARSRGVAAQLNAGVKHLLKKNKVQVFEGSGRLAG SGTIRLEQKDG-VSEIQSTHIILATGARARAMAPVEPDGRLVWSYKEAMTPERMPKSLLI VGSGAIGIEFASFYRSLGAEVTVVEVRDRVLPVEDAEVSAFAHKAFERQGMKLLTSSSVV SLQKQADSVIAVIDTKGTTTEIRADRVIAAVGIVGNVENLGLEGTGVQVENTHIVTDAWC

	QTGEPGVYAIGDVAGAPWLAHKASHEGILCVERIAGVDGIHPLDKTRIPGCTYSRPQIAS IGLTEAQAKERGYELKVGRFPFMGNGKAIALGEPEGFIKTVFDAKTGELLGAHMVGAEVT ELIQGFSIGKTLETTEAELMHTVFPHPTLSEMLHEATLAAYGRAIHT >AIO06092.1 dihydrolipoamide dehydrogenase family protein [Thauera aromatica] MTDNNSYDLIVVGAGPGGYVAAIRAAQLGMKTAVVEREHLGGICLNWGCIPTKALLRSAE VGRLARHAAEYGVSVPEPKFDLERIVQRSRAIAAQLNGGIRHLLNKNKVSVIEGEARLAG AGRVAVTRGGADAGTYAAPHLILATGARARQLPGLEDDGRLVWTYRKAMTPDVLPKSLLI VGSGAIGIEFASFYHALGSQVTVVEVMDRILPVEDEDISALARKAFEDQGMRILTGAKAS IARKSAECVTVRIEAGGAAEELTVDRVIVAVGISPNTENLGLEHTRVRLERGHIVTDPWC RTDEPGLYAIGDVTRPPWLAHKASHEAMICVEAIAGLADVHPLELRNIPGCTYSHPQIAS VGLTERKAREQGHEVRVGRFPFVGNGKAIALGEPEGLVKTVFDARSGELLGAHMIGAEVT ELIQGYTLARTLEATEAELIATVFPHPTLSETMHEAVLAAYGRAIHI
HHQ dehydrogenase large subunit (BtdhL)	ABK58630.1
HHQ dehydrogenase small subunit (BtdhS)	ABK58631.1
Phloroglucinol Pathway	
Phloroglucinol Reductase	>WP_014184752.1 SDR family oxidoreductase [Desulfosporosinus orientis] MVDIQ--FVNNLFDVKDKVALITGATGALGKAISFGYGLAGMKIFVTGRSGEKCKALCDE LEAQGIECGYSIGDPAVEADVIKVVEDAVQKFGEINVLLTAAGYNHPQPIVDQDLAEWKK IMDSDVQGTWLFCKYAGQQMIERGKGGKVILVSSARSKMGMAGYTGYCTAKAGIDLMAQS LACEWTAKYKINVNTINPTVFRSDLTEWMFDPESPVYANFLKRLPVGRLGEPEDFIGPCI FLASNASDFMTGANVATEGGYWAN >WP_021630531.1 SDR family oxidoreductase [Clostridium sp. ATCC BAA-442] MVNVKKEFVDNMFSVKGKVALVTGATGALGCVLSKAYGYAGAKVFMTGRNEKKLQALEAE FKAEGIDCAYGVADPADEAQVDAMITACVAQYGEVNILAVTHGFNKPQNILEQSVADWQY IMDADCKSVYVVCKYVAQQMVDQGKGGKIVVVTSQRSKRGMAGYTGYCTSKGGADLMVSS MACDLSAKYGINVNSICPTVFRSDLTEWMFDPESAVYQNFLKREPIGRLAEPEDFVGYAL FLSSDASNYITGANCDCSGGYLTC >WP_027868985.1 SDR family oxidoreductase [Eubacterium sp. AB3007] MVNVEKSFVNNMFSVEGKVALVTGATGALGCVLSKAYGYAGAKVFMTGRNAEKLQKLQDE FEAEGIDCAYFVADPQKEEDVKALIAACVEKYGEVNILAICHGYNKPANILDQSVEDWQF IMDADCKSVYIVCKYVAEQMVEQGKGGKMVVVTSQRSKRGMAGYTGYCTSKGGADLMVSS MACDLTAKYGINVNSICPTVFRSELTEWMFDPDSEVYKNFLKREPIGRLAEPYDFVGFAL FLSSEASDFMTGGNYDCSGGYLTC

Table 3.4. Enzymes in 159R homologous to *Sphingomonas paucimobilis* SYK-6 involved in lignin degradation or metabolism

Enzyme in SYK-6	Bit Score	E-value	% Identity	159R Gene annotation	Gene ID
Beta-etherase (ligF)	45.8	7e-07	40	glutathione S-transferase	2788607536
2-keto-4-carboxy-3-hexenedioate hydratase (ligJ)	437	1e-155	60	4-oxalomesaconate hydratase	2788602671
4-carboxy-4-hydroxy-2-oxoadipate aldolase (ligK)	231	9e-78	59	4-carboxy-4-hydroxy-2-oxoadipate aldolase	2788602672
LigR protein (ligR)	273	3e-89	40	transcriptional regulator /LysR family transcriptional regulator	2788606035
Vanillin dehydrogenase (ligV)	325	3e-107	41	aldehyde dehydrogenase (NAD+)	2788604477

Table 3.5. Enzymes in 159R involved in lignin degradation or metabolism. Enzyme annotation was predicted by the Department of Energy- Joint Genome Institute (DOE-JGI) Microbial Genome Annotation Pipeline (MGAP v.4) (129).

Gene Product	IMG JGI Gene ID
4,5-DOPA dioxygenase extradiol	2788604550
benzoate membrane transport protein	2788605192
feruloyl esterase	2788602630
4-hydroxybenzoate polyprenyltransferase	2788605631
AAHS family 4-hydroxybenzoate transporter-like MFS transporter	2788601817

p-hydroxybenzoic acid efflux pump subunit AaeAB	2788604686, 2788604687
vanillate O-demethylase ferredoxin subunit	2788604204, 2788606120
vanillate O-demethylase monooxygenase subunit	2788606119
2-succinylbenzoyl-CoA synthetase	2788607257
O-succinylbenzoate synthase	2788607258
glutathione S-transferase	788605280, 2788602381, 2788606111, 2788603289, 2788603705, 2788605754, 2788605671, 2788603656, 2788603688, 2788605210, 2788603033, 2788607494, 2788604573, 2788604878, 2788606764, 2788604627
xylulokinase	2788605934, 2788605929, 2788605827, 2788606718, 2788603328, 2788604167, 2788604787, 2788606042
alpha-D-xyloside xylohydrolase	2788605868, 2788606010
GH43 family beta-xylosidase	2788606510
xylose isomerase, xylose isomerase-like TIM barrel protein	2788602980, 2788602989, 2788603070
2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase in catechol pathway	2788602115, 2788604948, 2788602827, 2788605041
4-carboxymuconolactone decarboxylase	2788607467, 2788605411
phenylpropionate dioxygenase-like ring-hydroxylating dioxygenase large terminal subunit	2788605051, 2788604207
catechol 2,3-dioxygenase, catechol 2,3-dioxygenase-like lactoylglutathione lyase family enzyme	2788605046, 2788606693, 2788603188, 2788605200, 2788605050, 2788606195

3.7 Genetic Potential for Host-symbiont Interactions

Sodalis species are predominantly insect endosymbionts that range from recent to ancient origin (98). By comparing genomes of different stages, it is possible to understand how host-symbiont associations evolve over time (99). Phylogenetic and genomic evidence presented in this study suggest that strain 159R is an antecedent to both the symbiont *Sodalis*-clade as well as free-living *S. praecaptivus.* To look further into the genetic potential strain 159R has to associate with hosts, we compared genes of *Sodalis* endosymbionts with strain 159R that were related to host-symbiont interactions. Previously, genes have been identified in *S. glossinidius morisitans* for outer membrane proteins, OmpA and OmpR, as well as a type III secretion system (T3SS) that are critical for infection (97,99,130). Strain 159R also contained these genes as well as genes encoding for 3 chitinases, a chitin deacetylase, and a collagenase-like PrtC family protease, suggesting that it has the capability to live as an insect endosymbiont. Further work should be completed to determine what other genes are necessary for insect colonization. Since strain 159R also has pathways for catabolism of plant metabolites, it would also be of interest to investigate any plant host associations as well.

3.8 Description of *Sodalis* sp. strain 159R

Cells are facultative, Gram negative, rod-shaped cells, and can grow up to 37°C. On R2A media, colonies are non-pigmented, opaque circular colonies with shiny surfaces. According to Biolog GN2, cells can assimilate α-D-glucose-1 phosphate, α-D-glucose, α-D-lactose, D-glucose-6-phosphate, D-fructose, D-

galactonic acid lactone, D-galactose, D-gluconic acid, D-glucuronic acid, D-mannitol, D-mannose, D-serine, D-sorbitol, D-trehalose, D,L-α-glycerol phosphate, D,L-lactic Acid, glycerol, L-aspartic acid, maltose, N-acetyl-D-galactosamine, N-acetyl-D-glucosamine, pyruvic acid methyl ester, succinic acid, and mono-methyl succinate.

The type strain is 159RT (=DSM tbd, =ATCC tbd), which was enriched from temperate soil in Petersham, MA, onto organosolv lignin under anoxic conditions. The genome of the type strain is characterized by the size of 6.38Mbp and a G+C content of 54.9 mol%.

Data availability. This Whole Genome Shotgun project has been deposited in GenBank under the accession no. SJOI00000000.

CHAPTER 4

IRON CHELATOR-MEDIATED ANOXIC BIOTRANSFORMATION OF LIGNIN BY NOVEL SP., *TOLUMONAS LIGNOLYTICA* BRL6-1

4.1 Abstract

Lignin is a recalcitrant biopolymer that can comprise up to 30% of plant biomass. Current pretreatment methods to remove lignin are environmentally unfriendly and costly for industrial applications such as biofuel production or paper mill pulping. An alternative and greener approach is biopulping, which uses microbes and their enzymes to break down lignin and has potential to add value to lignin from lignocellulose. Here we investigate the physiology and lignin biotransformation mechanisms of a novel isolate, *Tolumonas lignolytica* BRL6-1, under anoxic conditions. *Tolumonas lignolytica* BRL6-1 is a facultative anaerobic bacterium that was isolated from tropical forest soils on lignin as a sole carbon source. To determine the role of lignin in BRL6-1 metabolism, we compared physiological and biochemical changes when the cells were grown anaerobically in either lignin amended or un-amended conditions. In the presence of lignin, BRL6-1 had a higher biomass and shorter lag phase compared to un-amended conditions, with 14% of the upregulated proteins by $\log_2$ fold-change of 2 or greater relating to Fe^{2+} transport in early exponential phase. Ferrozine assays of the <10kDa supernatant fractions confirmed that Fe(III) was bound to lignin and reduced to Fe(II) only in the presence of BRL6-1, suggesting redox activity by the cells. In

addition, electron paramagnetic resonance (EPR) detected radical molecules >10kDa supernatant fractions in lignin amended conditions.

From our findings, we hypothesized that BRL6-1 is producing a small molecule or protein that acts as both an iron chelator and redox agent under anoxic conditions to obtain the iron bound to lignin. Arnow assays identified catechol-like siderophores in <10kDa supernatant fractions only in lignin amended conditions. However, concentrations of these compounds did not change over the course of BRL6-1 growth and were similar to abiotic controls, suggesting that the compounds were lignin derived and unrelated to BRL6-1 metabolism. Alternatively, BRL6-1 may be utilizing an anaerobic radical enzyme that is interacting with the lignin and iron. Secretome (extracellular enzyme) analysis showed an extra band at 20kDa in lignin amended conditions. LC-MS/MS analysis identified the presence of a protein of unknown function but had homology to enzymes in the radical SAM superfamily, suggesting that it may have a role in radical formation in lignin amended conditions. Finally, we tested to see if low molecular weight (LMW) lignin fractions were being produced from BRL6-1 interacting with lignin. Fourier transform ion cyclotron resonance (FTICR) mass spectrometer analysis of <10kDa supernatant fractions did not detect LMW lignin derivatives in the presence of BRL6-1. These results suggest that if lignin biotransformation is occurring, it is within the larger polymer structure and should be further studied to determine what linkages and subunits are being targeted.

Overall this investigation suggests that BRL6-1 is using a protein similar to the radical SAM superfamily to interact with the Fe(III) bound to lignin and reducing

it to Fe(II) for cellular use, increasing BRL6-1 fitness under lignin amended conditions. This interaction potentially generates organic free radicals and causes a radical cascade which could modify and depolymerize lignin. This mechanism would be similar to previously described aerobic chelator-mediated Fenton chemistry or radical producing lignolytic enzymes, such as lignin peroxidases, but under anoxic conditions.

4.2 Introduction

The industrial processing of lignocellulosic material produces $5x10^6$ metric tons of lignin annually worldwide (7). Lignin is the largest renewable source of aromatics that can be used for products such as flavors, fragrances, dyes, and other valuable secondary metabolites (18,19). However, it is considered an 'untapped" resource due to the difficulty of removing lignin from lignocellulosic material and conversion to desired downstream products (35,131).

Investigation of microbial mediated processes for the depolymerization of lignin have focused predominantly on aerobic fungi and bacteria (20,21,35,127,132). Under oxic conditions, enzymes such as laccases and peroxidases produce oxidants that diffuse into and reduce the lignin complex (132,133), causing bond scission reactions between lignin subunits. For organisms that lack lignolytic enzymes such as brown-rot fungi as well as bacteria like *Pantoea ananatis* Sd-1 and *Cupriavidus basilensis* B-8, chelator-mediated Fenton chemistry (CMF) is used to depolymerize lignin (25,26,134,135). In this mechanism, the microorganism produces an iron reducer molecule, a chelator molecule, and H_2O_2.

Once the chelator binds to Fe(III) in the environment, it then reacts with the iron reducer molecule to reduce Fe(III) to Fe(II). Fe(II) then reacts with H_2O_2 to create •OH radicals. Similarly to oxidants formed by laccases and peroxidases, the •OH radicals disrupt the lignin structure, causing bond scissions of subunits (135). Both lignolytic enzyme and chelator mediated lignin depolymerization are promising for industries that rely lignocellulosic feedstocks (127). For example, the use of Fenton chemistry and aerobic bacterium, *C. basilensis* B-8, for lignocellulosic processing has been studied on rice straw, showing a synergistic relationship in lignin depolymerization and cellulose yield (134). However, there are limitations to these processes that hinder them to be competitive on the market. Both aerobic fungi and bacteria require constant aeration and mixing, making it very costly to maintain the cultures (29). Mass production of fungal or bacterial lignolytic enzymes are also not possible due to lacking a method of recycling the enzymes after one use, low substrate specificity, and low redox potential (23).

Though originally thought to be impossible 30 years ago (136), anaerobic bacteria could offer a solution to issues presented for aerobic microorganisms. Bacterial anaerobic extracellular lignin depolymerization has been previously studied in *Klebsiella* sp. strain BRL6-2 and *Enterobacter lignolytica* SCF-1 (30,31,137). Based on genome analysis, BRL6-2 is hypothesized to use lignin as an electron acceptor for energy production (30). Support for this mechanism is also seen with humic substances, which are lignin rich (138), acting as electron acceptors for bacteria in sediments and anoxic waters (139). RNAseq analysis comparing SCF1 growth in lignin amended and un-amended conditions suggested

various enzymes that may be responsible for lignin depolymerization, including alcohol dehydrogenases (31). However, the exact mechanism has yet to be elucidated. By identifying additional anaerobic bacteria capable of degrading lignin, mechanisms and their regulation can be uncovered and further developed for lignin depolymerization and valorization applications.

Tolumonas lignolytica BRL6-1 is a novel, facultative anaerobic bacterium that was isolated from tropical forest soils on lignin as sole carbon source (140). In the presence of lignin, BRL6-1 has a shorter lag phase and a higher biomass (140). However, the mechanism of lignin modification and how it benefits cell growth is not well understood. We hypothesize that when grown anaerobically in the presence of lignin, BRL6-1 produces an extracellular protein that acts as both iron chelator and redox agent. This protein potentially generates organic free radicals and causes a radical cascade that modifies and depolymerizes lignin. Our work aims to elucidate the molecular mechanism of anaerobic lignin modification by *Tolumonas lignolytica* BRL6-1 as a way to valorize lignin for industries relying on lignocellulose as their raw material.

4.3 Materials and Methods

4.3.1 Culturing Tolumonas lignolytica BRL6-1

To study lignin modification under anoxic conditions, *Tolumonas lignolytica* BRL6-1 was grown in 0.04% D-glucose as the primary carbon source amended or un-amended with 0.1% alkali lignin, low sulfonate (Sigma Aldrich, CAS Number 8068-05-1). Cultures grew on modified CCMA media consisting of (per liter) 2.25 g

NaCl, 0.5 g NH_4Cl, 0.227 g KH_2PO_4, 0.348 g K_2HPO_4, 5 mg $MgSO_4 \cdot 7H_2O$, 2.5 mg $CaCl_2 \cdot 2H_2O$, 0.01 mL SL-10 trace elements, and 0.01 mL Thauer's vitamins (141–143) . The D-glucose concentration was 0.2% for the ferrozine assays, Arnow assays, proteome and secretome analysis, described in more detail below. Cultures grew at 30°C anaerobically in triplicate and uninoculated bottles serve as abiotic controls. Iron amended cultures had an additional 38 ppb Fe(II) added to the media as $FeCl_2 \cdot 4H_2O$.

To study the physiological response of BRL6-1 in the presence of lignin, growth in lignin-amended medium, lignin unamended medium, and lignin unamended medium supplemented with an additional 38 ppb Fe were monitored by measuring cell concentration by adsorption (OD_{600}). Bacterial growth curves were analyzed with gcFit function via *grofit* package in R (144). Calculated average lag phase, maximum growth rate (μ Max), and maximum cell growth (A) were based on the Gompertz Model.

4.3.2 Proteomic Analysis for Cell Pellet and Secretome

To identify proteins differentially expressed during lignin-amended growth, biomass was collected at early and late logarithmic growth phase from cultures grown in the presence or absence of lignin and sent to Pacific Northwest National Laboratory (PNNL) for LC-MS/MS preparation and analysis. Proteins were extracted from cell pellet fractions using methanol/chloroform and analyzed with LC-MS/MS (Joshua Adkins, PNNL personal communication). Raw mass spectrometry data were searched with MS-GF+ against NCBI RefSeq *Tolumonas* sp. BRL6-1 database

(October 2014 version) in addition with bovine/porcine trypsin and other common contaminants such as keratin sequences (3164 total sequences). Searching parameters required tryptic digestion of at least one of the peptide ends (partially tryptic), <10 ppm peptide mass tolerance and methionine oxidation as variable modification. The identified MS/MS spectra were filtered with an MS-GF+ score of 1e-09 resulting in ≤ 1.0% false discovery rate (FDR) at the protein level. The count of spectra attributed to each individual protein within each experimental condition is as a value for quantitative analysis.

Supernatant fractions from late logarithmic growth phase were collected to identify differentially expressed proteins in the secretome during lignin-amended growth. Cell-free supernatant was generated by collecting 100 mL cultures at the end of late logarithmic growth phase and vacuum filtrated with a 0.45 μm filter. Cell-free supernatant was centrifuged using a 10 kDa filter & the >10 kDa fraction was further concentrated using TCA/DOC precipitation (41). Samples were run on a 15% SDS-PAGE gel and silver stained. Bands of interest from both lignin amended and unamended samples were cut out at 20 kDa, 37 kDa, and 50 kDa. Using an in-gel tryptic digest kit (Thermo Fisher, Catalog #89871), samples were prepared as described by the manufacturer for LC-MS/MS analysis. LC-MS/MS analysis was completed by the Mass Spectrometry Center at University of Massachusetts Amherst (Stephen Eyles, UMass Amherst personal communication). Raw mass spectrometry data was search with MS/GF+ against the NCBI RefSeq *T. lignolytica* BRL6-1 database (2016 version).

4.3.3 Analysis of Proteomic Data

Spectral counts from the cell pellet proteomics between lignin amended and un-amended conditions were compared using msms.edgeR function via *msms.Tests* package in R (145). The post-test effect size filter of msms.edgeR deemed proteins differentially expressed if proteins had p-values <0.05, absolute values of $\log_2$ fold-change >1 or < -1, and total spectral counts >2 across biological replicates (145).

4.3.4 Ferrozine and Arnow Assays

In order to confirm that iron was bound to lignin, Fe^{2+} and Fe^{3+} concentrations were measured with ferrozine assays. Supernatant from lignin amended and unamended cultures were harvested by removing in 15 mL aliquots of culture from serum bottles under anoxic conditions during lag phase, late logarithmic growth phase, and mid-stationary growth phase. Samples were filtered under anoxic conditions through a 0.45 μm filter to remove biomass and then ultrafiltrated with a 10 kDa Amicon filter. Filtrate (<10 kDa) was tested in triplicate for total iron using a ferrozine assay as previously described (146). Briefly, in a 96-well plate under anoxic conditions, 225 μL of sample with 15 μL 1 M ascorbic acid were added, followed by 60 μL of 50 mg ferrozine/mL and 500 mM potassium acetate buffer, pH 5.5. Plates were wrapped in tin foil and incubated for 135 min at 37ºC before being read at 562 nm with a plate spectrophotometer. A separate set of plates had ascorbic acid substituted with water to calculate Fe^{2+} in the media. For both ferrozine assays a standard curve of Fe^{2+} was completed as well as controls with samples taken from the abiotic cultures.

To determine if BRL6-1 is producing catechol-like compounds, Arnow assays were completed as described (147) with the same fractions as the ferrozine assays to detect catechol-like siderophores known as catecholates. Briefly, 1 mL of a <10 kDa fraction sample was combined with 1 mL 0.5 M HCl, 1 mL nitrite-molybdate reagent, and 1 mL 1 M NaOH. Reactions was incubated for 5 min before being diluted 5-fold with water in a 96-well plate and light absorbance read at 510 nm. Samples from the biotic replicates as well as abiotic controls were completed in triplicate.

4.3.5 Inductively Coupled Plasma (ICP) Spectroscopy of Kraft Alkali Lignin Substrate

One gram of alkali lignin, low sulfonate (Sigma Aldrich, CAS Number 8068-05-1) was sent in triplicate to the University of Massachusetts Amherst Soil and Plant Nutrient Testing Laboratory. Lignin was acid wet digested in nitric acid, hydrochloric acid, and hydrogen peroxide in a block digester and measured with ICP Spectroscopy to determine the total P, K, Ca, Mg, Zn, Cu, Mn, Fe, and B.

4.3.6 Nuclear Magnetic Resonance (NMR) Analysis

Supernatants from lignin amended and unamended cultures were harvested anaerobically at mid-stationary growth phase by aseptically removing 15 mL aliquots from serum bottle cultures under anoxic conditions. Samples were filtered under anoxic conditions through a 0.45 μm filter to remove biomass and then ultrafiltered with a 10 kDa Amicon filter at 4,000×*g*. To avoid filter contaminants, such as glycerol, from interfering with NMR analysis, filters were pre-washed 5×

with N_2 purged sterile water. The < 10 kDa supernatant fractions were sent to PNNL for proton nuclear magnetic resonance (NMR) metabolite analysis. At PNNL, samples were diluted by 10% (v/v) with 5 mM 2,2-dimethyl-2-silapentane-5-sulfonate-d6 (DSS) as an internal standard. All NMR spectra were collected using a Varian Direct Drive 600 MHz NMR spectrometer equipped with a 5 mm triple-resonance salt-tolerant cold probe. The 1D 1H NMR spectra of all samples were processed, assigned, and analyzed by using Chenomx NMR Suite 8.3 with quantification based on spectral intensities relative to the internal standard. Candidate metabolites present in each of the complex mixture were determined by matching the chemical shift, J-coupling, and intensity information of experimental NMR signals against the NMR signals of standard metabolites in the Chenomx library. The 1D 1H spectra were collected following standard Chenomx data collection guidelines (148), employing a 1D NOESY presaturation experiment with 65536 complex points and at least 512 scans at 298 K.

4.3.7 Fourier Transform Ion Cyclotron Resonance Mass Spectrometer (FTICR-MS) Analysis

The same sample fractions (<10 kDa) for NMR analysis described above also were sent to PNNL for FTICR-MS analysis. Samples were directly infused into a 15 Tesla Bruker Solarix XR FTICR mass spectrometer using negative mode electrospray ionization. The flow rate was 3 ul/min, with an ion accumulation time of 50 ms and time of flight of 750 ms. The free induction decay was recorded into a 4MWord time domain of 1.4 sec yielding a resolving power, after magnitude mode Fourier

transform, of 470,000 at m/z 400. Spectra were recorded between m/z 153 and 1000. Processing parameters, including zero filling and apodisation settings, were instrument defaults.

Spectra were internally calibrated against homologous series of CHO and CHOS species confirmed by isotopic fine structure analysis using Bruker DataAnalysis 5.0, with peaklists exported to Formularity for exact mass to formula assignment. Formularity used a peak alignment tolerance of 1 ppm, mass accuracy tolerance of 0.25 ppm, and elemental constraints of O>0, N<=2, S<=3, P<2. Only singly charged, deprotonated species were assigned. Isotopologues of multiply charged ions were removed through filtering of mass defects greater than 0.3.

4.3.8 Electron Paramagnetic Resonance (EPR) Analysis

The >10 kDa fractions from the same set of samples for NMR analysis described above were sent to PNNL for EPR analysis. Only one biological replicate and one abiotic control from lignin amended conditions were analyzed. Spectra were acquired on a Bruker Elexsys 580 spectrometer equipped with a SHQE resonator and a Bruker continuous flow liquid nitrogen cryostat (VT 4131). Spectra at temperatures between liquid samples for frozen solution experiments were loaded in 4 mm OD × 3 mm ID FEP tubes (Wilmad) and flash frozen in liquid nitrogen. Microwave frequency was typically ~9.34 GHz and a microwave power of 0.2 mW. The field was swept from 0 to 5000 G in 83 s and modulated at a frequency of 100 kHz with 10 G amplitude. A time constant of 82 ms was employed.

4.4 Results and Discussion

To determine the role of lignin in the anaerobic metabolism and growth of *T. lignolytica* BRL6-1, we compared cultures grown with glucose that were either amended or unamended with lignin. Our growth results support previous findings that under lignin amended conditions, BRL6-1 fitness improves, having a shorter lag phase and higher biomass (140) (**Table 4.1**). To explain this change in growth, it was originally hypothesized that lignin may serve as a secondary carbon source as well as a potential energy source (140). Comparing protein expression between lignin amended conditions to unamended, our msms.edgeR analysis (145) resulted in a total of 41 proteins were significantly up-regulated and 101 down-regulated in early exponential phase and a total of 9 proteins were significantly up-regulated and 9 proteins down-regulated in late exponential phase (**Fig. 4.1**).

Table 4.1. Changes in average lag phase, maximum growth rate (μ Max), and maximum cell growth (A) of *T. lignolytica* BRL6-1 growth in lignin amended, un-amended conditions, and un-amended conditions with 38ppb iron addition. t-Test assuming unequal variances was used to compare conditions; p-value = 0.03 and 0.04 for lag phase and A, respectively.

Condition	Lag Phase (hrs)	μ Max (OD_{600} per hr)	A (OD_{600})
Lignin Amended	5.0^{b} (±0.6)	0.039^{a} (±0.03)	0.140^{b} (±0.013)
Lignin Un-Amended	11.0^{a} (±4)	0.030^{a} (±0.008)	0.124^{a}(±0.003)
Lignin Un-Amended + 38ppb Fe	10.5^{a} (±5)	0.072^{a} (±0.03)	0.124^{a}(±0.007)

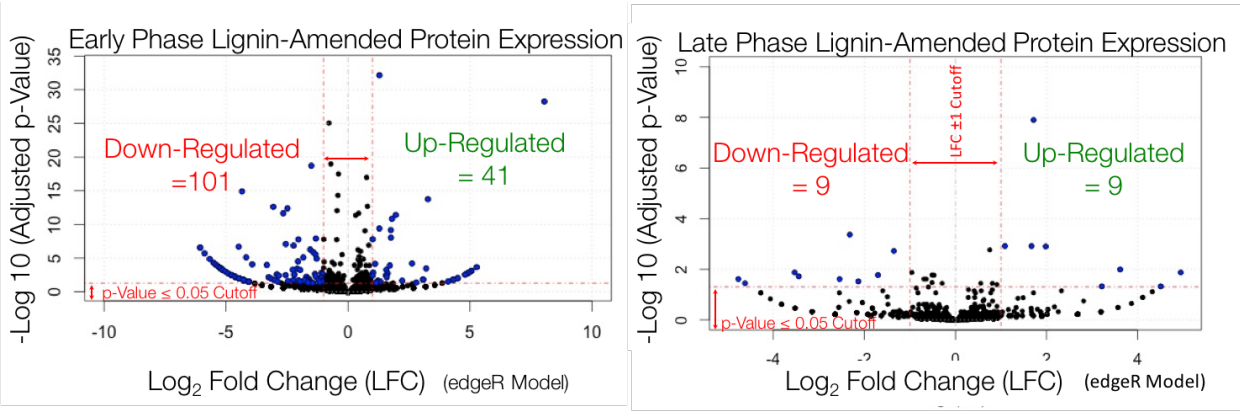


Figure 4.1. Proteomic analysis results of *T. lignolytica* BRL6-1 grown in lignin amended versus unamended conditions. Blue dots represent significantly expressed proteins in lignin amended conditions whereas black dots denote represent proteins that did not change between lignin amended and un-amended.

In both early and late exponential phase, the most significantly up-regulated protein in lignin amended conditions was annotated as a carboxymuconolactone decarboxylase (CMD) family protein with an alkylhydroperoxidase (AhpD) domain and CXXC motif (log_2 fold-change of 8 and 7, respectively) (**Fig 4.2A**). Based on the CXXC motif, it is thought that the protein detected in BRL6-1 has AhpD-like activity (149). AhpD is part of an antioxidant defense system that forms a complex with peroxiredoxin, AhpC. Its function is to restore the enzyme activity of AhpC via reduction (150); however, looking further into the genome, BRL6-1 contains a gene annotated as AhpF, which is an alternative alkyl hydroperoxide reductase to AhpD as seen in *Salmonella typhimurium* (63.5% sequence identity with NCBI BLASTp) (151). Based on this information, AhpC likely forms a complex with AhpF. To support this, BPROM program analysis placed AhpC and AhpF downstream of the same predicted promoter in the BRL6-1 genome (152). Additionally our msms.edgeR analysis determined AhpCF protein expression was not significantly different between lignin amended and un-amended conditions. Therefore, this up-

regulated AhpD-like protein in the presence of lignin could be serving another role with its reducing activity instead of restoring AhpC activity.

The mechanism that AhpD uses to reduce its substrates is a proton relay system (145). This mechanism has been described previously in lignin degrading enzyme, LigL, in *Sphingomonas paucimobilis* SYK-6. LigL is part of the first degradation step of lignin-derivative, (αS, βR)-GGE, via stereospecific oxidation of the benzylic alcohol (153,154). The proton relay mechanism has also been described for p-Cresol methylhydroxylase (PCMH) in *Pseudomonas* species to degrade phenol p-cresol as well as p-hydroxybenzyl alcohol (155). Therefore, this AhpD-like protein could be reducing lignin-derived compounds in the cell using this mechanism. Support for cellular aromatic compound uptake for the AhpD-like enzyme to act upon includes the up-regulation of a C4-dicarboxylate ABC transporter protein in both early and late exponential phase (log_2 fold-change of 1.2 and 4.9, respectively). This transporter has an 87% sequence identity to transporter, DctA, in *Pseudomonas chlororaphis* O6 that was found to be essential for benzoate uptake (156).

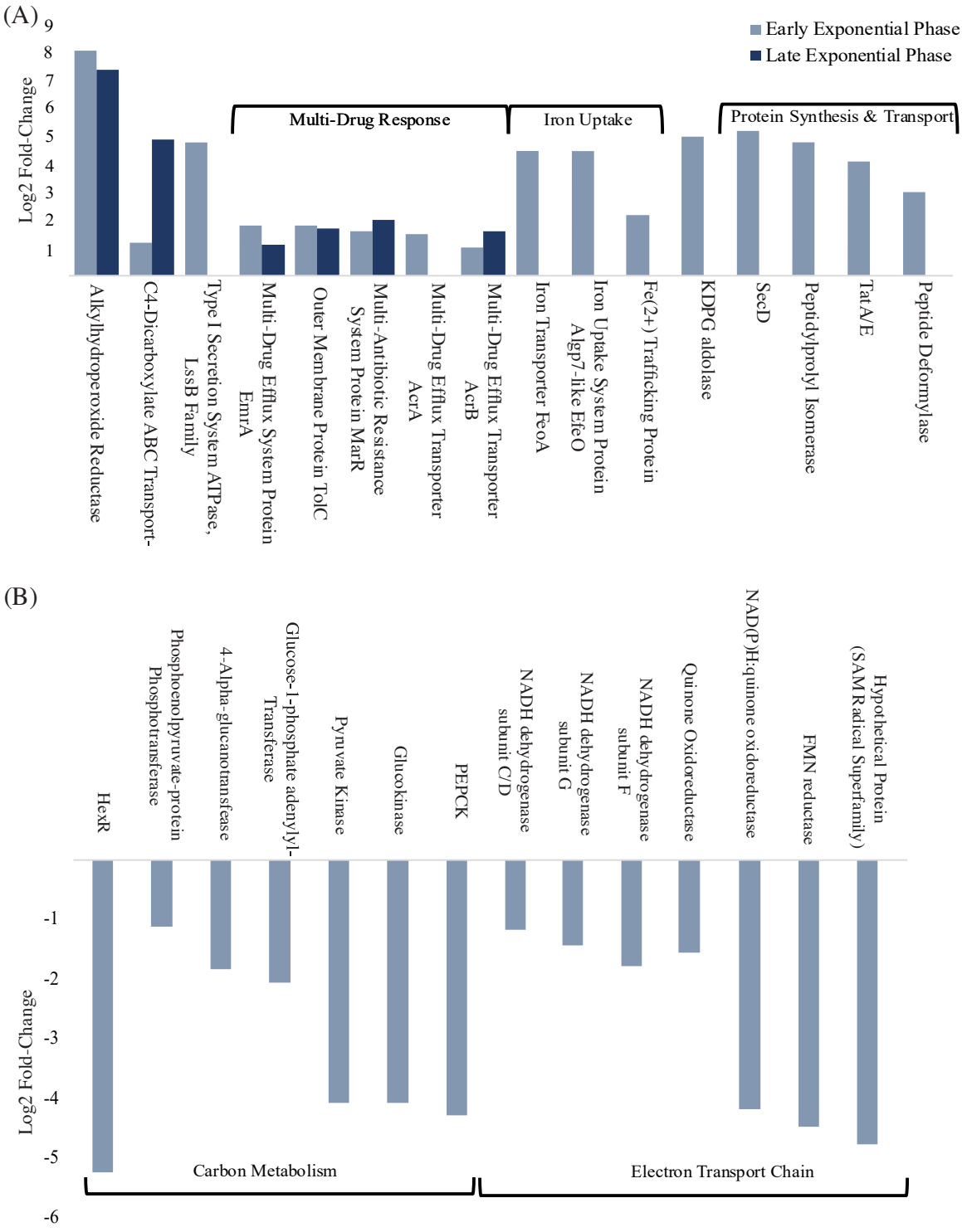


Figure 4.2. (**A**) Significantly up-regulated and (**B**) down-regulated protein expression of BRL6-1 under lignin amended conditions compared to lignin un-amended ($P < 0.05$). Abbreviations are the following: phosphoenolpyruvate carboxykinase (PEPCK); 2-keto-3-deoxy-6-phosphogluconate (KDPG).

With potential evidence that BRL6-1 is up-taking benzoate compounds and degrading them via AphD, we then compared changes in protein expression relating to C metabolism in the presence of lignin. In lignin amended conditions, a HexR transcriptional factor was significantly down-regulated in BRL6-1 by a $\log_2$ fold-change of -5 during early logarithmic growth phase (**Fig 4.2B**). HexR is known as a global central carbon metabolism regulator that represses the transcription of glucose-related genes (157,158). It is negatively affected by the Entner-Doudoroff (ED) keto-deoxy-phosphogluconate (KDPG) aldolase intermediate, 2-keto-3-deoxy-6-phosphogluconate (KDPG) (22). BRL6-1's glucokinase and pyruvate kinase, which are responsible for the first and last step of glycolysis, respectively, were also down-regulated by a $\log_2$ fold-change of -4, whereas a KDPG aldolase was significantly up-regulated by a $\log_2$ fold-change of 5 (**Fig 4.2**). These findings suggest that in the presence of lignin, there is a higher conversion of KDPG to pyruvate by the KDGP adolase. With lower concentrations of KDPG present in BRL6-1, HexR is then able to repress glucose from being converted to glucose-6-phosphate as well as repress the conversion of phosphoenolpyruvic acid (PEP) to pyruvate. Furthermore, phosphoenolpyruvate carboxykinase (PEPCK), the rate limiting enzyme for gluconeogenesis, was also down-regulated by a $\log_2$ fold-change of -4, indicating that PEP was not being funneled into gluconeogenesis (23). Metabolomic analysis supported the shift seen in C related proteins, with glucose consumption higher in lignin amended conditions as well as having a higher production of pyruvate, lactate, and formate (**Fig. 4.3**). This shift in C metabolism may be due to BRL6-1

potentially producing more extracellular proteins in the presence of lignin, as there is a significant increase in BRL6-1 protein expression relating to protein synthesis and transport (**Fig 4.2A**). Since protein synthesis can be growth-limiting, the ED pathway is energetically favorable for facultative anaerobic bacteria over glycolysis, requiring less enzymes that need to be produced and therefore can support growth (159).

Despite the observed shifts in C metabolism-related proteins described above, there was no significant up-regulation of enzymes related to aromatic metabolism (140). Additionally, NMR analysis did not detect any monomers present in the <10 kDa supernatant. This suggests that either no monomers are being cleaved from lignin in the presence of BRL6-1, monomer production is below detection, or that lignin was depolymerized into high molecular weight polymers and present only in the >10 kDa fraction which was not analyzed.

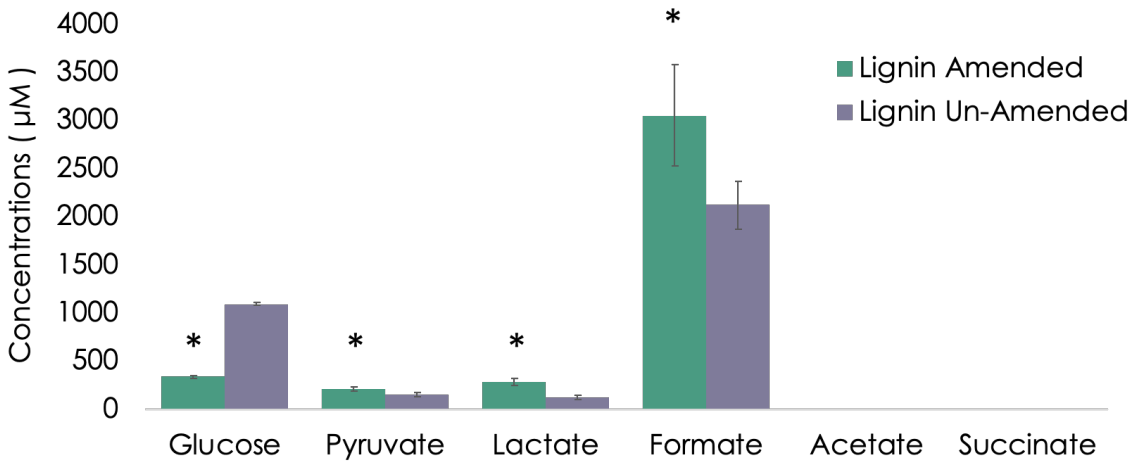


Figure 4.3. Primary carbon source (glucose) and metabolite (pyruvate, lactate, formate, acetate, and succinate) concentrations (μM) at late exponential phase of BRL6-1 grown either in lignin amended or un-amended conditions.

To determine if lignin had a role in energy production, we searched for enzymes that were related to the electron transport chain that were significantly expressed higher or lower in lignin amended conditions compared to un-amended. During early exponential phase, 6% of all significantly downregulated proteins (log_2 fold-change >1, $p < 0.05$) in the presence of lignin were NADH dehydrogenase subunits (log_2 fold-change between -1.2 to -4) and one flavin mononucleotide (log_2 fold-change of -4) (**Fig. 4.2B**). This was surprising since organisms such as *E. lignolytica* SCF-1 had upregulation in NADH dehydrogenase and other electron transport chain enzymes in the presence of lignin (137). Upon further examination of proteins that were significantly up-regulated by a log_2 fold-change >2 in the presence of lignin during early exponential phase, 14% were related to Fe^{2+} uptake (**Figure 4.2A**). We hypothesized that BRL6-1 could be obtaining energy using iron redox and that lignin could play a role due to its strong affinity for iron (160). The complex between lignin and iron makes iron more soluble in the environment but not necessarily more bio-available for cellular use (161). This considered, BRL6-1 may have a mechanism that is disrupting the lignin-iron association. By doing so, BRL6-1 could obtain both iron and a potential carbon source faster than cells in unamended conditions, explaining the ability of BRL6-1 to exit lag phase more quickly in the presence of lignin (162).

To further investigate the relationship between lignin, iron, and BRL6-1 fitness, we first asked if lignin was contributing to a higher iron concentration aside from the SL-10 minerals we added to the media. ICP Spectrophotometry showed that the lignin substrate contained 38 ppb iron (**Table 4.2**). To test whether the 38

ppb Fe was benefitting BRL6-1 fitness, we completed Fe addition growth curve experiments and monitored BRL6-1 growth. There was no significant difference between lignin amended conditions and lignin unamended conditions with iron addition (**Table 4.1**), suggesting that additional iron alone was not enough to benefit BRL6-1 growth.

To test if lignin has a strong affinity for iron in our system, we completed a ferrozine assay for <10 kDa fractions of supernatant from lag phase, late logarithmic growth phase, and mid-stationary growth phase. We expected that if the iron was binding to lignin in the media, we should see less bioavailable iron in the supernatant of lignin amended compared to lignin unamended conditions. In lignin amended conditions, Fe(III) was not detectable in the <10 kDa fractions throughout the entire growth curve whereas 292 ppb Fe(III) was detected during lag phase in lignin unamended conditions. As bacterial biomass increased overtime, Fe(II) accumulated in both conditions to similar concentrations (**Fig 4.4**). There was no change in Fe(II/III) concentrations in abiotic controls. This suggests that Fe(III) was bound to the lignin and was reduced to Fe(II) by BRL6-1. Electron paramagnetic resonance (EPR) analysis further supported these findings, with detection of Fe(III) in the >10 kDa supernatant fractions which decreased in concentration from lag phase to late stationary phase (**Fig 4.5**).

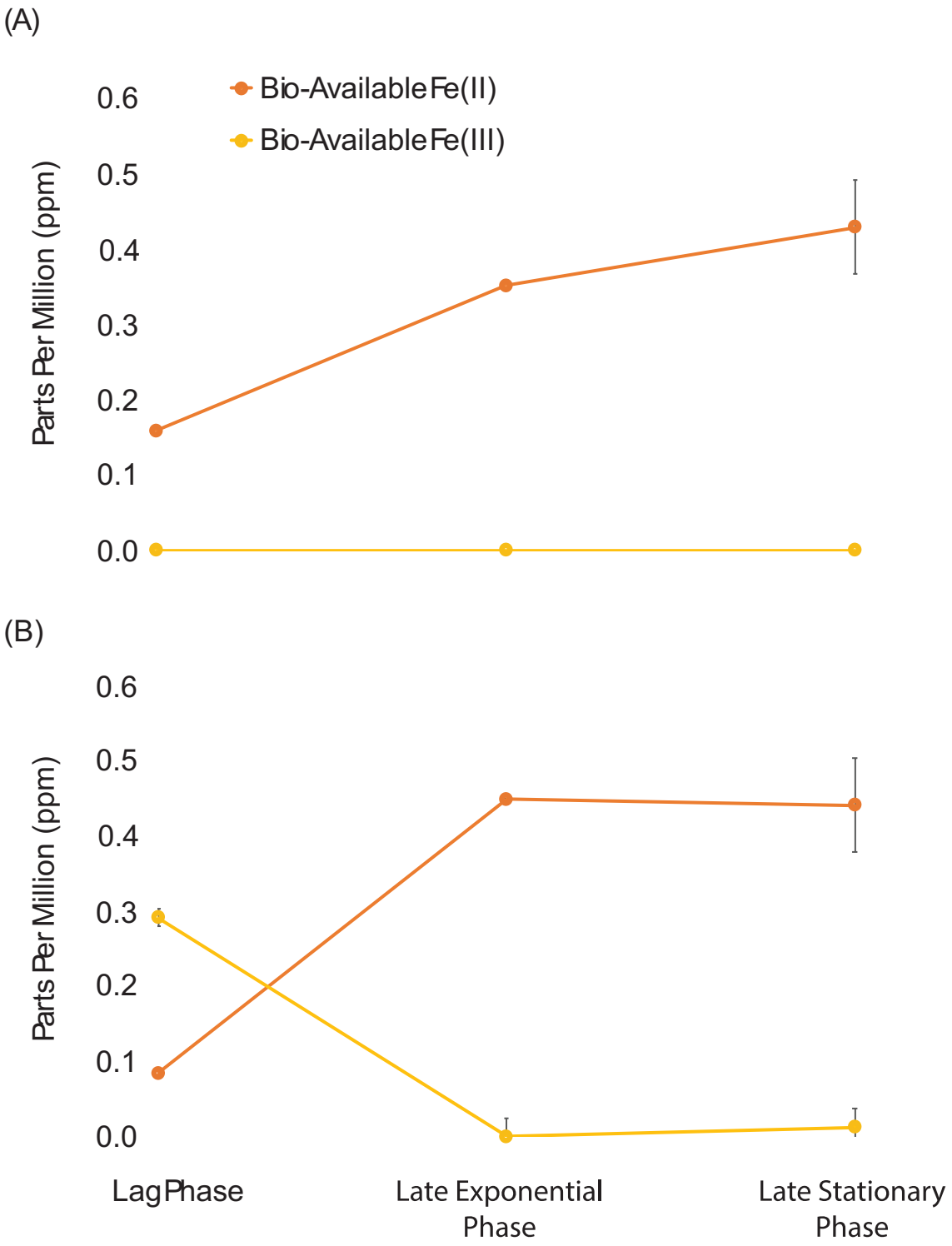


Figure 4.4. Bio-available Fe(II) (orange) and Fe(III) (yellow) concentrations in parts per million (ppm) at lag phase, late exponential phase, and late stationary phase of BRL6-1 grown in lignin amended conditions (**A**) and un-amended conditions (**B**).

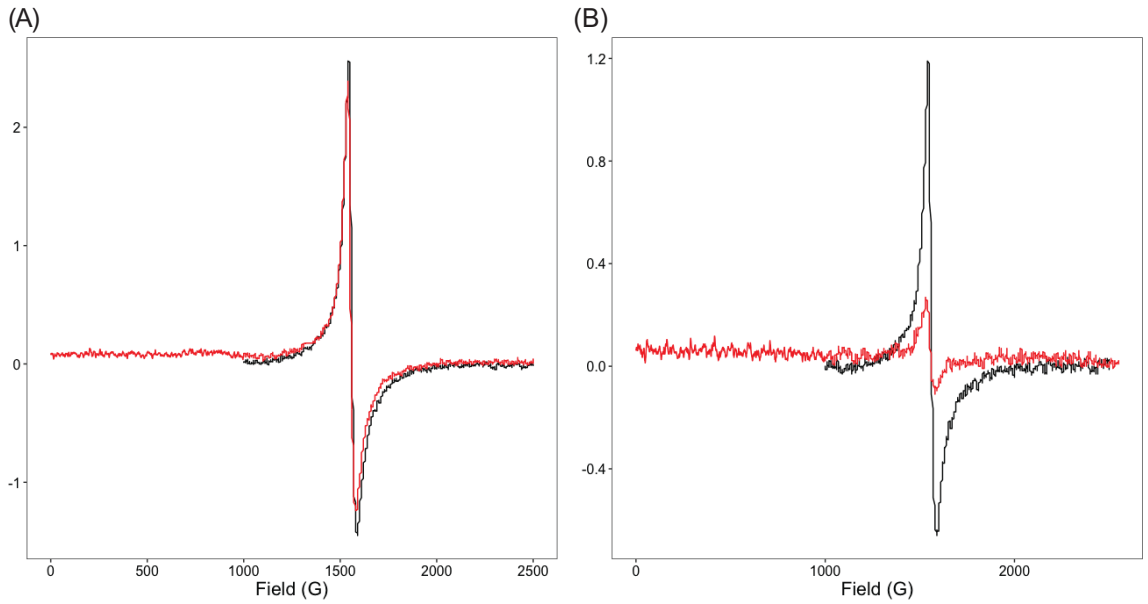


Figure 4.5. EPR measurement of Fe(III) for >10kDa supernatant fraction of abiotic control (**A**) and biological sample (**B**). Samples harvested at lag phase is in black and sample harvested in late stationary phase is in red. Field G is magnetic field strength and the y-axis is signal strength for Fe(III). The Fe(III) concentration does not change in signal in the abiotic control throughout the growth curve whereas Fe(III) decreases over time in the biological sample.

With evidence for Fe(III) bound to lignin and being reduction to Fe(II) based on ferrozine assays, we next wanted to determine the mechanism that BRL6-1 may use to accomplish this redox. One potential was the use of siderophores, which are organic molecules used by bacteria to chelate Fe(III) in the environment (163). The most common siderophore used by bacteria are catecholates, which rely on hydroxyl groups of the catechol rings to form the iron chelation center (147). To see if BRL6-1 produces this type of siderophore, we completed Arnow assays on <10 kDa supernatant fractions from lignin amended and unamended cultures during lag phase, late logarithmic growth phase, and mid-stationary growth phase. Catecholate detection was seen only in lignin amended conditions; however, there was no change in concentration of catechol over the course of the growth curve (**Fig**

4.6). Additionally, abiotic controls of the lignin amended conditions had similar concentrations to biotic replicates. This is likely due to Arnow assays being non-specific between catecholates and compounds containing catechol, such as soluble lignin (164), making it difficult to differentiate sources as well as any small changes in concentrations. BRL6-1 may also be producing other groups of siderophores such as hydroxamates or carboxylates (163), which would need to be detected with a Csáky assay or the use of phenolphthalein and sodium hydroxide, respectively (165,166).

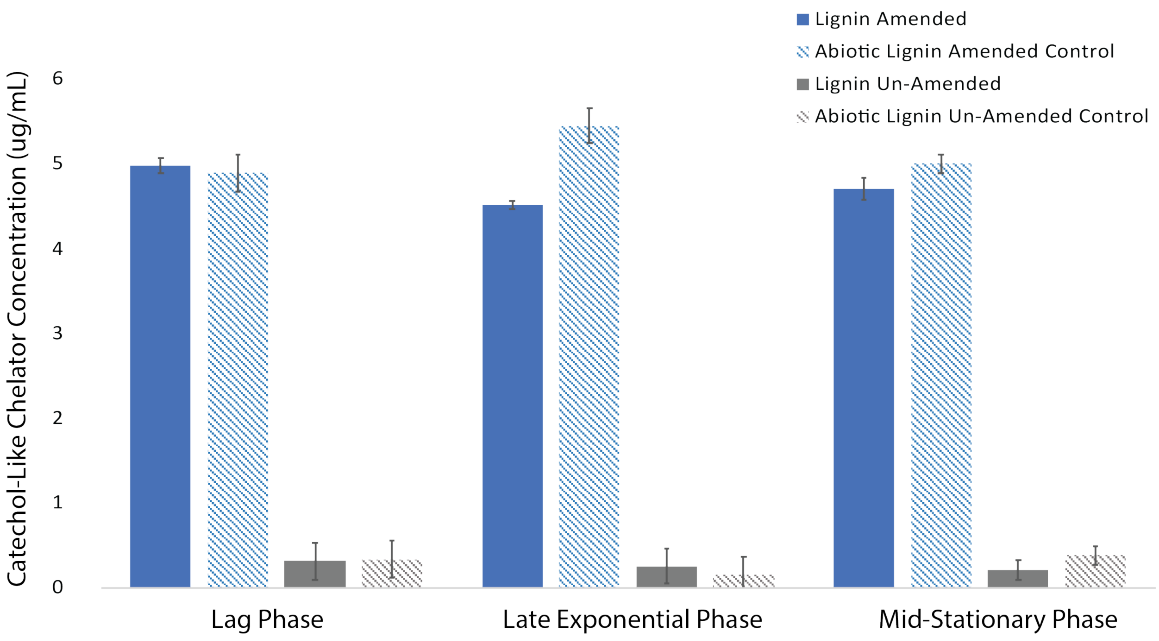


Figure 4.6. Catechol-like chelator concentrations (μg/mL) at lag, late exponential, and mid-stationary phase of BRL6-1growth under lignin amended (blue) and un-amended conditions (gray). Abiotic controls are striped for both conditions.

We also investigated to see if BRL6-1 secreted lignolytic enzymes or iron reducing proteins in the presence of lignin as seen for aerobic fungi and bacteria

(25,26,135). Samples from late stationary phase were run on an SDS-PAGE and silver stained to detect differential banding between the two growth conditions. Differential banding was at 20kDa under lignin amended conditions (**Fig. 4.7**). In addition, because BRL6-1 has had previously predicted peroxidases in its genome (140) and that lignin peroxidases are 35-48 kDa (167), we also were interested in the bands at 37 and 50kDa. Therefore, bands were cut out at 50 kDa, 37 kDa , and 20 kDa for both conditions to identify the proteins present.

A protein originally annotated as hypothetical, WP_024871222.1 was detected in all three lignin-amended biological replicates with a predicted size of 20.8 kDa. There were no conserved domains detected in this protein, but Position-Specific Iterated (PSI) BLAST analysis of the protein identified homology to hypothetical protein from *Alteromonadales bacterium* BS08 (53% Identity; E-value 4e-61). BS08 was isolated from the gut of *Bankia setacea*, also known as the shipworm, that digests wood as a food source (168). Additionally, WP_024871222 had homology to enzymes in the radical SAM superfamily (37.7% Identity; E-value 0.06). Therefore, it is possible, that WP_024871222.1 has a role in lignin modification via radical formation.

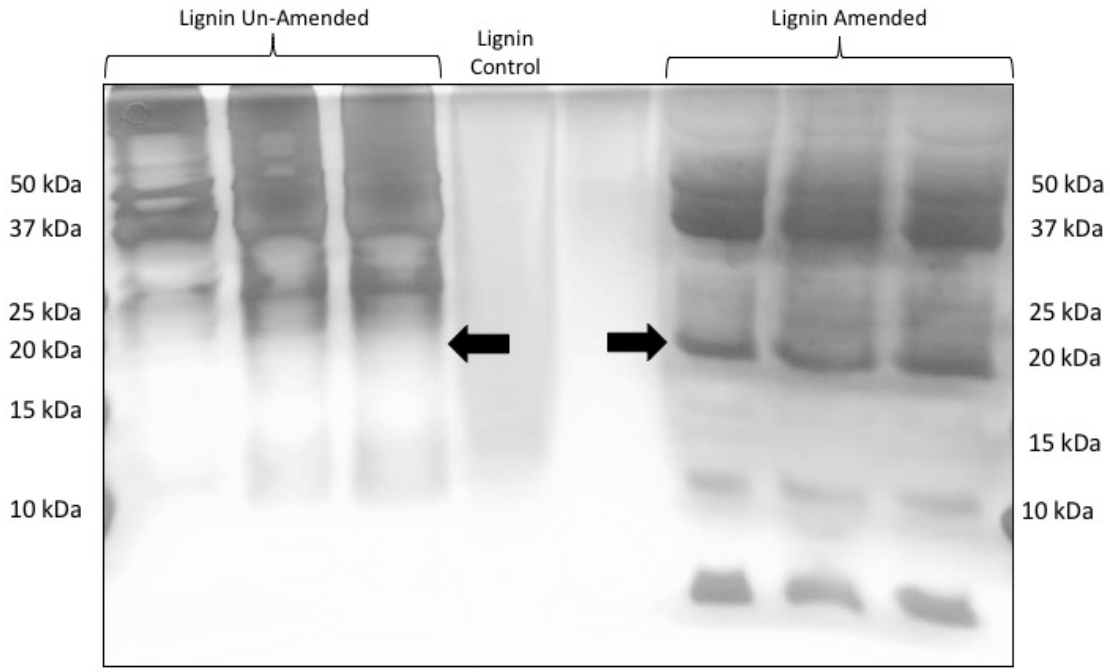


Figure 4.7. SDS-PAGE of *T. lignolytica* BRL6-1 grown amended and un-amended on lignin. Arrows showing differential banding at 20kDa.

To determine how lignin structure was altered by the presence of BRL6-1, <10 kDa fractions of supernatant from lag phase and late stationary phase were analyzed with FTICR-MS (**Fig 4.8**). FTICR-MS spectra yielded 6,794 peaks (including peaks that were unassigned) between samples, with 5,276 common peaks shared between lag phase and late stationary phase. The mean standard deviation of peak intensities for lag phase samples is 1.95e6 and for late stationary phase samples, 1.48e6, both within the maximum level of the noise (1,000,000 to 2,000,000 typically), meaning that the differences in peak intensity within the replicates are likely predominantly due to noise. Based on these results, there was no evidence of enzymatic lignin depolymerization in the presence of BRL6-1. It is possible however, that BRL6-1 is transiently depolymerizing the lignin structure as seen for some brown rot fungi with chelator-mediated Fenton chemistry (169). Since the average

molecular weight of Sigma Aldrich Kraft lignin is 10 kDa based on manufacturer information, majority of the lignin biopolymer and any alterations present would be in the >10 kDa fraction. Therefore, changes in linkages or side chains cannot be detected with FTICR-MS in the <10 kDa fractions. The reasoning for not analyzing the >10 kDa fractions was that the highest published mass of a lignin compound observed by FTICR-MS was reported to be 4 kDa (170) and may not be the best method to analyze the larger polymers of lignin for structural changes. Alternatively, we suggest that NMR analysis be used for the >10 kDa fraction in a future studies for lignin structural changes (171).

To verify that BRL6-1 produced organic radicals in the presence of lignin, we analyzed >10 kDa fractions of one biotic and abiotic samples with EPR (**Fig 4.9**). Organic radicals were detected in lignin amended conditions with differences in intensity between biotic and abiotic controls. However, further analysis is required to determine the cause for changes being seen between abiotic and biotic samples as well as from lag phase and late stationary phase. Future directions also include identifying the radical compound.

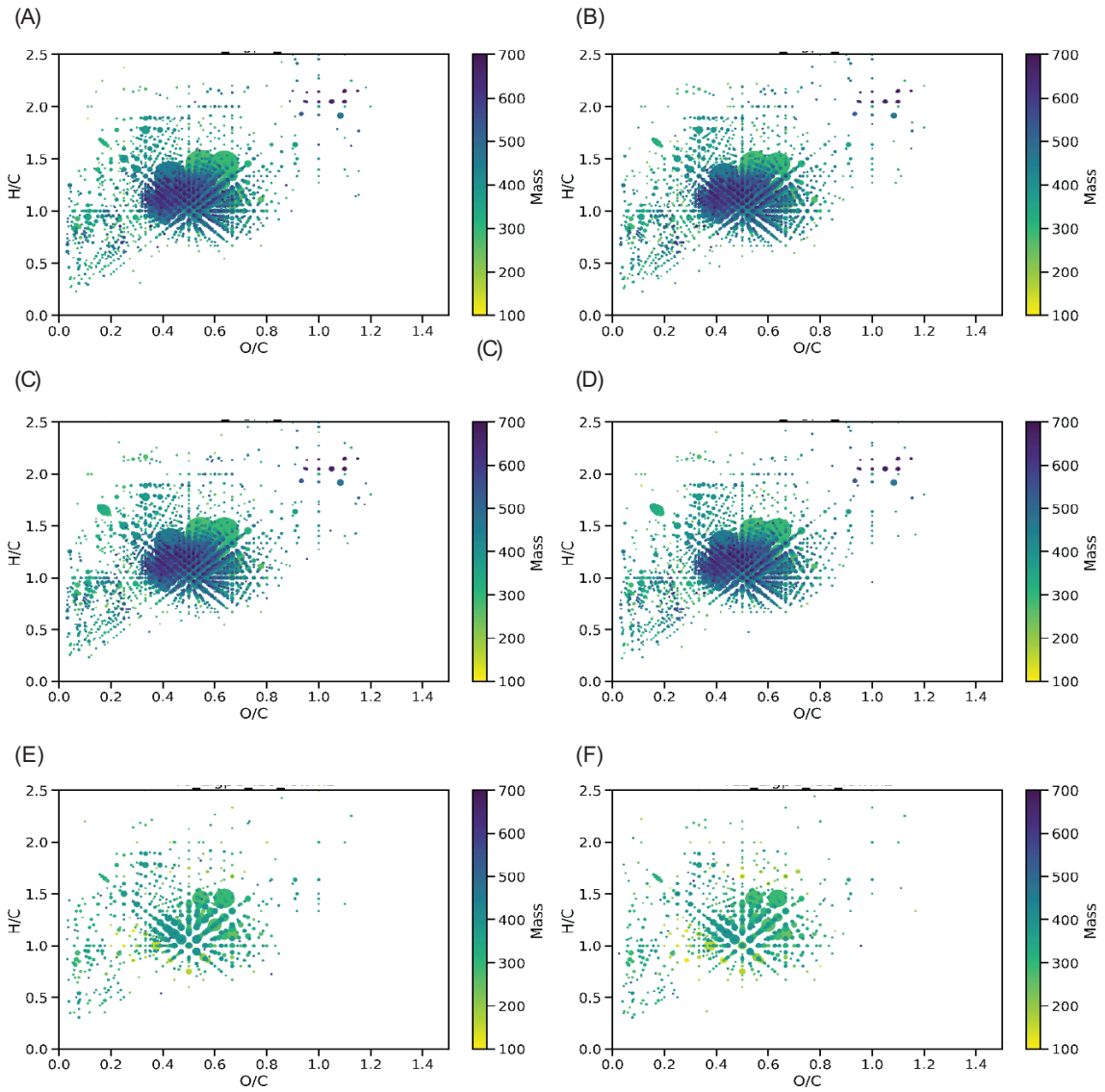


Figure 4.8. Mass changes of lignin before (A, C, and E) and after (B, D, and F) BRL6-1 growth based on FTICR-MA analysis. Depicted as van Krevelen diagrams where each point represents an assigned monoisotopic peak, with its position calculated from the hydrogen-to-carbon ratio (H/C) and oxygen-to-carbon ratio (O/C). Peak intensity is indicated by size of circle and color reflects mass.

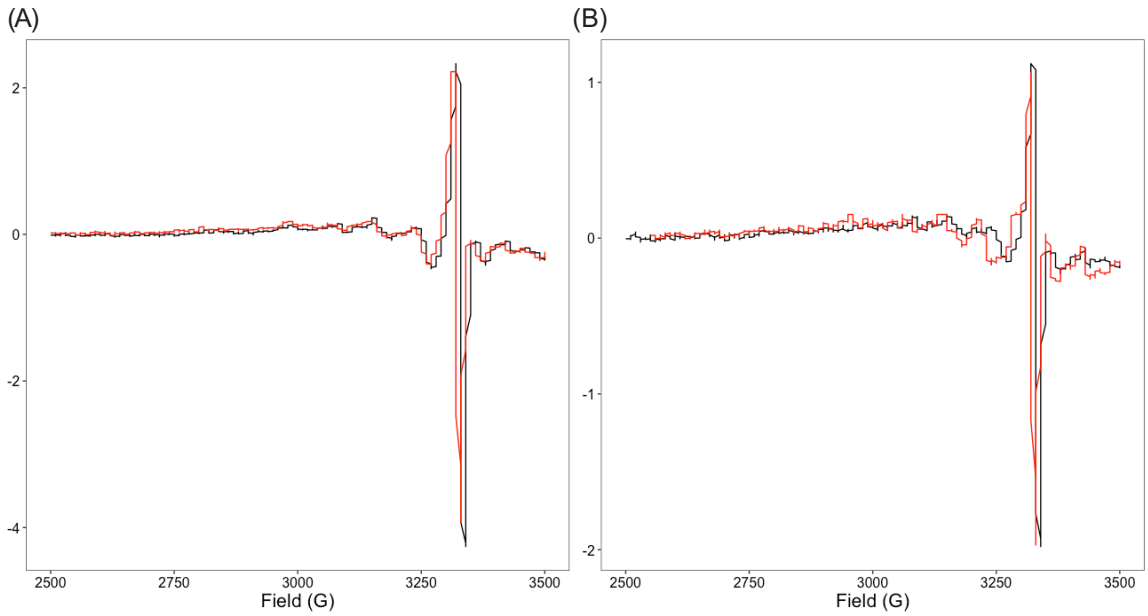


Figure 4.9. EPR measurement of organic radical signals for >10kDa supernatant fraction of abiotic control (**A**) and biological sample (**B**). Sample harvested at lag phase is in black and sample harvested in late stationary phase is in red. Field G is magnetic field strength and the y-axis is signal. A signal for organic radicals is detected in both the abiotic control and biological sample; however, no difference is detected between time points.

4.5 Conclusion

Our analysis supports the hypothesis that *T. lignolytica* BRL6-1 is producing a protein that acts as both an iron chelator and redox agent under anoxic conditions to obtain the iron bound to lignin. Ferrozine and EPR analysis demonstrated that Fe(III) is stripped from lignin and reduced to Fe(II) in the presence of BRL6-1. In addition, organic radicals were detected in the lignin amended conditions based on EPR – though their chemical identify has yet to be identified. BRL6-1's mechanism of radical formation would be similar to that of Fenton reactions where ferrous iron acts as a catalyst to generate free radicals from hydrogen peroxide (25), but instead free radicals are formed from organic compounds, which are in turn able to modify and depolymerize lignin, making it available as a potential carbon source for BRL6-

1. Therefore, BRL6-1 possesses a promising mechanism for industrial applications to remove lignin from lignocellulosic material that would be more cost effective than aerobic microbial mechanisms.

CHAPTER 5

SUMMARY

Lignin is the second most abundant carbon polymer on earth and despite having more fuel value than cellulose, it currently is considered a waste byproduct in many industrial lignocellulose applications (172). For example, in the paper pulping industry about 98% of lignin is burned with the remaining being discarded through waste water effluent (7). The reason behind lignin being "untapped" resource thrown away is due to the recalcitrant nature of the polymer, making it difficult to separate and process for valuable downstream products such as fragrance, dyes, fuel, and other secondary chemical metabolites (35,131). From an environmental perspective, lignin as well as other aromatics like PAHs and xenobiotics that are discarded through industrial effluent can cause eutrophication of water systems and have long term effects on both ecosystems and human health (43). By implementing microbes to depolymerize and convert lignin and other harmful aromatic compounds to valuable products, it possible to reduce the environmental impact as well as economically benefit from a more sustainable production of aromatic chemicals (18).

Investigation of microbial mediated processes for the depolymerization of lignin have focused predominantly on aerobic fungi and bacteria (20,21,35,127,132). Enzymes such as laccases and peroxidases (132,133) as well as chelator-mediated Fenton chemistry (CMF) are used by aerobic microbes to depolymerize lignin (25,26,134,135). These mechanisms have been studied in the

lab as well as small scale production sites to determine the feasibility of applying them towards lignin valorization (23,29,133). However, there are limitations of these microbial processes that prevent it from outcompeting other lignin removal processes. Both aerobic fungi and bacteria require constant aeration and mixing, making it very costly to maintain the cultures (29). Mass production of fungal or bacterial lignolytic enzymes are also not possible due to lacking a method of recycling the enzymes after one use, low substrate specificity, and low redox potential (23). This dissertation investigates anaerobic bacteria as a promising alternative source of enzymes and microbes that are applicable to consolidated depolymerization of lignin and its conversion to valuable byproducts.

In the first project we asked (1) which bacteria have the capability for anaerobic aromatic metabolism, and (2) is vertical inheritance or horizontal transfer driving the phylogeny of anaerobic aromatic metabolic pathways? Seven of the nine known anaerobic aromatic central intermediate pathways (17,32) were analyzed to determine if the they were phylogenetically conserved. Our results determined that benzoyl-CoA catabolism under anoxic conditions had a strong phylogenetic signal (Fritz and Purvis D) and a moderate clade depth (consenTRAIT $\boldsymbol{\tau}_D$) that was significantly non-random, supporting that vertical inheritance has had a stronger role in its phylogeny. Conversely, resorcinol, HHQ, and HBA pathways have strong evidence for horizontal gene transfer and microdiversity driving their evolution, likely due to increased anthropogenic inputs of aromatic contaminants (88). With this information, the next steps forward would be to test if benzoyl-CoA can be predicted for uncharacterized taxa with phylogeny-based prediction

algorithms. Benzoyl-CoA is the most common intermediate for anaerobic bacteria, with many lignin derivatives being funneled into benzoyl-CoA via peripheral pathways (17). Therefore, identification of novel isolates that are predicted to metabolize benzoyl-CoA could lead to new peripheral mechanisms for biotechnological applications such as bio-pulping and lignin valorization.

For the second project, we enriched soil consortia on organosolv lignin as a sole carbon source under anoxic conditions to identify novel bacterial isolates with capabilities of anaerobic lignin depolymerization, catabolism, or both. To determine the lignolytic potential the bacteria had, isolates were screened for clearing zones on lignin-mimicking dyes, malachite green and Congo red. Strain 159R displayed no lignolytic activity and was less than 97% identical in rRNA gene sequence to its closest known relatives, *Sodalis praecaptivus* HS and *Sodalis glossinidius*. This suggested that strain 159R may be a novel *Sodalis* species and that it relies on other members in the microbial community to depolymerize lignin in the soil to acquire aromatic monomers as a carbon source (112). Genome sequencing revealed a genome size of 6.38 Mbp and a G+C content of 54.9 mol%. As expected, many genes relating to intracellular anaerobic and aerobic aromatic metabolism were present in its genome, including the genetic potential to catabolize vanillate and catechol. Pairwise whole genome average nucleotide identity (ANI) and estimated genome-sequence based digital DNA–DNA hybridization (dDDH) values further supported the rRNA gene sequence results that strain 159R represented a new *Sodalis* species. Our phylogenetic analysis revealed that the phylogenetic position of strain 159R is more distantly related to the *Sodalis* clade than close-relative, *Biostraticola tofi*.

Additionally, synteny block coverage was greater in genomes of *Sodalis* members and *B. tofi* compared to strain 159R, suggesting that the former are subsets of the strain 159R genome. However, percentage of conserved proteins (POCP) supported that strain 159R was indeed part of the *Sodalis* genus and that *B. tofi* may possibly be a *Sodalis* member too. Genome size and phylogenetic evidence suggest that strain 159R may be an evolutionary precursor to *Sodalis* endosymbionts as well as free-living *S. praecaptivus* HS, consistent with the genomic streamlining observed in the evolutionary adaptation of other organisms to obligate endosymbiosis (99,102,108). Future directions would be to experimentally confirm aromatic compound utilization under anoxic conditions with labelled substrate to identify proteins involved as potential targets for lignin valorization and biopulping applications.

The third project elucidated the role lignin has in *Tolumonas lignolytica* BL6-1 metabolism and fitness under anoxic conditions. In the presence of lignin, BRL6-1 had a higher biomass and shorter lag phase compared to un-amended conditions, with 14% of the upregulated proteins by $\log_2$ fold-change of 2 or greater relating to Fe^{2+} transport in early exponential phase. Transient iron accumulation in *Salmonella enterica* serovar Typhimurium is required in order for the cells to come out of lag phase (162) and so we hypothesized the up-regulation of iron enzymes might be due to BRL6-1 having the same iron requirement. However, lignin has a strong affinity for iron (160) and therefore we would expect that iron is less bioavailable to the cells in lignin amended conditions (161). Ferrozine assays of the <10kDa supernatant fractions confirmed that Fe(III) was bound to lignin, but it was reduced to Fe(II) when BRL6-1 was present, suggesting redox activity by the cells. To explain

this redox activity, we hypothesized that BRL6-1 is producing a small molecule or protein that acts as both an iron chelator and redox agent under anoxic conditions to obtain the iron bound to lignin. Secretome (extracellular enzyme) analysis coupled with LC-MS/MS identified the presence of a protein of unknown function but had homology to enzymes in the radical SAM superfamily, suggesting that it may have a role in radical formation in lignin amended conditions. Protein isolation and characterization are needed to confirm that this protein interacts with the Fe(III) bound to lignin and reduces it to Fe(II) for cellular use. In addition, further analysis with electron paramagenetic resonance (EPR) as well as nuclear magnetic resonance (NMR) are necessary in order to confirm that organic radicals are being produced in the process of this enzymes obtaining iron and that these radicals altering lignin structure.

The work presented in this dissertation advances the effort in identifying isolates that can perform anaerobic lignin depolymerization and catabolism through bioinformatics as well as traditional culturing techniques. The first project lays the foundational work in support of using phylogeny-based prediction models to identify uncharacterized taxa that have the trait of anaerobic benzoyl-CoA metabolism. The second project and third project give examples of how different isolates can have separate roles in lignin depolymerization and utilization. Using methodological approaches such as protein expression, metabolite production, and chemical structural analysis of lignin can give a comprehensive outlook of how microbes interact with lignin and each other. This information can be used to develop and enhance anaerobic aromatic depolymerization and catabolism

mechanisms for biotechnological applications such as biopulping and biofuel production from lignocellulosic material.

APPENDIX A

PHYLOGENETIC CONSERVATION OF ANAEROBIC AROMATIC METABOLISM

SUPPLEMENTAL DATA

Seven of the nine known central intermediate pathways were analyzed to determine if anaerobic aromatic metabolism is phylogenetically conserved in Chapter II. Enzymes from the modified β-oxidation reaction, acyl-CoA hydratase, hydroxyacyl-CoA dehydrogenase, and oxoacyl-CoA hydrolase, were chosen for benzoyl-CoA and its analogs (50,51,64). For the phloroglucinol pathway, only one enzyme has been identified, the phloroglucinol reductase (65). Central intermediate resorcinol and analog, α-resorcylate each have enzymes identified (RehLS, resorcinol hydroxylase and DbhLS, 3,5-dihydroxybenzoate hydroxylase, respectively) that convert these compounds to hydroxyhydroquinone (HHQ)(17). HHQ is another central intermediate that is further converted by two enzymes in the oxidative pathway, BtdLS, an HHQ dehydrogenase, and BqdLMS, an HBQ dehydrogenase. DbhLS and RehLS are considered the "Resorcinol pathway" for this analysis and BtdLS and BqdLMS are grouped for the "HHQ pathway"

Once genomes were identified containing one or more of these enzymes based on HMMER analysis, they were further screened using an E-value cut-off based on the strict aerobe genus *Acinetobacter* that should not contain any of the enzymes of interest (68). Any organism whose protein had an E-value equal to or greater than any *Acinetobacter* species was removed (**Appendix A, Table 1**). The dataset was then screened for the presence or absence of genes related to anaerobic

respiration using IMG JGI Function Profile and selected KEGG IDs modified from the list of Llorens-Marès *et*

al. 2015 (**Appendix A, Table 2**). If an organism had at least one set of selected KEGG IDs, it was retained.

Table A.1. Sequences used in HMMER for each enzyme of interest (either multiple alignment sequences or NCBI Genbank ID), which pathway the enzyme is involved in, and the E-value cut-off used in creating the database of positive genotypes. E-value was based on *Acinetobacter* acting as an indicator that should not contain any of the enzymes of interest. Hydroxyhydroquinone (HHQ) enzymes were not able to be filtered using *Acinetobacter* due to the E-value being zero. Instead, enzymes of the HHQ pathway were later screened using KEGG IDs for anaerobic respiration (Appendix A, Table 2).

Enzyme	Sequence	E-value Cut-Off
Benzoyl Co-A Pathway		
BCA acyl-CoA hydratase	>SP|O87873|DCH_THAAR/7-257 LKVWLERDGSLLRLRLARPKANIVDAAMIAAMRQALGEHLQAPALRAVLLDAEGPHFSFGASVDEH MPDQCAQMLKSLHGLVREMLDSPVPILVALRGQCLGGGLEVAAAGNLLFAAPDAKFGQPEIRLGVF APAASCLLPPRVGQACAEDLLWSGRSIDGAEGHRIGLIDVLAEDPEAAALRWFDEHIARLSASSLRFA VRAARCDSVPRIKQKLDTVEALYLEELMASHDAVEGLKAFLEKRSANWENR >RF|YP_385104.1/6-256 LKVWLEKDGALLRLRLARPKANIVDAAMIAALQAALTEHLPSAKLRAVLLDAEGPHFSFGASVEEH MPESCAAMLQSLHALVIQMLESPVPVLVAVRGQCLGGGLEVVAAGNLIFAAPGAMLGQPEIKIGVF APAASCLLPERIGKTASEDLLFSGRSITAEEGFRIGLVTAVAEDPEQAAVAYFDEHLAGLSASSLRFAV RAARIGVLERTKTKIAAVEKLYLEELMATHDAVEGLNAFLGKRPAAWQDR >RF|YP_421505.1/9-259 LKVWKDREGKLLRLRLSRPKANIVDAEMIAALSAALGDAHEDSALRAVLIDHEGPHFSFGASVAEH MPDQCAAMLASLHKLVIAMVDFPLPILVAVRGQCLGGGLEVALAGHMMFVSPDAKLGQPEIVLGVF APAASCLLPERMPRVAAEDLLYSGRSIDGAEAARLGIANAVVDDPENAALAWFDNGPAKHSAASLR FAVKAARLGMNERVKAKIAEVEALYLNGLMATHDAVEGLNAFLEKRPALWEDR	$3e^{-31}$
BCA hydroxyacyl-CoA dehydrogenase	>OMNI|NTL01AE3009/8-353 TWQMTEPGK-LQKTRVPMPELGSGDVVVKIAGCGVCHTDLSYFYMGVPTVQKPPLSLGHEISGTII--- GGEASMIGKEVIVPAVIPCGECELCKTGRGNRCLAQKMPGNSMGIYGGYSSHIVAQSKYLCVVEN---- RGDTPLEHLAVVADAVTTPYQAAVRADLKKDDLVIVVGAAGGVGSFMVQTAKGMGAKAVIGIDIN EEKLEMMKGFGADFIINPKDK-SAKEVKELFKGFCKE RGLPSNYGWKIFEVTGSKPGQELALSLLSFTGKLVIVGYGTAETNYMLSKLMAFDAEIIGTWGCPPD RYAAVRDMCLDGRIQLGPFVETRPMSQIEHVFDEAHHGKLKRRVILTP >gi|19571180/20-368 RWMMTSPGAPMVRAEFEIGELSADQVVVAVAGCGVCHTDLGYYYDSVRTNHALPLALGHEISGRV VQAGANAAQWLGRAVIVPAVMPCGTCELCTSGHGTICRDQVMPGNDIQ-- GGFASHVVVPARGLCPVDEARLAAAGLQLADVSVVADAVTTPYQA VLQAGVEPGDVAVVIGV-GGVGGYAVQIANAFGA-SVVAIDVDPAKLEMMSKHGAALTLNAREI- SGRDLKKAIEAHAKANGLRLT- RWKIFECSGTGAGQTSAYGLLTHGATLAVVGFTMDKVEVRLSNLMAFHARALGNWGCLPEYYPAA LDLVLDKKIDLASFIERHPLDQIGEVFAAAHAHKLTRRAILTP >OMNI|NTL06MM2144/25-374	$3e^{-41}$

	RWMMTGVGQPMVKEAMEIAAPGAGEVLVEVAGCGVCHTDLDYYYNGVRTNHALPLALGHEISGR VIQAGAGAESWVGKAVIISAVIPCGQCDLCKRGKGTICRSQKMPGNDLQ-- GGFATHITVPANGLCAVDEARLKAAGLELSEVSVVADALTTPYQAAVQAGIGQGDLVIVIGC- GGVGGYSVQVASAMGA- TVVALDIDPVKLEAVKAAGAKLTLNPKDFPSTREIKKEIGAFAKAQGLRST- EWIIMECSGSVPGQQSAFDLMVHGCTICVVGYTMNKAEFRLSNLMAFHARALGNWGCPPDLYPGA LDLVLSGKINVKNFVERRPLDSINDTFAAVHDHKLSRRAVLCP	
BCA oxoacyl-CoA hydrolase	>OMNI|NTL01AE3010/12-371 IKDHALMGEEHFGTEAPSVL- FEKRPVTDPQGNVVPGLYAAWIILNNPKQYNSYTTEMVKAIIAGFQRASSDRTIVAAVFTAVGDKAFC TGGNTAEYASYYAQRPNEYGEYMDLFNAMVDGILNCKKPTICRVNGMRVGGGQEIGMATDLTITSD MAIFGQAGPKHGSAPDGGSTDFLPWMLNMEDAMYNCISCEPWSAYKMKSKNLITKVVPVLKKDGE WVRNPLVRTDAYVDD-GELV YGEPVAADKAKAAKELIAQCTTDFAKLDEAVDALVWKFANLFPQCLIKSIDGIRGKKKFFWDQMKLA NRHWLAANMNHEAYLGFTAFNN-KKATGKDVIDFIKFRQLVAEGHAFDDAFAEQVL >OMNI|NTL01GM2088/16-376 LNDHNLIDRE- VESLCDGMVKYEKRPAKRHDGSVAEGIYNAWIILDNPKQYNSYTTDMVKAIILAFRRASVDRSVNAV VF TGVGDKAFCTGGNTKEYAEYYAGNPQEYRQYMRLFNDMVSAILGCDKAVISRVNGMRIGGGQEIGM ACDFSIAQDLANFGQAGPKHGSAAIGGATDFLPLMVGCEQAMVSGTLCEPFSAHKAARLGIICDVVPA LKVGGKFVANPTVVTDRYLDEYGRVVHGEFKAGAAFKEGQGQIKEGEIDLSLLDEKVESLCTKLLETF PECMTKSLEELRKPKLHAWNLNKENSRAWLALNMMNEA RTGFRAFNEGTKETGRE-IDFVKLRQGLAKGTPWTEELIESLM >OMNI|NTL06MM2143/17-372 LNDHNLV----PTTVVPGVL- YEKRPAKRADGTVAEGLYNAWITLDNQKQYNSYTTDMVKGVIMAFRDASNARDVSSVVF TGAGDKAFCTGGNTKEYAEYYAGNPQEYRQYMRLFNDMVSAILGCDKPVICRVNGMRIGGGQEIGM AADFSVAQDLAKFGQAGPKHGSAPIGGATDFLPVMIGCEQAMVSGSLCEPWSAHKAYRTGIIMDLVP ALKVDGKFVANPLVITDRYLDEFGKIVHGESKTGAELAAGKELLKKGTIDLSLLDAKVEEICAKILHTF PDCFTKTIQELRKPKLNAWNANKENSRDWLGLNMMTEARTGFRAFNEGPKE-DRE- IDFVALRQALAKGAPWTPELIESLI >gi|3724166/17-373 LVDHNLV----PETVCPGVL- YEKRPARNLKGEVVPGLYNVWISLDNPKQYNSYTTDMVKGLILAFRAASCARDVASVVF TAVGDKAFCTGGNTKEYAEYYAGNPQEYRQYMRLFNDMVSAILGCDKPVICRVNGMRIGGGQEIG MAADFTVAQDLANFGQAGPKHGSAAIGGATDFLPLMIGCEQAMVSGTLCEPFSAHKANRLGICMQI VPALKVDGKFIANPLVVTDRYLDEFGRIIHGEFKTGDELAAGKELMKRGEIDLSLLDEAVEKLCAKLI STFPECLTKSFEELRKPKLDAWNRNKENSRAWLALNMMNEARTGFRAFNEGNKETGRE- IEFTDLRQALAKGMPWTPELIESLM	$3e^{-31}$
3-Methylbenzoyl-CoA Pathway		
3-MBA hydratase	CCH23021.1	$1.5e^{-80}$
3-MBA dehydrogenase	CCH23023.1	$7.9e^{-93}$
3-MBA hydrolase	CCH23022.1	$8.0e^{-94}$

4-Methylbenzoyl-CoA Pathway		
4-MBA hydratase	AIW63094.1	$3e^{-82}$
4-MBA dehydrogenase	AIW63095.1	$7.4e^{-83}$
4-MBA hydrolase	AIW63096.1	$1.3e^{-94}$
3-Hydroxybenzoyl-CoA Pathway		
HBA acyl-CoA hydratase	>WP_050418522.1 enoyl-CoA hydratase/isomerase family protein [Azoarcus sp. CIB] MISLRIEDS----------VATVTLCRAPV- NAINEEWIAAFDRILAELEHTPRVNVLWIRSAERVFCAGADL- DVIGSLFATEAGRVQMIAITRRMQQLYARLERLPQVTVAEIGGAAMGGGFELALACDLRVVADSAK VGLPEARLGLLPAA- GGTQRMTRICGEAVARRLILGAEVVGGVDAVKLGCAHWVAPAAELEEFTRGVVTRIAALPALALSE CKRCITVAVEGD-EDGYQVELAGSAALLADGETQQRVRAFLNR-------- >WP_011236223.1 enoyl-CoA hydratase/isomerase family protein [Aromatoleum aromaticum] MISLTIEAS----------VATVTLCRSPV-NAINEEWIEQLDRILAEIERTPRVNVLWIRSGERVFCAGADL- ELIRSLFDSETGRRQMIAMTRRMQEVYARLERLPQVSVVEIGGAAMGGGFELALACDLRVVADSARI GLPEARLGLLPAA- GGTQRMTRICGEAVARRLILGAEVIGGAEAVALGCAHWVAPAAELESVARAVVERIAALPGTALAE CKRCIDVAVAAE-ENGFEVELSGSAALLADAETQRRVQRFLDKQRQ----- >CAC28159.1 putative hydrolase [Thauera aromatica] MSVVLVEQPTPD--------- VAVVRLNRPDARNALNQEVRSALAEHFDRLGQAAEVRCIVLTGGERCFAAAPDIRAM------ADAG-- AIEIMLRQTQRLWQAIAACPKPVIAAVNGYAWGGGCELAMHADIIIAGEGASFCQPEVKVGIMPGA- GGTQRLTRAVGKFQAMKMVLTGLPVSARERLAMGLASEVVADDAVQARALELARHIATLPPLAIA QIKEVLLAGQDASLDTALMLERKAFQLLFASADQKEGMRAFLEKRPPVFRGG >CAC28155.1 unnamed protein product [Thauera aromatica] MYKLKAADWHPEHFKLEVANRVATITLNRPDRKNPLTFESYAELRDTFHKFQYVDDVRSIVITGAG GNFCSGGDVHDIIGPLTKMDMN-- GLLTFTRMTGNLVKEMRTCPQPIISAIDGICAGAGAIVSMASDMRYATPDAKTAFLFVRVGLAGCDM GACAILPRIIGHGRASELLYTGRVMSAQEGQAWGYFNDLVAPDQVLAKAQEMALSLANGPAFAHA MTKKCLHQEWDMSIEQALETEAEAQAICMQTQDFTRAYNAFVAKQKPVFEGN >WP_050418021.1 enoyl-CoA hydratase family protein [Azoarcus sp. CIB] MYKLKAAEWRPEHFKLEVADRVATITLNRPERKNPLTFESYAELRDTFIKLQYAEDVRAVVMTGAG GNFCSGGDVHDIIGPLTKMDMT-- GLLAFTRMTGNLVKEMRNCPQPIISAVDGVCAGAGAIITMASDLRYATPEAKTAFLFVRVGLAGCD MGACSILPRIIGQGRASELLYTGRSMSAEEGRAWGYFNDVVPAEKVLAKAQEMALSLANGPAFAHS VTKKCLHQEWNQTIEQALETEAEAQAICMQTEDFTRAYNAFVNKQVPKFEGN >WP_011236224.1 enoyl-CoA hydratase family protein [Aromatoleum aromaticum] MYKLKAAEWRPEHFKLEVADRVATITLNRPERKNPLTFESYAELRDTFHKLQYVDDVRTVVITGAG GNFCSGGDVHDIIGPLTKMDMN-- GLLTFTRMTGNLVKEMRNCPQPIISAVDGICAGAGAIVSMASDLRYATPEAKTAFLFVRVGLAGCDM GACSILPRIIGHGRASELLYTGRSMSAEEGRAWGYFNDIVPAEKVLGRAQEMALSLANGPAFAHSMT KKCLHQEWNQTIEQALETEAEAQAICMQTQDFTRAYNAFVNKQVPKFEGN	$1e^{-79}$

HBA hydroxyacyl-CoA dehydrogenase	>WP_050418028.1 SDR family oxidoreductase [Azoarcus sp. CIB] MTADSGRALAGKHVVITGGGRGIGAAIAAALSAQGARLTLMGRNRGQLEER-- AAVLRTLGGESCEVHCEAVDVADEASVVSAFAAAAKRLGPVAVLVNNAGQAGSAPFLRTESALWQ QMLAVNLTGTYLATRAALPDMLAAG- WGRIINVASTAGEKGYPYVTAYCAAKHGVIGLTRSLALELAHKHVTVNAVCPGYTDTDIVRDAVTN IREKTGRSEAEALAELAKHNPQGRLVRPEEVANAVLWLCLPGSDAITGQAISVSGGEVM-- >CAC28156.1 putative alcohol dehydrogenase [Thauera aromatica] --MTHSRALSGKHAVITGGGRGIGAAIAHSLAEQGAAVTLMGRTLPRLEQQ--AEELRAFSQ---- VHCEAVDVAQADSVAAAFAAAQARLGPVDILVNNAGQALSAPFVKTDPALWQQMLDVNLTGVFL GTRAVLPGMLAAG- WGRVINITSTAGQKGYPYVSAYCAAKHGVIGLTRALALETARKNVTVNAVCPGYTDTDIVRDSVSNI QTKTGRSEAEALAELTRFNPQGRLVRPQEVANAVLWLCLPGSEAITGQSISVAGGEMM-- >WP_041646819.1 SDR family oxidoreductase [Aromatoleum aromaticum] -----MRELSGKHAVVTGGGRGIGAAIAQRLAEQGACVTLMGRRREPLEER-- ADALRALIGVHCDMHCEAVDVADPASVAAAFDAAARRFGPVSILVNNAGQASSAPFVKTDLALWQ RMLDVNLTGTYLGTKAVLSGMLAAG- WGRIVNVASTAGQKGYPYVSAYCAAKHGVIGMTRALALELAQKNITVNAVCPGYTDTDIVREAITN IRAKTGRSEAEAQGELAKHNPQGRLVRPDEVANAVLWLCLPGAEAITGQAISVSGGEVM-- >CAC28154.1 putative alcohol dehydrogenase [Thauera aromatica] ------MRLEGKTAVVTGGASGIGRATAETLAAAGAHVVI----- GDLDQEKGAAVAAAIRESGRKADYFPLDVTSLDSVGVFAKAVEENGLEVDIVVNVAGWGKIQPFM ENSPDFWRKVIDLNLLGPVAVTHAFLGGMIARGRGGKVITVASDAGRVGSTGETVYSGAKGGAIAF GKALAREMARYKINVNSVCPGPTDTPLLAAVPEKHQE---------- AFVKATPMRRLGKPSEIADAVLFFASSDSDFITGQVLSVSGGMTMVG >WP_011236225.1 SDR family oxidoreductase [Aromatoleum aromaticum] ------MRLDGKTAVVTGGASGIGLATAETLARAGAYVLI----- GDIDEQKGAAVAGALCEQQLGVDFIRLDVTDLDSIAAFKDEAYRRRPQIDIVANVAGWGKIQPFMEN TPDFWRKVIDLNLLGPVAVSHAFLPQMIERG- AGKIVTVASDAGRVGSLGETVYSGAKGGAIAFTKSLAREVARYNINVNCVCPGPTDTPLLQAVPEKH RE----------AFVKATPMRRLAKPSELADAVLFFASDRASFITGQVISVSGGLTLAG >WP_050418022.1 SDR family oxidoreductase [Azoarcus sp. CIB] ------MNLQGKTAVVTGGASGIGYATAETLARAGAKVVI----- GDIDAAKGAAAAGMLAEQHLDVDFVRLDVTDIDSIHAFRDETYRRHPQVDIVANVAGWGKIQPFME NTPDFWRKVIDLNLLGPVAVSHAFLQQMIERG- SGKIVTVSSDAGRVGSLGETVYSGAKGGAIAFTKSLAREVARYNINVNCVCPGPTDTPLLQAVPEKH RE----------AFVKATPMRRLAKPSELADAVLFFASDRASFITGQVISVSGGLTLAG	$2.2e^{-80}$
HBA oxoacyl-CoA hydrolase	>CAC28157.1 putative acyl-CoA dehydrogenase [Thauera aromatica] MSEKSYLEWPFFEDRHRKLEAELDSWATNNISEHH- GELDSACRELVAKLGAAGWLRYCVGGTSYGGEHETIDTRSICLLRETLARHSGLADFAFGMQGLGS GAITLHGSDAQKREYLPRVASGQALAAFALSEPGSGSDVAAMACSARLDGEYYVLDGEKSWISNGG IADFYVVFARTGEAPGARGLSAFIVDADTPGLEIAERIEVIAPHPLARLRFTDCRVHKSAMLGTPGLGF KVAMQTLDIFRTSVAAAALGFSRRALDEALRRATTREMFQQKLADFQITQVKLAQMATSVDISALLT YRAAWRRDQGHKVTREAAMAKMTATESAQQVIDSAVQIWGGCGVVSNHPVELLYREIRALRIYEG ATEVQQLIIARQTLTAYEDS--- >WP_050418027.1 acyl-CoA dehydrogenase [Azoarcus sp. CIB]	$1.1e^{-75}$

	MSDRSYLEWPFFEERHRGMQVELEAWAAAHIDGHPHGDLDDACRELVRKLGADGWLRYMVGGTA YGGRHDTIDTRAVCLLRETLARHSGLADFALGMQGLGSGAITLHGTDAQKRKYLSEVAAGRAIPAF ALSEPDSGSDVAAMACSARRDGNDYVLDGEKTWISNGGIADFYVVFARTGEAPGARGLSAFIVEAN LPGFEIAERIDVIAPHPLARLRFTGCRVPAANLLGAPGQGFKVAMQTLDIFRTSVAAAALGFARRALD EGLRRATTRDMFGKKLADFQITQAKLAQMATHVDTAALLTYRAAWMRDQGKNITGAAAMAKMTS TETAQQVIDAAVQLWGGCGVVSEHPVERLYREIRALRIYEGATEVQQLIIARQTLSAWEQEQAV >WP_011236231.1 acyl-CoA dehydrogenase [Aromatoleum aromaticum] MSDQTYLEWPFFDEPHRQLQIELEAWASANVTEHHGSDLDTACRELVAKFGAAGWLRYVVGGTAY GGCHDVIDTRAVCLLRETLGRHSGLADFAFGMQGLGSGAITLHGTDAQKRDYLPRVASGRAIAAFA LSEPGSGSDVAAMACSARQDGDEYVIDGEKTWISNGGIADFYVVFARTGEAAGSRGLSAFIVDADRP GLEIAERIDVIAPHPLARLRFRECRVPKSCLLGVPGQGFKVAMQTLDIFRTSVAAAALGFARRALDEA LKRATTRDMFGQKLADFQITQAKLAQMATAVDTSALLTYRAAWLRDQGQTITGAAAMAKMTSTET AQQVIDAAVQMWGGCGVVSDHPVERLYREIRSLRIYEGATEVQQLIIARQTLSAYERQQEH	
Resorcinol Pathway		
3,5-dihydroxybenzoate hydroxylase large subunit (DbhL)	AIO06084.1	$6.6e^{-256}$
3,5-dihydroxybenzoate hydroxylase small subunit (DbhS)	AIO06085.1	$7.1e^{-75}$
Resorcinol hydroxylase large subunit (RehL)	ABK58620.1	$1.8e^{-218}$
Resorcinol hydroxylase large subunit (RehS)	ABK58619.1	$1.9e^{-56}$
Hydroxyhydroquinone Pathway		
Benzoquinone Dehydrogenase BqdL	>AIO06095.1 benzoquinone dehydrogenase alpha subunit [Thauera aromatica] MPKTIDLHYHAPWQEVVATADDWDHLGSATVLRMLHHLHLVRAFEETVLELDGEGLVHGPAHSSI GQDGGAVGAVSLLRSSDLITGSHRGHHQFLAKCLAHLDRGEADPRRTPLSEGVRTMLYRALAEILGL ADGYCRGRGGSMHLRWAEAGALGTNAIVGGGVPLATGAAWACKRRGAGDVAFTFLGDGAVNIGA VPESMNLAALWSLPVCFFIENNGYAVSTKLSEETRETRLSSRGGAYGIPALRVDGMDPVAVRVATQ MALDAMRAGQGPYIIEAEVYRYFHHGGGLPGSAFGYRSKDEEAAWRARDPLACLARGMIERDWLS ADEDATLRAGARACMVEIAARLTEKDGSKRRIVPALWPQATFRDEGVRGDLAELAGVRCEELETAS GKVGEVKFISAVAGVMARRMESDERIFCLGEDIHKLNGGTNGATRGLAARFPDRIVPTPIAEQGFVG LAGGVAMEGHYRPVVELMYADFALVAADPLFNQIGKARHMFGGDMAVPLVLRSKCAIGTGYGSQ HSMDPAGLYAMWPGWRIVAPSTPFDYVGLMNSALQCDDPVLVIEHVGLYNTTAPGPLEDFDYYIPL GKAKVVRPGTALTVLTYLAMTPLAVKVADELGVDAEVIDLRSLDRAGIDWETIGDSVRKTNNVVVL EQGSQTASYGAMLADEVQRRLFDHLDQPVKRIHGGEAAPNVSKVLERAAFVGAEEVRAGFIEVLAD AGRPLAQTAPALG----- >ABK58621.1 dehydrogenase [Azoarcus anaerobius]	

	MPRITNLDYAEPWIELASTPQDWKKLGKTELLRVLYYHHLVRAFEEAVLNLEKLGLVHGPAHSSIGQ EGGAVGSVMLLNSSDMITGAHRGHHQFLVKGMQHIDSPSYDPRAAPLPEEVQTFLYRTLAEILGLSD GFCKGRGGSMHLRWVEAGAMGTNAIVGGGVPIANGLAWAQKRRNKGEVTFTFFGDGGMNIGAVP ESMNLAALWNLPICFFIENNGYAVSTTLEEETRETRLSSRGGAYAIPAWRVDGMDPVAVRLASEAAI ERMRAGKGPTIIEAVLYRYFHHGGSVAGSAFGYRKKDEESSWIAKDPLDRTVREMINLQWLTADEN TAIRRHCESAMQGIVERLVEGEGSKRRIRAELWPKPEFRDQGLRGDLSEFKDARFEELETASGPVGD VKFVDAVARVMGRRMETDERVFCMGEDIHRLKGGTNGATKGLAERFPDRIIPAPIAEQGFVGLAGG VAQDGQYRPVVELMYSDFALVAADQLFNQIGKARHMFGGDSAVPLVLRTKCAIGTGYGSQHSMDP AGMYAMWPGWRIVAPSTPFDYVGLMNSALKCEDPVLVIEHTDLYNTTDQGPLEDLDYCIELGKAK VVRKGSAFTVLTYLAMTPLALKVADEMGLDVEIIDLRSLDRAGIDWATIGESIRKTNNVVVLEQGPL TVSYGAMLTDEIQRRFFDYLDQPVQRIHGGESSPSVSKVLERAAFVGAEEIRAGFTRMMADMGQPLP ATPSPAGNSITA	
Benzoquinone Dehydrogenase BqdS	>AIO06106.1 benzoquinone dehydrogenase small subunit [Thauera aromatica] MPVEILMPSTGASMSEGNILRWLKQEGEAVERGEALLEIETDKAVVEAVTPARGILGKILAAGGSEG VKVDSVVGLIAVDGEDPVALAGAVLAGATPAGSAPAGAATVATA--------- AGEASPAEVQRRIPASPLARRLARETGVDLAAVRGRGPHGRVLRADVESVARQAAAAAAPGGAAPL LAATVAAAGTAVPSAAGAAFEDIPHSAMRRVIAQRLGEAKRTVPHFYLSLDCAVDALLALRAQINA QLDAQVGAQVGAQVGAHPDGGKLSVNDFIVKAVALALRRVPGCNAAWTEAAVRRFAEVDIAVAV ATPGGLITPIVRHADDKSLGSLSAEIRALAGRAREGRLKPEEYQGGGFTLSNLGMYGIREFAAIINPPQ ACILAVGACEQRPVVRDGSLAVATLMSCTLSVDHRVVDGAQAAEFLAEFRRLIENPLAILV >ABK58622.1 dihydrolipoamide acetyltransferase [Azoarcus anaerobius] ------MPSVSTSMTEGTLARWLKKDGETVAKGEVIAEIETDKAILEVEAEAEGIFKAFVADGAT-- VKVGEPMGALLAPGETLGGTISAAQSAAAPTAAAVGGETAVAVAVAAPAAAPSTGHAPAAHDGTRI FASPLARSLALLHGLDLVNISGSGPQGRIVKRDIEA- AMSAQRPASGAVAAPVAEAPVKAPQPAAPQAAGAGYELIPHSSMRRVIAQRLSESKQQVPHFYLTV DCRLDKLLALRQQVN----------GSLPD- VKVSVNDFIVKAVAAAMKRVPATNASWSDEGVRRYRDIDISVAVATPNGLITPVVRQADAKSVGTI SAEVKDLAERARQGKLKPDEYQGGGFTISNLGMYGVRDFAAIINPPQACILAVGTAEKRPVIEDGAIV PATVMTCTLSVDHRVVDGAVGAEFLAAFKALLETPLGLLV	
Benzoquinone Dehydrogenase BqdM	>ABK58623.1 putative dehydrogenase E3 component [Azoarcus anaerobius] - MAQEKFDLTVIGGGPGGYVAAIRAAQLGLRTALIEREHLGGICLNWGCIPTKALLRSAEIFDHFKHAG DFGLEVQGASFDLQKIVARSRGVAAQLNAGVKHLLKKNKVQVFEGSGRLAG SGTIRLEQKDG- VSEIQSTHIILATGARARAMAPVEPDGRLVWSYKEAMTPERMPKSLLIVGSGAIGIEFASFYRSLGAE VTVVEVRDRVLPVEDAEVSAFAHKAFERQGMKLLTSSSVVSLQKQADSVIAVIDTKGTTTEIRADRV IAAVGIVGNVENLGLEGTGVQVENTHIVTDAWCQTGEPGVYAIGDVAGAPWLAHKASHEGILCVER IAGVDGIHPLDKTRIPGCTYSRPQIASIGLTEAQAKERGYELKVGRFPFMGNGKAIALGEPEGFIKTVF DAKTGELLGAHMVGAEVTELIQGFSIGKTLETTEAELMHTVFPHPTLSEMLHEATLAAYGRAIHT >AIO06092.1 dihydrolipoamide dehydrogenase family protein [Thauera aromatica] MTDNNSYDLIVVGAGPGGYVAAIRAAQLGMKTAVVEREHLGGICLNWGCIPTKALLRSAEVGRLAR HAAEYGVSVPEPKFDLERIVQRSRAIAAQLNGGIRHLLNKNKVSVIEGEARLAGAGRVAVTRGGAD AGTYAAPHLILATGARARQLPGLEDDGRLVWTYRKAMTPDVLPKSLLIVGSGAIGIEFASFYHALGS QVTVVEVMDRILPVEDEDISALARKAFEDQGMRILTGAKASIARKSAECVTVRIEAGGAAEELTVDR VIVAVGISPNTENLGLEHTRVRLERGHIVTDPWCRTDEPGLYAIGDVTRPPWLAHKASHEAMICVEAI AGLADVHPLELRNIPGCTYSHPQIAS	

	VGLTERKAREQGHEVRVGRFPFVGNGKAIALGEPEGLVKTVFDARSGELLGAHMIGAEVTELIQGYT LARTLEATEAELIATVFPHPTLSETMHEAVLAAYGRAIHI	
HHQ dehydrogenase large subunit (BtdhL)	ABK58630.1	
HHQ dehydrogenase small subunit (BtdhS)	ABK58631.1	
Phloroglucinol Pathway		
Phloroglucinol Reductase	>WP_014184752.1 SDR family oxidoreductase [Desulfosporosinus orientis] MVDIQ-- FVNNLFDVKDKVALITGATGALGKAISFGYGLAGMKIFVTGRSGEKCKALCDELEAQGIECGYSIGD PAVEADVIKVVEDAVQKFGEINVLLTAAGYNHPQPIVDQDLAEWKKIMDSDVQGTWLFCKYAGQQ MIERGKGGKVILVSSARSKMGMAGYTGYCTAKAGIDLMAQSLACEWTAKYKINVNTINPTVFRSDL TEWMFDPESPVYANFLKRLPVGRLGEPEDFIGPCIFLASNASDFMTGANVATEGGYWAN >WP_021630531.1 SDR family oxidoreductase [Clostridium sp. ATCC BAA-442] MVNVKKEFVDNMFSVKGKVALVTGATGALGCVLSKAYGYAGAKVFMTGRNEKKLQALEAEFKAE GIDCAYGVADPADEAQVDAMITACVAQYGEVNILAVTHGFNKPQNILEQSVADWQYIMDADCKSV YVVCKYVAQQMVDQGKGGKIVVVTSQRSKRGMAGYTGYCTSKGGADLMVSSMACDLSAKYGIN VNSICPTVFRSDLTEWMFDPESAVYQNFLKREPIGRLAEPEDFVGYALFLSSDASNYITGANCDCSGG YLTC >WP_027868985.1 SDR family oxidoreductase [Eubacterium sp. AB3007] MVNVEKSFVNNMFSVEGKVALVTGATGALGCVLSKAYGYAGAKVFMTGRNAEKLQKLQDEFEAE GIDCAYFVADPQKEEDVKALIAACVEKYGEVNILAICHGYNKPANILDQSVEDWQFIMDADCKSVYI VCKYVAEQMVEQGKGGKMVVVTSQRSKRGMAGYTGYCTSKGGADLMVSSMACDLTAKYGINVN SICPTVFRSELTEWMFDPDSEVYKNFLKREPIGRLAEPYDFVGFALFLSSEASDFMTGGNYDCSGGYL TC	$2.2e^{-46}$

Table A.2. Marker genes used to indicate that organisms are capable of anaerobic respiration (Modified from Llorens-Marès *et al.* 2015).

Step	**KEGG/TIGRfam ID**	**Gene**
Anaerobic C fixation: Arnon pathway	K00174 K00175 K00244 K01648	2-oxoglutarate:ferredoxin oxidoreductase subunit alpha; 2-oxoglutarate:ferredoxin oxidoreductase subunit beta
Anaerobic C fixation: Reductive citric acid cycle	K00194 K00197	CO dehydrogenase subunit delta; CO dehydrogenase subunit gamma
Fermentation	K00016	L-lactate dehydrogenase

Methanogenesis	K00400 K00401	coenzyme M methyl reductase beta subunit (mcrB); methyl coenzyme M reductase system, component A2
Anammox	K10535	hydroxylamine oxidoreductase/hydrazine oxidoreductase (hao/hzo)
Denitrification	K00376 K02305 K04561	nitrous oxide reductase (nosZ); nitric-oxide reductase (norC); nitric-oxide reductase (norB)
N-fixation	K00531 K02586 K02588 K02591	Nitrogenase; nitrogenase molybdenum-iron protein alpha chain (nifD); nitrogenase iron protein (nifH); nitrogenase molybdenum-iron protein beta chain (nifK)
Selenium Respiration	K17050 K17051 K17052	SerABC
Assimilatory Sulfate Reduction	K00860 K00956 K00957	adenylylsulfate kinase (cysC); sulfate adenylyltransferase subunit 1 (cysN); sulfate adenylyltransferase subunit 2 (cysD)
Dissimilatory Sulfate Reduction	K00394 K00395 K11180	adenylylsulfate reductase subunit A (aprA); adenylylsulfate reductase subunit B (aprB); sulfite reductase (dsrA)
Polysulfide Reduction	K08352	polysulfide reductase chain A (psrA)
Iron Respiration	TIGR03058 TIGR03509 TIGR03507	MtrABC, OmcAB

APPENDIX B

TOLUMONAS LIGNOLYTICA BRL6-1 ANAEROBIC GROWTH ON LIGNIN DERIVED MONOMERS

The goal of this project is to develop a metabolic map in *Tolumonas lignolytica* BRL6-1 using ^{13}C-labelled lignin model compounds suspected to be capable of supporting microbial growth. Our contribution was to find anoxic conditions that would support strain growth on lignin-derived monomers as sole carbon source. To do so, we performed growth curves, analyzing growth on monomers as the sole carbon source or in addition to glucose. In all anoxic conditions tested, BRL6-1 was unable to grow in the presence of monomers, regardless of pH or glucose addition (**Fig. 1**). In most anoxic conditions tested, the presence of the monomers was toxic to cell growth.

BRL6-1 was grown on modified CCMA media consisting of 2.25g L^{-1} NaCL, 0.5g L^{-1} NH_4Cl, 0.227g L^{-1} KH_2PO_4, 0.348g L^{-1} K_2HPO_4, 5mg L^{-1} $MgSO_4*7H_2O$, 2.5mg L^{-1} $CaCL_2*2H_2O$, 0.01 mL L^{-1} SL-10 trace elements, 0.01mL L^{-1} Thauer's vitamins, and 30mM of PIPES (pH 7) or MES (pH 5.5). Growth was monitored by absorption (OD_{600}) as a measure of cell density. Lignin monomers at pH 7 included benzoic acid, ferulic acid, guaiacol, vanillic acid, 4-hydroxybenzoic acid, and 3,4 dihydroxyphenylacetic acid (DOPAC). Lignin monomers at pH 5 included vanillic acid, 4-hydroxybenzoic acid, and 3,4 dihydroxyphenylacetic acid (DOPAC). All monomers were added to media at a final concentration of 5 mM. Heat labile compounds were made as liquid stocks, which were filter sterilized, not autoclaved.

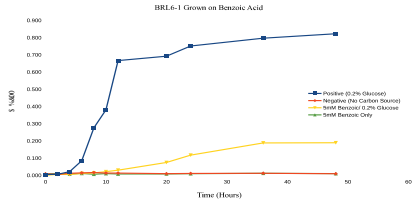
For all growth experiments, a positive control was grown in parallel containing 0.2% glucose only.

We have been working with lignin derived model compounds known to be capable supporting microbial growth in previous literature; however, these are only 6 monomers of a highly complex biopolymer. Some monomers are known to be toxic depending on the microbe (173) and/or require specific transporters (174), such as a C4-dicarboxylic acid/H+ symporter required for *Pseudomonas chlororaphis* O6 required for benzoate uptake (156) or *Rhodopseudomonas palustris* that has various transporters that each have specificity to different benzoate derivatives depending on the side chains (175).

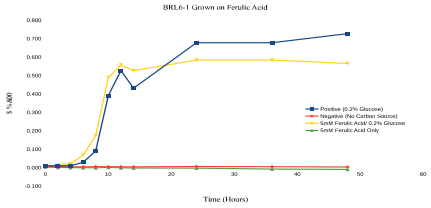
Monomer growth was predicted based on genome sequence analysis, so lack of growth on monomers was a surprise. It is possible that monomer concentrations in nature are much lower than in our growth conditions, meaning that metabolism of lignin monomers in solution is possible but not robust. If this is the case, then an experimental approach using monomers in culture may not be conducive to stable isotope probing and downstream analyses.

An alternative approach to defining metabolic pathways of lignin degradation is to grow BRL6-1 in media amended with ^{13}C-lignin. Looking at the metabolites present in the biomass may give insight to what is being taken up by the cell to more efficiently target monomer candidates.

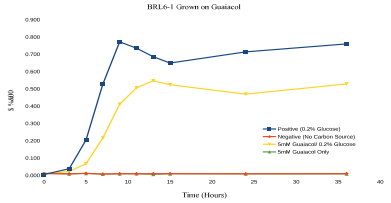
(A) BRL6-1 Grown with 0.2% Glucose and 5mM Benzoic Acid



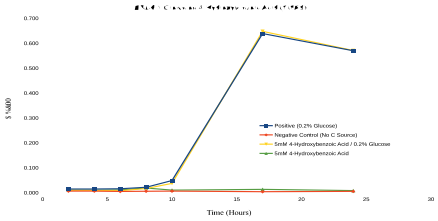
(B) BRL6-1 Grown with 0.2% Glucose and 5mM Ferulic Acid



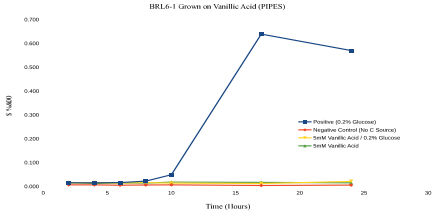
(C) BRL6-1 Grown with 0.2% Glucose and 5mM Guaiacol



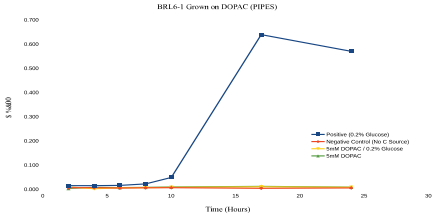
(D) BRL6-1 Grown with 0.2% Glucose and 5mM 4-Hydroxybenzoic Acid (pH 7)



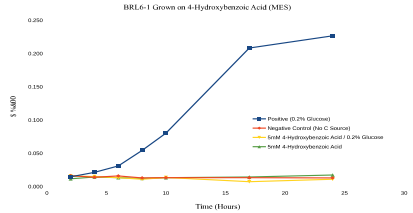
(E) BRL6-1 Grown with 0.2% Glucose and 5mM Vanillic Acid (pH 7)



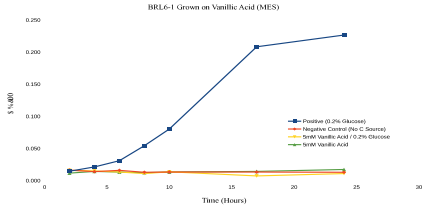
(F) BRL6-1 Grown with 0.2% Glucose and 5mM DOPAC (pH 7)



(G) BRL6-1 Grown with 0.2% Glucose and 5mM 4-Hydroxybenzoic Acid (pH 5.5)



(H) BRL6-1 Grown with 0.2% Glucose and 5mM Vanillic Acid (pH 5.5)



(I) BRL6-1 Grown with 0.2% Glucose and 5mM DOPAC (pH 5.5)

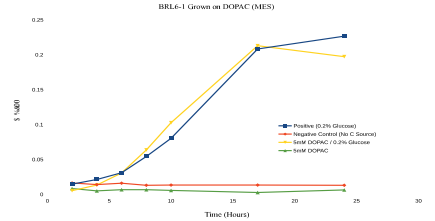


Figure B.1. Anaerobic growth of *Tolumonas lignolytica* BRL6-1 with glucose as the primary carbon source and amended with lignin derived monomers or lignin derived monomers as the sole carbon source. Media was either at a pH 7 (buffered with 30mM PIPES) and 5.5 (buffered with 50mM MES). Lignin monomers at pH 7 included benzoic acid (**A**), ferulic acid (**B**), guaiacol (**C**), 4-hydroxybenzoic acid (**D**), vanillic acid (**E**), and 3,4 dihydroxyphenylacetic acid (DOPAC; **F**). Lignin monomers at pH 5 included 4-hydroxybenzoic acid (**G**), vanillic acid (**H**), and 3,4 dihydroxyphenylacetic acid (DOPAC; **I**). Growth was monitored by absorption (OD_{600}) as a measure of cell density.

BIBLIOGRAPHY

1. US EPA. US Environmental Protection Agency: Municipal Solid Waste (MSW) in the United States: Facts and Figures [Internet]. 2015. Available from: https://www.epa.gov/sites/production/files/2018-07/documents/2015_smm_msw_factsheet_07242018_fnl_508_002.pdf

2. Ashrafi O, Yerushalmi L, Haghighat F. Wastewater treatment in the pulp-and-paper industry: A review of treatment processes and the associated greenhouse gas emission. J Environ Manage. 2015 Aug 1; 158:146–57.

3. Haq I, Raj A. Pulp and Paper Mill Wastewater: Ecotoxicological Effects and Bioremediation Approaches for Environmental Safety. In: Bharagava RN, Saxena G, editors. Bioremediation of Industrial Waste for Environmental Safety: Volume II: Biological Agents and Methods for Industrial Waste Management [Internet]. Singapore: Springer Singapore; 2019 [cited 2019 Jul 26]. p. 333–56. Available from: https://doi.org/10.1007/978-981-13-3426-9_14

4. Gomes FJB, Santos FA, Colodette JL, Demuner IF, Batalha LAR. Literature Review on Biorefinery Processes Integrated to the Pulp Industry. Nat Resour. 2014;05(09):419–32.

5. US EPA. US Environmental Protection Agency: Available and Emerging Technologies for Reducing Greenhouse Gas Emissions from the Pulp and Paper Manufacturing Industry. 2010.

6. Tran H, Vakkilainnen EK. 1 Advances in the Kraft Chemcial Recovery Process. 2006.

7. Pollegioni L, Tonin F, Rosini E. Lignin-degrading enzymes. FEBS J. 2015 Apr;282(7):1190–213.

8. Breen A, Singleton FL. Fungi in lignocellulose breakdown and biopulping. Curr Opin Biotechnol. 1999 Jun;10(3):252–8.

9. Chandra R, editor. Advances in Biodegradation and Bioremediation of Industrial Waste [Internet]. CRC Press; 2015 [cited 2016 Jan 21]. Available from: http://www.crcnetbase.com/doi/book/10.1201/b18218

10. Giles RL, Zackeru JC, Galloway ER, Elliott GD, Parrow MW. Single versus simultaneous species treatment of wood with Ceriporiopsis subvermispora and Postia placenta for ethanol applications, with observations on interspecific growth inhibition. Int Biodeterior Biodegrad. 2015 Apr; 99:66–72.

11. Costa S, Dedola D, Pellizzari S, Blo R, Rugiero I, Pedrini P, et al. Lignin Biodegradation in Pulp-and-Paper Mill Wastewater by Selected White Rot Fungi. Water. 2017 Dec 2;9(12):935.

12. Dashtban M, Chen S, Song R, Qin W. Lignin in Paper Mill Sludge is degraded by White-Rot Fungi in Submerged. In 2015.

13. Li S, Jiang L, Zhang H, Li Z, Wang X. Recycling of lignin: A new methodology for production of water reducing agent with paper mill sludge. 2009;62(5):5.

14. Rashid GMM, Durán-Peña MJ, Rahmanpour R, Sapsford D, Bugg TDH. Delignification and enhanced gas release from soil containing lignocellulose by treatment with bacterial lignin degraders. J Appl Microbiol. 2017;123(1):159–71.

15. Suvorova IA, Gelfand MS. Comparative Genomic Analysis of the Regulation of Aromatic Metabolism in Betaproteobacteria. Front Microbiol [Internet]. 2019 [cited 2019 Apr 25];10. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2019.00642/full

16. Brzeszcz J, Kaszycki P. Aerobic bacteria degrading both n-alkanes and aromatic hydrocarbons: an undervalued strategy for metabolic diversity and flexibility. Biodegradation. 2018 Aug 1;29(4):359–407.

17. Durante-Rodríguez G, Gómez-Álvarez H, Blázquez B, Fernández-Llamosas H, Martín-Moldes Z, Sanz D, et al. Chapter 13. Anaerobic Pathways for the Catabolism of Aromatic Compounds. In: Beckham GT, editor. Energy and Environment Series [Internet]. Cambridge: Royal Society of Chemistry; 2018 [cited 2019 Jul 13]. p. 333–90. Available from: http://ebook.rsc.org/?DOI=10.1039/9781788010351-00333

18. Salvachúa D, Karp EM, Nimlos CT, Vardon DR, Beckham GT. Towards lignin consolidated bioprocessing: simultaneous lignin depolymerization and product generation by bacteria. Green Chem. 2015 Nov 2;17(11):4951–67.

19. Ren T, Qi W, Su R, He Z. Promising Techniques for Depolymerization of Lignin into Value-added Chemicals. ChemCatChem. 2019;11(2):639–54.

20. Bugg TDH, Ahmad M, Hardiman EM, Rahmanpour R. Pathways for degradation of lignin in bacteria and fungi. Nat Prod Rep. 2011 Nov 15;28(12):1883–96.

21. Cragg SM, Beckham GT, Bruce NC, Bugg TD, Distel DL, Dupree P, et al. Lignocellulose degradation mechanisms across the Tree of Life. Curr Opin Chem Biol. 2015 Dec 1;29:108–19.

22. Harwood CS, Parales and RE. THE β-KETOADIPATE PATHWAY AND THE BIOLOGY OF SELF-IDENTITY. Annu Rev Microbiol. 1996;50(1):553–90.

23. Singh G, Arya SK. Utility of laccase in pulp and paper industry: A progressive step towards the green technology. Int J Biol Macromol. 2019 Aug 1;134:1070–84.

24. Singh P, Sulaiman O, Hashim R, Rupani PF, Peng LC. Biopulping of lignocellulosic material using different fungal species: a review. Rev Environ Sci Biotechnol. 2010 Feb 25;9(2):141–51.

25. Ma J, Zhang K, Huang M, Hector SB, Liu B, Tong C, et al. Involvement of Fenton chemistry in rice straw degradation by the lignocellulolytic bacterium Pantoea ananatis Sd-1. Biotechnol Biofuels. 2016;9:211.

26. Wang L, Yan W, Chen J, Huang F, Gao P. Function of the iron-binding chelator produced by Coriolus versicolor in lignin biodegradation. Sci China C Life Sci. 2008 Mar 1;51(3):214.

27. de Gonzalo G, Colpa DI, Habib MHM, Fraaije MW. Bacterial enzymes involved in lignin degradation. J Biotechnol. 2016 Oct 20;236(Supplement C):110–9.

28. Bugg TD, Rahmanpour R. Enzymatic conversion of lignin into renewable chemicals. Curr Opin Chem Biol. 2015 Dec 1;29:10–7.

29. Scott GM, Akhtar M, Swaney RE, Houtman CJ. Recent developments in biopulping technology at Madison, WI. Prog Biotechnol-Amst-. 2002;61–72.

30. Woo HL, Ballor NR, Hazen TC, Fortney JL, Simmons B, Davenport KW, et al. Complete genome sequence of the lignin-degrading bacterium Klebsiella sp. strain BRL6-2. Stand Genomic Sci. 2014 Dec 8;9:19.

31. Orellana R, Chaput G, Markillie LM, Mitchell H, Gaffrey M, Orr G, et al. Multi-time series RNA-seq analysis of Enterobacter lignolyticus SCF1 during growth in lignin-amended medium. PloS One. 2017;12(10):e0186440.

32. Carmona M, Zamarro MT, Blázquez B, Durante-Rodríguez G, Juárez JF, Valderrama JA, et al. Anaerobic catabolism of aromatic compounds: a genetic and genomic view. Microbiol Mol Biol Rev MMBR. 2009 Mar;73(1):71–133.

33. Butler JE, He Q, Nevin KP, He Z, Zhou J, Lovley DR. Genomic and microarray analysis of aromatics degradation in Geobacter metallireducens and comparison to a Geobacter isolate from a contaminated field site. BMC Genomics. 2007 Jun 19;8:180.

34. Goberna M, Verdú M. Predicting microbial traits with phylogenies. ISME J. 2016 Apr;10(4):959–67.

35. Brink DP, Ravi K, Lidén G, Gorwa-Grauslund MF. Mapping the diversity of microbial lignin catabolism: experiences from the eLignin database. Appl Microbiol Biotechnol. 2019 May 1;103(10):3979–4002.

36. Ehrenfreund P, Rasmussen S, Cleaves J, Chen L. Experimentally Tracing the Key Steps in the Origin of Life: The Aromatic World. Astrobiology. 2006 Jun 1;6(3):490–520.

37. González-Gaya B, Martínez-Varela A, Vila-Costa M, Casal P, Cerro-Gálvez E, Berrojalbiz N, et al. Biodegradation as an important sink of aromatic hydrocarbons in the oceans. Nat Geosci. 2019 Feb;12(2):119–25.

38. He S, Lau MP, Linz AM, Roden EE, McMahon KD. Extracellular electron transfer may be an overlooked contribution to pelagic respiration in humic-rich freshwater lakes. 2018 Aug 15 [cited 2018 Nov 24]; Available from: http://biorxiv.org/lookup/doi/10.1101/392027

39. Cabaniss SE, Madey G, Leff L, Maurice PA, Wetzel R. A Stochastic Model for the Synthesis and Degradation of Natural Organic Matter. Part I. Data Structures and Reaction Kinetics. Biogeochemistry. 2005;76(2):319–47.

40. US EPA. 2000 Nonmethane Organic Compounds (NMOC) and Speciated Nonmethane Organic Compounds (SNMOC) Monitoring Program. 2000.

41. Thevenot M, Dignac M-F, Rumpel C. Fate of lignins in soils: A review. Soil Biol Biochem. 2010 Aug;42(8):1200–11.

42. Potapowicz J, Szumińska D, Szopińska M, Polkowska Ż. The influence of global climate change on the environmental fate of anthropogenic pollution released from the permafrost: Part I. Case study of Antarctica. Sci Total Environ. 2019 Feb 15;651:1534–48.

43. Mishra VK, Singh G, Shukla R. Chapter 6 - Impact of Xenobiotics Under a Changing Climate Scenario. In: Choudhary KK, Kumar A, Singh AK, editors. Climate Change and Agricultural Ecosystems [Internet]. Woodhead Publishing; 2019 [cited 2019 Jul 6]. p. 133–51. Available from: http://www.sciencedirect.com/science/article/pii/B9780128164839000062

44. Biswas B, Qi F, Biswas JK, Wijayawardena A, Khan MAI, Naidu R. The Fate of Chemical Pollutants with Soil Properties and Processes in the Climate Change Paradigm—A Review. Soil Syst. 2018 Sep;2(3):51.

45. Holopainen JK, Virjamo V, Ghimire RP, Blande JD, Julkunen-Tiitto R, Kivimäenpää M. Climate Change Effects on Secondary Compounds of Forest Trees in the Northern Hemisphere. Front Plant Sci [Internet]. 2018 Oct 2 [cited 2019 Jul 6];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6176061/

46. Richardson CJ (ORCID:0000000283736587), Flanagan N, Wang H, Ho M, Chanton J, Cooper B. Phenolic compounds and black carbon feedback controls on peat decomposition and carbon accumulation in southeastern peatlands under regimes of seasonal drought, drainage and frequent fire [Internet]. Duke Univ., Durham, NC (United States); 2018 Dec [cited 2019 Jul 6]. Report No.: DOE-Duke-1000. Available from: https://www.osti.gov/biblio/1488733

47. Mukamuhirwa A, Persson Hovmalm H, Bolinsson H, Ortiz R, Nyamangyoku O, Johansson E. Concurrent Drought and Temperature Stress in Rice—A Possible Result of the Predicted Climate Change: Effects on Yield Attributes, Eating Characteristics, and Health Promoting Compounds. Int J Environ Res Public Health. 2019 Jan;16(6):1043.

48. Henrichs SM, Reeburgh WS. Anaerobic mineralization of marine sediment organic matter: Rates and the role of anaerobic processes in the oceanic carbon economy. Geomicrobiol J. 1987 Jan 1;5(3–4):191–237.

49. Ghattas A-K, Fischer F, Wick A, Ternes TA. Anaerobic biodegradation of (emerging) organic contaminants in the aquatic environment. Water Res. 2017 Jun 1;116:268–95.

50. Porter AW, Young LY. Benzoyl-CoA, a universal biomarker for anaerobic degradation of aromatic compounds. Adv Appl Microbiol. 2014;88:167–203.

51. Kuntze K, Shinoda Y, Moutakki H, McInerney MJ, Vogt C, Richnow H-H, et al. 6-Oxocyclohex-1-ene-1-carbonyl-coenzyme A hydrolases from obligately anaerobic bacteria: characterization and identification of its gene as a functional marker for aromatic compounds degrading anaerobes. Environ Microbiol. 2008 Jun 1;10(6):1547–56.

52. Rayu S. The interactions between xenobiotics and soil microbial communities. In 2016.

53. Maphosa F, Lieten SH, Dinkla I, Stams AJ, Smidt H, Fennell DE. Ecogenomics of microbial communities in bioremediation of chlorinated contaminated sites. Front Microbiol [Internet]. 2012 Oct 2 [cited 2019 Jul 10];3. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3462421/

54. Bush T, Diao M, Allen RJ, Sinnige R, Muyzer G, Huisman J. Oxic-anoxic regime shifts mediated by feedbacks between biogeochemical processes and microbial community dynamics. Nat Commun. 2017 Oct 6;8(1):789.

55. Kreft J-U, Plugge CM, Prats C, Leveau JHJ, Zhang W, Hellweger FL. From Genes to Ecosystems in Microbiology: Modeling Approaches and the Importance of Individuality. Front Microbiol [Internet]. 2017 Nov 27 [cited 2019 Jul 10];8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5711835/

56. Díaz E. Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility. Int Microbiol Off J Span Soc Microbiol. 2004 Sep;7(3):173–80.

57. Zamarro MT, Barragán MJL, Carmona M, García JL, Díaz E. Engineering a bzd cassette for the anaerobic bioconversion of aromatic compounds. Microb Biotechnol. 2017 Jul 24;10(6):1418–25.

58. Van Assche A, Álvarez-Pérez S, de Breij A, De Brabanter J, Willems KA, Dijkshoorn L, et al. Phylogenetic signal in phenotypic traits related to carbon source assimilation and chemical sensitivity in Acinetobacter species. Appl Microbiol Biotechnol. 2017 Jan;101(1):367–79.

59. Martiny AC, Treseder K, Pusch G. Phylogenetic conservatism of functional traits in microorganisms. ISME J. 2013 Apr;7(4):830–8.

60. Morrissey EM, Mau RL, Schwartz E, Caporaso JG, Dijkstra P, van Gestel N, et al. Phylogenetic organization of bacterial activity. ISME J. 2016 Sep;10(9):2336–40.

61. Isobe K, Allison SD, Khalili B, Martiny AC, Martiny JBH. Phylogenetic conservation of bacterial responses to soil nitrogen addition across continents. Nat Commun. 2019 Jun 7;10(1):2499.

62. Guénard G, von der Ohe PC, de Zwart D, Legendre P, Lek S. Using phylogenetic information to predict species tolerances to toxic chemicals. Ecol Appl. 2011 Dec 1;21(8):3178–90.

63. Fritz SA, Purvis A. Selectivity in Mammalian Extinction Risk and Threat Types: a New Measure of Phylogenetic Signal Strength in Binary Traits. Conserv Biol. 2010;24(4):1042–51.

64. Porter AW, Young LY. The bamA gene for anaerobic ring fission is widely distributed in the environment. Front Microbiol [Internet]. 2013 Oct 10;4. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3794298/

65. Conradt D, Hermann B, Gerhardt S, Einsle O, Müller M. Biocatalytic Properties and Structural Analysis of Phloroglucinol Reductases. Angew Chem Int Ed. 2016 Dec 12;55(50):15531–4.

66. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.

67. Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, et al. TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res. 2001 Jan 1;29(1):41–3.

68. Abdel-El-Haleem A-E-H. Acinetobacter: environmental and biotechnological applications. Afr J Biotechnol. 2003 Apr 30;2(4):71–4.

69. Tandon D, Haque MM, Mande SS. Inferring Intra-Community Microbial Interaction Patterns from Metagenomic Datasets Using Associative Rule Mining Techniques. PLoS ONE [Internet]. 2016 Apr 28 [cited 2019 Jul 29];11(4). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4849775/

70. Hahsler M, Grün B, Hornik K. arules - A Computational Environment for Mining Association Rules and Frequent Item Sets. J Stat Softw. 2005 Sep 29;14(1):1–25.

71. Csuos M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics. 2010 Aug 1;26(15):1910–2.

72. Felsenstein J. Phylogenies from Molecular Sequences: Inference and Reliability. Annu Rev Genet. 1988;22(1):521–65.

73. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016 Jul 8;44(Web Server issue):W242–5.

74. Orme D, Freckleton R, Thomas G, Petzoldt T. The caper package: comparative analysis of phylogenetics and evolution in R. R Package Version. 2013;5(2):1–36.

75. Mendler K, Chen H, Parks DH, Lobb B, Hug LA, Doxey AC. AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. Nucleic Acids Res. 2019 May 21;47(9):4442–8.

76. Markowitz VM, Chen I-MA, Chu K, Szeto E, Palaniappan K, Pillay M, et al. IMG/M 4 version of the integrated metagenome comparative analysis system. Nucleic Acids Res. 2014 Jan;42(Database issue):D568-573.

77. Nakamura FM, Germano MG, Tsai SM. Capacity of Aromatic Compound Degradation by Bacteria from Amazon Dark Earth. In 2014.

78. Yang J, Yang Y, Wu W-M, Zhao J, Jiang L. Evidence of Polyethylene Biodegradation by Bacterial Strains from the Guts of Plastic-Eating Waxworms. Environ Sci Technol. 2014 Dec 2;48(23):13776–84.

79. Viñas M, Sabaté J, Espuny MJ, Solanas AM. Bacterial Community Dynamics and Polycyclic Aromatic Hydrocarbon Degradation during Bioremediation of Heavily Creosote-Contaminated Soil. Appl Environ Microbiol. 2005 Nov;71(11):7008–18.

80. Pacheco-Sánchez D, Molina-Fuentes Á, Marín P, Medina-Bellver J-I, González-López Ó, Marqués S. The Azoarcus anaerobius 1,3-Dihydroxybenzene (Resorcinol) Anaerobic Degradation Pathway Is Controlled by the Coordinated Activity of Two Enhancer-Binding Proteins. Appl Env Microbiol. 2017 May 1;83(9):e03042-16.

81. Reams AB, Neidle EL. Selection for Gene Clustering by Tandem Duplication. Annu Rev Microbiol. 2004 Oct;58(1):119–42.

82. Lahme S, Eberlein C, Jarling R, Kube M, Boll M, Wilkes H, et al. Anaerobic degradation of 4-methylbenzoate via a specific 4-methylbenzoyl-CoA pathway. Environ Microbiol. 2012 May 1;14(5):1118–32.

83. Kavakiotis I, Tzanis G, Vlahavas I. Mining frequent patterns and association rules from biological data. Biol Knowl Discov Handb Preprocessing Min Postprocessing Biol Data Wiley Book Ser Bioinforma Comput Tech Eng Jersey USA Wiley-Blackwell John Wiley Sons Ltd. 2014;

84. Poudel S, Colman DR, Fixen KR, Ledbetter RN, Zheng Y, Pence N, et al. Electron Transfer to Nitrogenase in Different Genomic and Metabolic Backgrounds. J Bacteriol [Internet]. 2018 Apr 24 [cited 2019 Jan 15];200(10). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5915786/

85. Kanak A. Benzene and Beyond: Mechanisms of Novel Anaerobic Aromatic Degradation Pathways in Geobacter daltonii. 2014;86.

86. Juhasz AL, Stanley GA, Britz ML. Metabolite repression inhibits degradation of benzo[a. J Ind Microbiol Biotechnol. 28(2):88–96.

87. Zimmerman AE, Martiny AC, Allison SD. Microdiversity of extracellular enzyme genes among sequenced prokaryotic genomes. ISME J. 2013 Jun;7(6):1187–99.

88. Suresh S, Srivastava VC, Mishra IM. Adsorption of catechol, resorcinol, hydroquinone, and their derivatives: a review. Int J Energy Environ Eng. 2012 Nov 5;3(1):32.

89. Cabral L, Pereira de Sousa ST, Júnior GVL, Hawley E, Andreote FD, Hess M, et al. Microbial functional responses to long-term anthropogenic impact in mangrove soils. Ecotoxicol Environ Saf. 2018 Sep 30;160:231–9.

90. Moore LR, Rocap G, Chisholm SW. Physiology and molecular phylogeny of coexisting Prochlorococcus ecotypes. Nature. 1998 Jun 4;393(6684):464–7.

91. Jaspers E, Overmann J. Ecological Significance of Microdiversity: Identical 16S rRNA Gene Sequences Can Be Found in Bacteria with Highly Divergent Genomes and Ecophysiologies. Appl Environ Microbiol. 2004 Aug 1;70(8):4831–9.

92. Dale C, Maudlin I. Sodalis gen. nov. and Sodalis glossinidius sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly Glossina morsitans morsitans. Int J Syst Bacteriol. 1999 Jan;49 Pt 1:267–75.

93. Eleftherianos I, Atri J, Accetta J, Castillo JC. Endosymbiotic bacteria in insects: guardians of the immune system? Front Physiol [Internet]. 2013 Mar 15 [cited 2019 Jul 28];4. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3597943/

94. Grünwald S, Pilhofer M, Höll W. Microbial associations in gut systems of wood- and bark-inhabiting longhorned beetles [Coleoptera: Cerambycidae]. Syst Appl Microbiol. 2010 Jan 1;33(1):25–34.

95. Šochová E, Husník F, Nováková E, Halajian A, Hypša V. Arsenophonus and Sodalis replacements shape evolution of symbiosis in louse flies. PeerJ. 2017 Dec 11;5:e4099.

96. Rubin BER, Sanders JG, Turner KM, Pierce NE, Kocher SD. Social behaviour in bees influences the abundance of *Sodalis* (Enterobacteriaceae) symbionts. R Soc Open Sci. 2018 Jul;5(7):180369.

97. Maltz MA, Weiss BL, O'Neill M, Wu Y, Aksoy S. OmpA-Mediated Biofilm Formation Is Essential for the Commensal Bacterium Sodalis glossinidius To Colonize the Tsetse Fly Gut. Appl Environ Microbiol. 2012 Nov 1;78(21):7760–8.

98. Chari A, Oakeson KF, Enomoto S, Jackson DG, Fisher MA, Dale C. Phenotypic characterization of Sodalis praecaptivus sp. nov., a close non-insect-associated member of the Sodalis-allied lineage of insect endosymbionts. Int J Syst Evol Microbiol. 2015 May;65(Pt 5):1400–5.

99. Oakeson KF, Gil R, Clayton AL, Dunn DM, von Niederhausern AC, Hamil C, et al. Genome Degeneration and Adaptation in a Nascent Stage of Symbiosis. Genome Biol Evol. 2014 Jan 8;6(1):76–93.

100. Boyd BM, Allen JM, Koga R, Fukatsu T, Sweet AD, Johnson KP, et al. Two Bacterial Genera, Sodalis and Rickettsia, Associated with the Seal Louse Proechinophthirus fluctus (Phthiraptera: Anoplura). Appl Env Microbiol. 2016 Jun 1;82(11):3185–97.

101. Verbarg S, Frühling A, Cousin S, Brambilla E, Gronow S, Lünsdorf H, et al. Biostraticola tofi gen. nov., spec. nov., A Novel Member of the Family Enterobacteriaceae. Curr Microbiol. 2008 Jun;56(6):603–8.

102. Clayton AL, Oakeson KF, Gutin M, Pontes A, Dunn DM, von Niederhausern AC, et al. A Novel Human-Infection-Derived Bacterium Provides Insights into the Evolutionary Origins of Mutualistic Insect–Bacterial Symbioses. Guttman DS, editor. PLoS Genet. 2012 Nov 15;8(11):e1002990.

103. Baumgardner DJ. Soil-Related Bacterial and Fungal Infections. J Am Board Fam Med. 2012 Sep 1;25(5):734–44.

104. Raaijmakers JM, Paulitz TC, Steinberg C, Alabouvette C, Moënne-Loccoz Y. The rhizosphere: a playground and battlefield for soilborne pathogens and beneficial microorganisms. Plant Soil. 2009 Aug 1;321(1):341–61.

105. Pérez-Pantoja D, Donoso R, Agulló L, Córdova M, Seeger M, Pieper DH, et al. Genomic analysis of the potential for aromatic compounds biodegradation in Burkholderiales. Environ Microbiol. 2012;14(5):1091–117.

106. Tian J-H, Pourcher A-M, Bouchez T, Gelhaye E, Peu P. Occurrence of lignin degradation genotypes and phenotypes among prokaryotes. Appl Microbiol Biotechnol. 2014 Dec;98(23):9527–44.

107. Zhao X, Liu D. Kinetic Modeling and Mechanisms of Acid-Catalyzed Delignification of Sugarcane Bagasse by Aqueous Acetic Acid. BioEnergy Res. 2013 Jun 1;6:436–47.

108. Horn M, Collingro A, Schmitz-Esser S, Beier CL, Purkhold U, Fartmann B, et al. Illuminating the evolutionary history of chlamydiae. Science. 2004 Apr 30;304(5671):728–30.

109. Buyer JS. A Soil and Rhizosphere Microorganism Isolation and Enumeration Medium That Inhibits Bacillus mycoides. Appl Environ Microbiol. 1995 May 1;61(5):1839–42.

110. Cho J-C, Giovannoni SJ. Cultivation and Growth Characteristics of a Diverse Group of Oligotrophic Marine Gammaproteobacteria. Appl Environ Microbiol. 2004 Jan 1;70(1):432–40.

111. Bandounas L, Wierckx NJ, de Winde JH, Ruijssenaars HJ. Isolation and characterization of novel bacterial strains exhibiting ligninolytic potential. BMC Biotechnol. 2011;11:94.

112. Konstantinidis KT, Tiedje JM. Towards a Genome-Based Taxonomy for Prokaryotes. J Bacteriol. 2005 Sep;187(18):6258–64.

113. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009 Jan 2;323(5910):133–8.

114. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013 Jun;10(6):563–9.

115. Frank AC, Lobry JR. Oriloc: prediction of replication boundaries in unannotated bacterial chromosomes. Bioinforma Oxf Engl. 2000 Jun;16(6):560–1.

116. Chen I-MA, Markowitz VM, Chu K, Palaniappan K, Szeto E, Pillay M, et al. IMG/M: integrated genome and metagenome comparative data analysis system. Nucleic Acids Res. 2017 Jan 4;45(Database issue):D507–16.

117. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol. 2018;68(1):461–6.

118. Qin Q-L, Xie B-B, Zhang X-Y, Chen X-L, Zhou B-C, Zhou J, et al. A Proposed Genus Boundary for the Prokaryotes Based on Genomic Insights. J Bacteriol. 2014 Jun 15;196(12):2210–5.

119. Harris HMB, Bourin MJB, Claesson MJ, O'Toole PW. Phylogenomics and comparative genomics of Lactobacillus salivarius, a mammalian gut commensal. Microb Genomics [Internet]. 2017 [cited 2019 Apr 11];3(8). Available from: https://mgen.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000115

120. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol. 2018 Jul 6;36:566–9.

121. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE. 2010 Mar 10;5(3):e9490.

122. Segata N, Börnigen D, Morgan XC, Huttenhower C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun. 2013;4:2304.

123. Soderlund C, Bomhoff M, Nelson WM. SyMAP v3.4: a turnkey synteny system with application to plant genomes. Nucleic Acids Res. 2011 May 1;39(10):e68–e68.

124. Rosas-Pérez T, León AV-P de, Rosenblueth M, Ramírez-Puebla ST, Rincón-Rosales R, Martínez-Romero J, et al. The Symbiome of Llaveia Cochineals (Hemiptera: Coccoidea: Monophlebidae) Includes a Gammaproteobacterial Cosymbiont Sodalis TME1 and the Known Candidatus Walczuchella monophlebidarum. Insect Physiol Ecol [Internet]. 2017 Apr 12 [cited 2019 Jul 23]; Available from: https://www.intechopen.com/books/insect-physiology-and-ecology/the-symbiome-of-llaveia-cochineals-hemiptera-coccoidea-monophlebidae-includes-a-gammaproteobacterial

125. Cases I, de Lorenzo V, Ouzounis CA. Transcription regulation and environmental adaptation in bacteria. Trends Microbiol. 2003 Jun 1;11(6):248–53.

126. Hall RJ, Flanagan LA, Bottery MJ, Springthorpe V, Thorpe S, Darby AC, et al. A Tale of Three Species: Adaptation of Sodalis glossinidius to Tsetse Biology, Wigglesworthia Metabolism, and Host Diet. mBio. 2019 Feb 26;10(1):e02106-18.

127. Kamimura N, Sakamoto S, Mitsuda N, Masai E, Kajita S. Advances in microbial lignin degradation and its applications. Curr Opin Biotechnol. 2019 Apr 1;56:179–86.

128. Sato Y, Moriuchi H, Hishiyama S, Otsuka Y, Oshima K, Kasai D, et al. Identification of Three Alcohol Dehydrogenase Genes Involved in the Stereospecific Catabolism of Arylglycerol-β-Aryl Ether by Sphingobium sp. Strain SYK-6. Appl Environ Microbiol. 2009 Aug 15;75(16):5195–201.

129. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, et al. The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). Stand Genomic Sci. 2015 Oct 26;10(1):86.

130. Dale C, Jones T, Pontes M. Degenerative Evolution and Functional Diversification of Type-III Secretion Systems in the Insect Endosymbiont Sodalis glossinidius. Mol Biol Evol. 2005 Mar 1;22(3):758–66.

131. Holladay JE, White JF, Bozell JJ, Johnson D. Top Value-Added Chemicals from Biomass - Volume II—Results of Screening for Potential Candidates from Biorefinery Lignin [Internet]. 2007 Oct [cited 2019 Jul 29] p. PNNL-16983, 921839. Report No.: PNNL-16983, 921839. Available from: http://www.osti.gov/servlets/purl/921839/

132. Brown ME, Chang MC. Exploring bacterial lignin degradation. Curr Opin Chem Biol. 2014 Apr 1;19(Supplement C):1–7.

133. B Fisher A, S Fong S, 1 Integrative Life Sciences Program, School of Life Sciences, Virginia Commonwealth University, Richmond, Virginia. Lignin biodegradation and industrial implications. AIMS Bioeng. 2014;1(2):92–112.

134. Zhang K, Si M, Liu D, Zhuo S, Liu M, Liu H, et al. A bionic system with Fenton reaction and bacteria as a model for bioprocessing lignocellulosic biomass. Biotechnol Biofuels. 2018 Dec;11(1):31.

135. Xu G, Goodell B. Mechanisms of wood degradation by brown-rot fungi: chelator-mediated cellulose degradation and binding of iron by cellulose. J Biotechnol. 2001 Apr 27;87(1):43–57.

136. Kirk TK, Farrell RL. Enzymatic "combustion": the microbial degradation of lignin. Annu Rev Microbiol. 1987;41:465–505.

137. DeAngelis KM, Sharma D, Varney R, Simmons B, Isern NG, Markilllie LM, et al. Evidence supporting dissimilatory and assimilatory lignin degradation in Enterobacter lignolyticus SCF1. Front Microbiol [Internet]. 2013 Sep 19 [cited 2016 Jan 5];4. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3777014/

138. Ertel JR, Hedges JI. The lignin component of humic substances: Distribution among soil and sedimentary humic, fulvic, and base-insoluble fractions. Geochim Cosmochim Acta. 1984 Oct 1;48(10):2065–74.

139. He S, Lau MP, Linz AM, Roden EE, McMahon KD. Extracellular electron transfer may be an overlooked contribution to pelagic respiration in humic-rich freshwater lakes. bioRxiv. 2018 Aug 15;392027.

140. Billings AF, Fortney JL, Hazen TC, Simmons B, Davenport KW, Goodwin L, et al. Genome sequence and description of the anaerobic lignin-degrading bacterium Tolumonas lignolytica sp. nov. Stand Genomic Sci. 2015 Nov 19;10(1):106.

141. Janssen PH, Schuhmann A, Mörschel E, Rainey FA. Novel anaerobic ultramicrobacteria belonging to the Verrucomicrobiales lineage of bacterial descent isolated by dilution culture from anoxic rice paddy soil. Appl Environ Microbiol. 1997 Apr;63(4):1382–8.

142. Widdel F, Rabus R. Anaerobic biodegradation of saturated and aromatic hydrocarbons. Curr Opin Biotechnol. 2001 Jun 1;12(3):259–76.

143. Tschech A, Pfennig N. Growth yield increase linked to caffeate reduction in Acetobacterium woodii. Arch Microbiol. 1984 Feb 1;137(2):163–7.

144. Kahm M, Hasenbrink G, Lichtenberg-Fraté H, Ludwig J, Kschischo M. grofit : Fitting Biological Growth Curves with R. J Stat Softw [Internet]. 2010 [cited 2019 Aug 2];33(7). Available from: http://www.jstatsoft.org/v33/i07/

145. Gregori J, Villarreal L, Sánchez A, Baselga J, Villanueva J. An effect size filter improves the reproducibility in spectral counting-based comparative proteomics. J Proteomics. 2013 Dec 16;95:55–65.

146. Jeitner TM. Optimized ferrozine-based assay for dissolved iron. Anal Biochem. 2014 Jun;454:36–7.

147. Williams KM, Martin WE, Smith J, Williams BS, Garner BL. Production of Protocatechuic Acid in Bacillus Thuringiensis ATCC33679. Int J Mol Sci. 2012 Mar 21;13(3):3765–72.

148. Weljie AM, Newton J, Mercier P, Carlson E, Slupsky CM. Targeted Profiling: Quantitative Analysis of 1H NMR Metabolomics Data. Anal Chem. 2006 Jul 1;78(13):4430–42.

149. Chen X, Hu Y, Yang B, Gong X, Zhang N, Niu L, et al. Structure of lpg0406, a carboxymuconolactone decarboxylase family protein possibly involved in antioxidative response from Legionella pneumophila. Protein Sci Publ Protein Soc. 2015 Dec;24(12):2070–5.

150. Koshkin A, Zhou X, Kraus CN, Brenner JM, Bandyopadhyay P, Kuntz ID, et al. Inhibition of Mycobacterium tuberculosis AhpD, an Element of the Peroxiredoxin Defense against Oxidative Stress. Antimicrob Agents Chemother. 2004 Jul 1;48(7):2424–30.

151. Chuang M-H, Wu M-S, Lo W-L, Lin J-T, Wong C-H, Chiou S-H. The antioxidant protein alkylhydroperoxide reductase of Helicobacter pylori switches from a peroxide reductase to a molecular chaperone function. Proc Natl Acad Sci. 2006 Feb 21;103(8):2552–7.

152. Solovyev V, Salamov AA. Automatic Annotation of Microbial Genomes and Metagenomic Sequences 3 MATERIAL AND METHODS Learning Parameters and Prediction of Protein-Coding Genes. In 2013.

153. Pereira JH, Heins RA, Gall DL, McAndrew RP, Deng K, Holland KC, et al. Structural and Biochemical Characterization of the Early and Late Enzymes in the Lignin β-Aryl Ether Cleavage Pathway from Sphingobium sp. SYK-6. J Biol Chem. 2016 May 6;291(19):10228–38.

154. Oppermann U, Filling C, Hult M, Shafqat N, Wu X, Lindh M, et al. Short-chain dehydrogenases/reductases (SDR): the 2002 update. Chem Biol Interact. 2003 Feb;143–144:247–53.

155. Cunane LM, Chen Z-W, Shamala N, Mathews FS, Cronin CN, McIntire WS. Structures of the flavocytochrome p-cresol methylhydroxylase and its enzyme-substrate complex: gated substrate entry and proton relays support the proposed catalytic mechanism11Edited by D. C. Rees. J Mol Biol. 2000 Jan 14;295(2):357–74.

156. Nam HS, Spencer M, Anderson AJ, Cho BH, Kim YC. Transcriptional regulation and mutational analysis of a dctA gene encoding an organic acid transporter protein from Pseudomonas chlororaphis O6. Gene. 2003 Dec 24;323:125–31.

157. Antunes A, Golfieri G, Ferlicca F, Giuliani MM, Scarlato V, Delany I. HexR Controls Glucose-Responsive Genes and Central Carbon Metabolism in Neisseria meningitidis. Henkin TM, editor. J Bacteriol. 2016 Feb 15;198(4):644–54.

158. Leyn SA, Li X, Zheng Q, Novichkov PS, Reed S, Romine MF, et al. Control of Proteobacterial Central Carbon Metabolism by the HexR Transcriptional Regulator A CASE STUDY IN SHEWANELLA ONEIDENSIS. J Biol Chem. 2011 Oct 14;286(41):35782–94.

159. Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R. Glycolytic strategy as a tradeoff between energy yield and protein cost. Proc Natl Acad Sci. 2013 Jun 11;110(24):10039–44.

160. Guillon E, Merdy P, Aplincourt M, Dumonceau J, Vezin H. Structural Characterization and Iron(III) Binding Ability of Dimeric and Polymeric Lignin Models. J Colloid Interface Sci. 2001 Jul 1;239(1):39–48.

161. Fioroto AM, Nascimento AN, Oliveira PV. In Vitro Evaluation of Cu, Fe, and Zn Bioaccessibility in the Presence of Babassu Mesocarp. J Agric Food Chem. 2015 Jul 22;63(28):6331–7.

162. Rolfe MD, Rice CJ, Lucchini S, Pin C, Thompson A, Cameron ADS, et al. Lag Phase Is a Distinct Growth Phase That Prepares Bacteria for Exponential Growth and Involves Transient Metal Accumulation. J Bacteriol. 2012 Feb;194(3):686–701.

163. Ahmed E, Holmström SJM. Siderophores in environmental research: roles and applications. Microb Biotechnol. 2014 May 1;7(3):196–208.

164. Montazeri M, Eckelman MJ. Life Cycle Assessment of Catechols from Lignin Depolymerization. ACS Sustain Chem Eng. 2016 Mar 7;4(3):708–18.

165. García CA, Passerini De Rossi B, Alcaraz E, Vay C, Franco M. Siderophores of Stenotrophomonas maltophilia: detection and determination of their chemical nature. Rev Argent Microbiol. 2012 Sep;44(3):150–4.

166. Dave B, Dube H. Chemical characterization of fungal siderophores. 2000;

167. Janusz G, Pawlik A, Sulej J, Świderska-Burek U, Jarosz-Wilkołazka A, Paszczyński A. Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. FEMS Microbiol Rev. 2017 Nov;41(6):941–62.

168. Sipe AR, Wilbur AE, Cary SC. Bacterial Symbiont Transmission in the Wood-Boring Shipworm Bankia setacea (Bivalvia: Teredinidae). Appl Environ Microbiol. 2000 Apr 1;66(4):1685–91.

169. Hatakka A, Hammel KE. Fungal Biodegradation of Lignocelluloses. In: Hofrichter M, editor. Industrial Applications [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011 [cited 2019 Aug 1]. p. 319–40. Available from: http://link.springer.com/10.1007/978-3-642-11458-8_15

170. D'Auria M, Emanuele L, Racioppi R. FT–ICR–MS analysis of lignin. Nat Prod Res. 2012 Aug 1;26(15):1368–74.

171. Wen J-L, Sun S-L, Xue B-L, Sun R-C. Recent Advances in Characterization of Lignin Polymer by Solution-State Nuclear Magnetic Resonance (NMR) Methodology. Materials. 2013 Jan 23;6(1):359–91.

172. Zeikus JG. Lignin Metabolism and the Carbon Cycle. In: Alexander M, editor. Advances in Microbial Ecology [Internet]. Boston, MA: Springer US; 1981 [cited 2019 Jul 7]. p. 211–43. (Advances in Microbial Ecology). Available from: https://doi.org/10.1007/978-1-4615-8306-6_5

173. Borneman WS, Akin DE, VanEseltine WP. Effect of phenolic monomers on ruminal bacteria. Appl Environ Microbiol. 1986 Dec 1;52(6):1331–9.

174. Nichols NN, Harwood CS. PcaK, a high-affinity permease for the aromatic compounds 4-hydroxybenzoate and protocatechuate from Pseudomonas putida. J Bacteriol. 1997;179(16):5056–5061.

175. Michalska K, Chang C, Mack JC, Zerbs S, Joachimiak A, Collart FR. Characterization of Transport Proteins for Aromatic Compounds Derived from Lignin: Benzoate Derivative Binding Proteins. J Mol Biol [Internet]. 2012 Nov 2 [cited 2016 Apr 21];423(4). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3836681/