December 2020

Development of *Cutaneotrichosporon oleaginosus* to Convert Lignin-Derived Phenolics to Oleochemicals

Allison Yaguchi
Clemson University, yaguchial@icloud.com

Development of *Cutaneotrichosporon oleaginosus* to convert lignin-derived phenolics to oleochemicals.

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Chemical & Biomolecular Engineering

---

by
Allison L. Yaguchi
December 2020

Accepted by:
Dr. Mark A. Blenner, Committee Chair
Dr. Marc R. Birtwistle
Dr. Christopher A. Saski
Dr. Eric M. Davis

Abstract

Oleaginous yeasts have long been a target for developing industrial-scale biorefineries due to their ability to accumulate high amounts of lipids, synthesize complex chemicals and proteins, and robustly metabolize diverse feedstocks. In parallel, interest in lignocellulosic biomass as a feedstock has grown. While most processes focus on the carbohydrates from cellulose and hemicellulose, the most energy-dense biopolymer, lignin, remains underutilized.

This dissertation describes foundational work describing lignin conversion by *Cutaneotrichosporon oleaginosus*, a non-model oleaginous yeast known for its metabolism of alternative sugars, including xylose, and tolerance and metabolism toxic lignocellulosic hydrolysate inhibitors such as 5-HMF, furfural, acetic acid. This dissertation is the first to describe robust lipid production by this yeast while utilizing five aromatic substrates as the sole carbon source: phenol, resorcinol, *p*-hydroxybenzoic acid, *p*-coumaric acid, and ferulic acid. This yeast can also tolerate an alkaline pretreated lignin hydrolysate and remain oleaginous.

The genetic basis of yeast aromatic metabolism is poorly characterized, so a multi-omic approach was applied to improve the existing genome annotation and identify novel gene functions relevant to aromatic catabolism. Genes unique to and common across all six substrates mentioned build a roadmap for future engineering for robust lignin valorization. To this, a small, functional genetic toolkit was de-

veloped to improve the genetic accessibility of this non-model yeast. Together, this dissertation demonstrates that *C. oleaginosus* is poised to become a preferred host for lignocellulosic biomass to oleochemical conversion.

Dedication

This dissertation is dedicated to my loving family.

To my parents, Diane and Michael, *for giving me the space, opportunities, and encouragement to pursue my passions,*

and to my sister, Megan, *whose strength and resilience I will always admire.*

Acknowledgments

I have been incredibly fortunate to have many family, friends, mentors, and colleagues who have supported me over the years. I am deeply thankful and will keep their many lessons close to my heart. I specifically would like to acknowledge the following, in no particular order:

Dr. Mark Blenner, who has been my mentor and advisor for ~~7~~ 6.5 years, supported me throughout the entire Ph.D. process, guided me through many failed experiments, and provided invaluable career advice. Thank you for nurturing my independence to explore my research, preparing me for a future in academia, and being patient with me.

My dissertation committee: Drs. Birtwistle, Davis, and Saski, for their expertise, guidance, and words of encouragement.

My lab mates: Murri, Michael, Weigao, Max, Molly, Dyllan, Vijay, Stephen, Ayushi, and Josh. Thank you all for the countless research discussions, proof reading, coffee "collaborations", and, above all, your friendship. You have all pushed me to be better and encouraged me through discouraging times. I truly value our time together and consider you all my lab family. I also would like to thank all of the undergraduates and post-docs who have been a part of the group along the way. Though it has been nearly a decade since I started working with them, I acknowledge the Fong group for catalyzing my love of biomolecular engineering. Drs. Niti Vanee,

Adam Fisher, George McArthur, and Stephen Fong: you are where it all started!

The ChBE faculty, who have had a role in my personal and professional development. Thank you for taking valuable time out of your days to answer my questions about graduate school, post-PhD experiences, and the transition to academia and encourage me to pursue my passion for community outreach and education. Also, the ChBE staff, Joy Rodatz, Terri McAllister, Diana Stamey, and Roshell Hicks, for making Earle Hall a welcoming and happy place, and our technicians for keeping our building up and running. You all have a difficult job, yet always found ways to make mine easier.

My collaborators: Dr. Gregg Beckham and his group at NREL for providing lignin hydrolysates and for hosting me for a three month fellowship, Dr. Adam Guss and his team for providing transcriptomic support, Dr. Kyria Boundy-Mills for access to the Phaff collection, and Dr. Chris Saski for the opportunity to tangentially work on several different projects.

The funding agencies that have supported this body of work: The United States Department of Agriculture Sun Grant (2014-38502-22598), the Department of Education GAANN Fellowship (P200A180076), the Department of Energy SCGSR award, and Clemson University. I also acknowledge BioRender.com for the ability to generate high quality figures.

To all of my friends I have met here. Graduate school is a difficult journey and one I could not have made alone. Walking a path alongside you, no matter how brief, has been an honor and a pleasure. I am grateful for each of you, for all the joy and love you have brought into my life. I will always cherish you, my Clemson family.

Finally, my deep and sincere gratitude to my family for their endless love, laughter, and support. I am forever indebted to you for giving me the opportunities and experiences ("in my lifetime") that have made me who I am.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Background

1.1 Lignocellulosic biomass: Valorizing lignin

Lignocellulosic biomass, or dry plant matter, is the most abundant raw material on Earth and a valuable resource for the production of biofuels and chemicals. It is comprised of three biopolymers: cellulose, hemicellulose, and lignin (Figure 1.1). The ratio of these biopolymers changes between hardwoods, softwoods, and grasses (Table 1.1) [1]. Lignin provides strength and hydrophobicity to cell walls and protects the polysaccharides from microbial degradation [2].

Over the past several decades, valorization of lignocellulosic biomass has largely focused on biomass-derived sugars. Despite being the most energy dense component in biomass and the only renewable source of aromatic compounds, lignin is often neglected and instead burned for its process heat value. While lignin has specialty applications, including incorporation into carbon fibers, polyurethane foams, composites, and phenol-formaldehyde resins [6–9], only 2% is recovered for non-fuel pur-

Portions of this chapter are based, in part, on the following reviews authored by Yaguchi, et al.: [3–5].

Table 1.1. Lignocellulosic biomass composition across different sources [1].

Source		**Cellulose (%)**	**Hemicellulose (%)**	**Lignin (%)**
Hardwood	Poplar	50.8-53.3	26.2-28.7	15.5-16.3
	Oak	40.4	35.9	24.1
	Eucalyptus	54.1	18.4	21.5
Softwood	Pine	42.0-50.0	24.0-27.0	20.0
	Douglas fir	44.0	11.0	27.0
	Spruce	45.5	22.9	27.9
Agricultural waste	Wheat Straw	35.0-39.0	23.0-30.0	12.0-16.0
	Barley hull	34.0	36.0	13.8-19.0
	Barley straw	36.0-43.0	24.0-33.0	6.3-9.8
	Rice straw	29.2-34.7	23.0-25.9	17.0-19.0
	Rice husks	28.7-35.6	12.0-29.3	15.4-20.0
	Oat straw	31.0-35.0	20.0-26.0	10.0-15.0
	Ray straw	36.2-47.0	19.0-24.5	9.9-24.0
	Corn cobs	33.7-41.2	31.9-36.0	6.1-15.9
	Corn stalks	35.0-39.6	16.8-35.0	7.0-18.4
	Sugarcane bagasse	25.0-45.0	28.0-32.0	15.0-21.0
	Sorghum straw	32.0-35.0	24.0-27.0	15.0-21.0
Grasses	Grasses	25.0-40.0	25.0-50.0	10.0-30.0
	Switchgrass	35.0-40.0	25.0-30.0	15.0-20.0

poses [10–12].

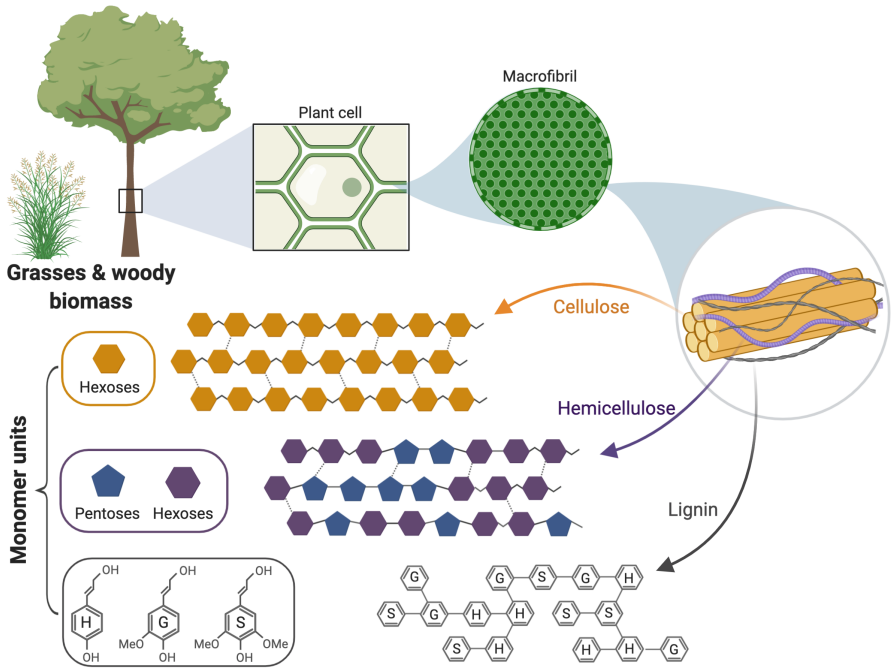


Figure 1.1. Lignocellulosic biomasss composition.

There are multiple sources of lignin, but the two primary sources are agricultural waste and Kraft lignin from paper mills [11–13]. Kraft pulping in paper mills alone produces approximately 130 million tons of lignin per year [12]. A single second-generation biofuel plant is estimated to produce 70,000 tons of lignin per year based on a 2,000 ton per day plant using corn stover as feedstock. A single plant would require 724,000 dry U.S. ton corn stover feedstock per year for the production of 31.3 million gallons of biofuel (the equivalent of 32.9 million gallons of gasoline) annually [14]. According to the U.S. Department of Agriculture and Renewable Fuels Association, there are 209 cellulosic biorefineries in the United States that produced 15,776 million gallons in 2019 [15, 16]. If every plant in the U.S. switched to biomass based feedstock to create the same annual output, approximately 35 million tons of

lignin would be produced. With the increasing use of lignocellulosic biomass, added to existing agricultural waste and Kraft lignin streams, lignin will be generated at rates that will far surpass its heating demand. Additionally, the manufacturing, operating, and capital costs associated with a cellulosic biorefinery requires a minimum selling price (MSP) of $5.35 per gallon [14]. With decreasing price of petroleum fuel, this MSP is not competitive in the current market. Thus, cost-reductions are imperative to realize a sustainable and profitable lignocellulosic bioeconomy, and lignin valorization has been identified as a linchpin towards this goal [14, 17–21]. Lignins' inherent heterogeneity and recalcitrance make it challenging to valorize; however, a unique opportunity is presented with microbes that natively metabolize lignin and lignin-related compounds. Focusing on biological systems allows the same valorization and applications found for cellulosic and hemicellulosic biomass fractions to be applied to lignin (Fig. 1.2).

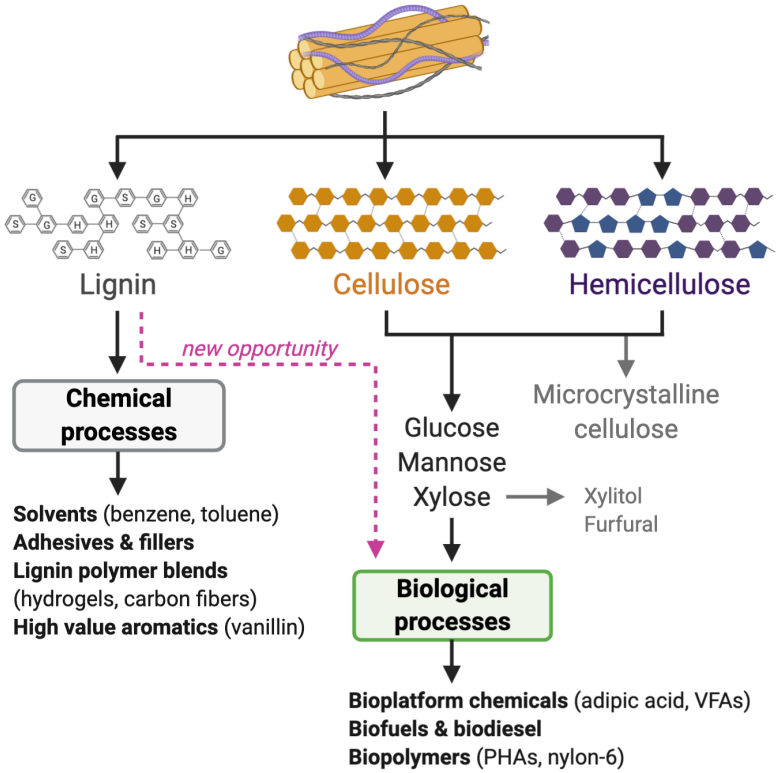


Figure 1.2. Valorization of lignin is essential for the lignocellulosic bioeconomy.

1.2 Lignin structure

Lignin is a complex, aromatic biopolymer comprised of three monomer units — the syringyl (S), guaiacyl (G), and *p*-hydroxyphenyl (H) subunits (Figure 1.3). It constitutes anywhere from 15-30% of biomass and the ratios of S, G, and H monomers varies between plant species and even among components (leaves versus bark) of the same plant [22]. Softwoods tend to have almost exclusively G units (G-lignin), hardwoods tend to have G and S units (G-S-lignin), and grasses tend to have all three monomer units (G-S-H-lignin) [23, 24].

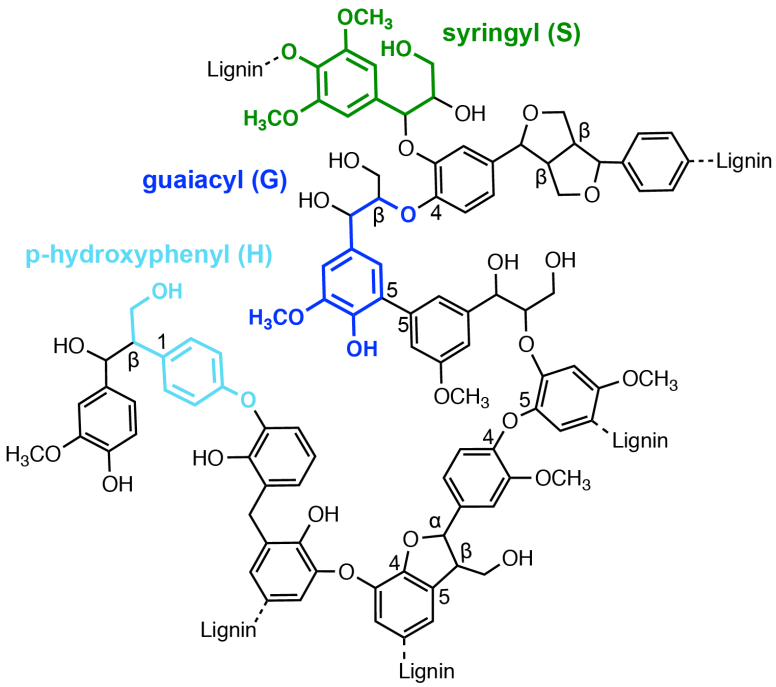


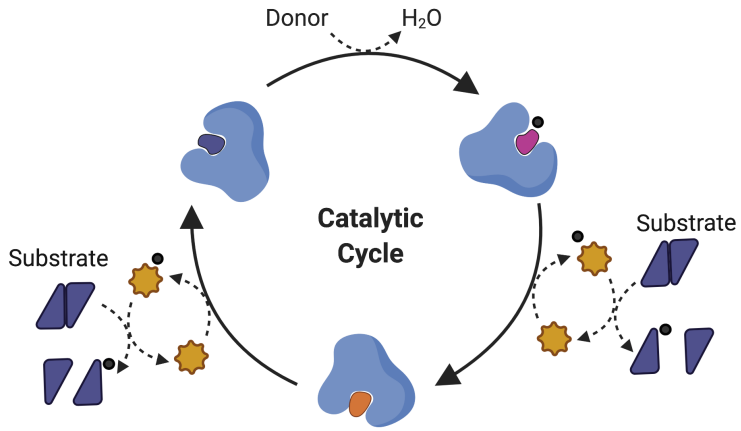
Figure 1.3. A model lignin polymer with the S, G, and H monomers and key bonds highlighted.

The S, G, and H monomers are linked together through a multitude of C-C (5-5', β-β', β-1) and C-O (β-O-4, α-O-4, 4-O-5) bonds (Figure 1.3). The linkages present in each lignin type is dependent largely on the type of biomass; however, nearly two-thirds of all bonds found in lignin are ether linkages. The predominant linkage is the β-O-4 (β-aryl ether) linkage, since the phenolic and β-hydrogens shared between the monomers participate in most of the bonding [25–27]. There are both biological and chemical means to depolymerize lignin, but many of them are designed to target the β-O-4 bonds, since these are most prevalent. The sections below discuss the current understanding of biological systems target bonds connecting the lignin polymer and process the resulting mixture of aromatic monomers.

1.3 Biotransformation of lignin

1.3.1 Lignin depolymerization

Fungal species known to be robust lignin depolymerizers include *Phanerochaete chrysosporium*, *Trametes versicolor*, *Pleurotus eryngii*, *Irpex lacteus*, and *Ceriporiopsis subvermispora* [28–31]. Of the many lignin degraders in nature, white-rot fungi are among the most efficient. They use numerous laccases and peroxidases to break down the different linkages connecting the S, G, and H lignin monomers. Fungi have a consortium of assisting enzymes, such as redox enzymes, that generate hydrogen peroxide and regenerate necessary cofactors and mediators. These auxiliary enzymes are often found in fungal secretomes and include glyoxal oxidases, aryl alcohol oxidases, pyranose oxidase, cellobiose dehydrogenase, benzoquinone reductases, and glucose oxidases [32, 33]. Other enzymes implicated in lignin depolymerization are glutathione S-transferases, etherases, thiolases, and cytochrome P450s [22, 34–37]. Bacterial species are less described for the lignin depolymerization, and delignification by bacteria is typically slower and less extensive than that of fungal systems. Notable strains that have been described for this unique trait include *Pseudomonas putida* mt-2, *Streptomyces viridosporus* T7A, *Thermobifida fusca*, *Rhodococcus opacus* PD630, and *Rhodococcus jostii* RHA1 [31, 38–44].



					Donor
DyP	Fe^{III}	$Fe^{IV\bullet+}$	Fe^{IV}	Veratryl alcohol, (non-)phenolics	H_2O_2
MnP	Fe^{III}	$Fe^{IV\bullet+}$	Fe^{IV}	Mn^{2+}/Mn^{3+}	H_2O_2
VP	Fe^{III}	$Fe^{IV\bullet+}$	Fe^{IV}	Mn^{2+}/Mn^{3+}, Veratryl alcohol, (non-)phenolics	H_2O_2
Laccase	$4(Cu^{2+})$	$4(Cu^{+})$	$2(Cu^{2+})2(Cu^{+})$	Phenolics, ABTS	O_2

Figure 1.4. Catalytic cycle for the main ligninases. Figure based off [45].

The major enzymes that depolymerize lignin are laccases, manganese-dependent peroxidases (MnP), lignin peroxidases (LiP), dye-decoloring peroxidases (DyP), and versatile peroxidases (VP). They have very similar catalytic cycles that involve oxidation and reduction of substrates and mediators (Fig. 1.4). The first ligninolytic enzyme ever isolated was a LiP from *P. chrysosporium* [46]. This enzyme was discovered to be well-suited to oxiding sites of high redox potential, such as β-O-4 linkages common in most lignin [46, 47]. This allows LiP to catalize oxidation of non-phenolic aromatic compounds to aryl-cation radicals, even in the absence of H_2O_2. MnP is the most ubiquitous lignin-modifying peroxidase in fungi. It has a lower redox potential than LiP, hence its dependence of a manganese cofactor to act on lignin. MnP is an enzyme that oxidizes Mn^{2+} to Mn^{3+}, which then catalyzes the oxidation of phenolic

compounds. VP, as its name suggests, has a versatile substrate preference due to its high redox potential and multiple sites for access to the catalytic heme group. They use substrates of both LiP and MnP enzymes through a similar mechanism. DyPs are present in both fungal and bacterial systems. These enzymes are named for their ability to oxidize anthraquinones as well as other peroxidase substrates. A notable problem with peroxidases is their tendency to simultaneously repolymerize the released aromatic monomers during depolymerization. Laccases are multi-copper oxidases found in both fungi and bacteria, though curiously, it is not produced by *P. chrysosporium*. They have low redox potential, with fungal laccases having higher redox potential than bacterial laccases. They oxidize phenolic compounds and ABTS as mediators, rather than H_2O_2 like peroxidases. Laccases do not suffer from repolymerization products as frequently as peroxidases, making these an attractive target for heterologous lignin biotransformation [36, 45, 48, 49].

1.3.2 Biological funneling

After depolymerization, released monomers are converted into key intermediates using funneling pathways (Fig. 1.5). These pathways are very well described for bacterial and some fungal species, but annotation lags significantly behind for yeast [50]. Funneling enzymes condense the released monomers into three main intermediates: gallic acid, protocatechuic acid (PCA), and catechol. Having a limited number of intermediates decreases the number of enzymes required to cleave the phenolic rings open.

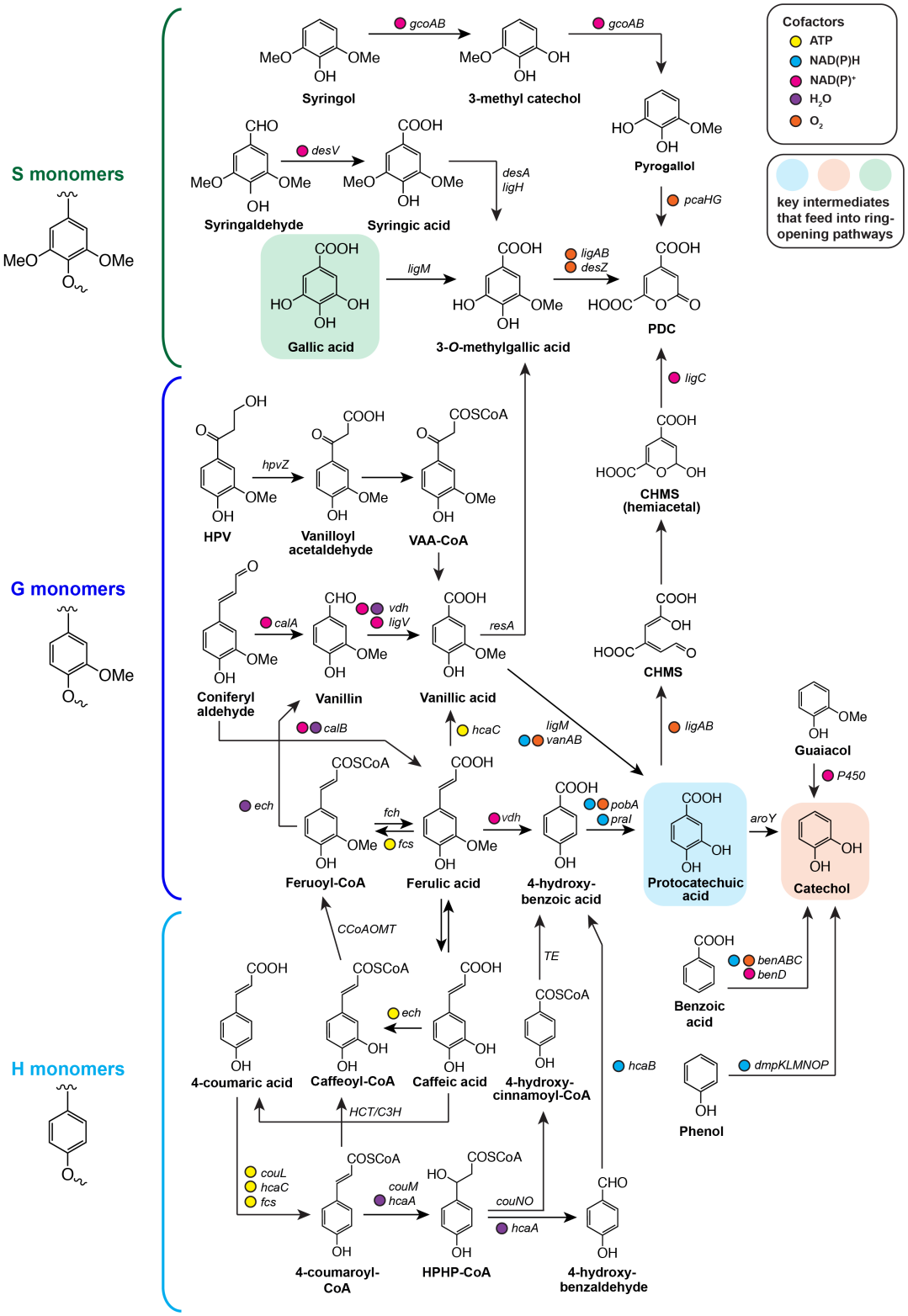


Figure 1.5. Funneling of S, G, and H monomers into key intermediates.

H monomers are the least substituted of the lignin monomeric units and are characterized by a hydroxyl group at the 3' carbon on the aromatic ring. A representative model compound is *p*-coumaric acid, as it is often released after lignin hydrolysis. H monomers are often funneled through PCA. In *Actinetobacter baylyi* ADP1, *p*-coumaric acid is converted through 4-hydroxybenzaldehyde to *p*-hydroxybenzoic acid (pHBA) via the *hca* operon. This is then converted to PCA by *pobA* [51]. By contrast, *Rhodococcus jostii* RH1 uses the *cou* operon to convert *p*-coumaric acid through 4-hydroxycinnamoyl-CoA. This is then converted to pHBA, which is then converted to PCA by *pobA* [52]. The presence of these genes were predicted and functionally conserved in *A. niger*. It is the only report of benzoic acid metabolism in filamentous fungi, and it was noted how little sequence similarity these genes shared with bacterial species [53]. PCA can be converted to catechol through a protocatechuate decarboxylase, *aroY*. This gene has been identified in many organisms including *Enterobacter cloacae*, *Klebsiella pneumoniae* sp. *pneumoniae*, *Cutaneotrichosporon cutaneum*, *Aspergillus niger* [54–57] and is a notorious bottleneck [58–61]. Phenol and guaiacol are converted directly to catechol. Guaiacol was converted to catechol using a cytochrome P450 monooxygenase, *gcoAB*, in *Amycolatopsis* sp. ATCC 39116 [62], a putative cytochrome P450 monooxygenase in *Rhodococcus rhodochrous* J3 [63], and *agcAB* in *Rhodococcus rhodocorus* EP4 [64]. Phenol is converted to catechol by a phenol hydroxylase.

G monomers have the same structure as H monomers with the addition of a single methoxy group, and the common model compounds are ferulic acid and vanillin. The funneling for G monomers is relatively similar to H monomers in that they funnel through PCA, but rather than pHBA directly upstream of PCA, G monomers tend to be converted to vanillic acid. Ferulic acid and vanillin are converted to vanillic acid by *fca* and *vdh* genes, respectively, in *Pseudomonas* species [65–67]. Vanillic acid is

converted to PCA by a vanillate demethylase. This reaction is catalyzed by *vanAB* in *Acinetobacter baylyi* ADP1 [68], *Pseudomonas* species [69, 70], and *Streptomyces* species [71] and *ligM* in *Sphingomonas* sp. SYK-6 [72].

The S monomers have methoxy groups at both the 3' and 5' positions on the aromatic ring. One or both of these methoxy groups are converted to hydroxy groups by demethylases, and ultimately funneled through gallic acid. Syringaldehyde is converted to syringic acid by *desV*, an aldehyde dehydrogenase, and then to 3-O-methylgallic acid (3-OMG) by *desA*, a syringate O-demethylase, from *Sphingomonas* sp. strain SYK-6. 3-OMG is then converted to gallic acid by *ligM*, a 3-O-methylgallate O-demethylase [72]. Often, the demethylation step is rate-limiting, causing a bottleneck immediately in funneling of S and G monomers, the primary constituents of lignin [73].

1.3.3 Aromatic ring cleavage

After monomer aromatic compounds are funneled into the key intermediates, protocatechuic acid, gallic acid, and catechol, they are then cleaved by dioxygenases at extra-diol (*meta*) or intradiol (*ortho*) positions (Fig. 1.6). Protocatechuic acid has three cleavage pathways: 4,5 *meta*-cleavage, 2,3 *meta*-cleavage, and an *ortho* cleavage pathway. Catechol has two cleavage routes - a single *meta*- and a single *ortho*-pathway. Gallic acid is cleaved in *ortho*- or *meta*-positions as well.

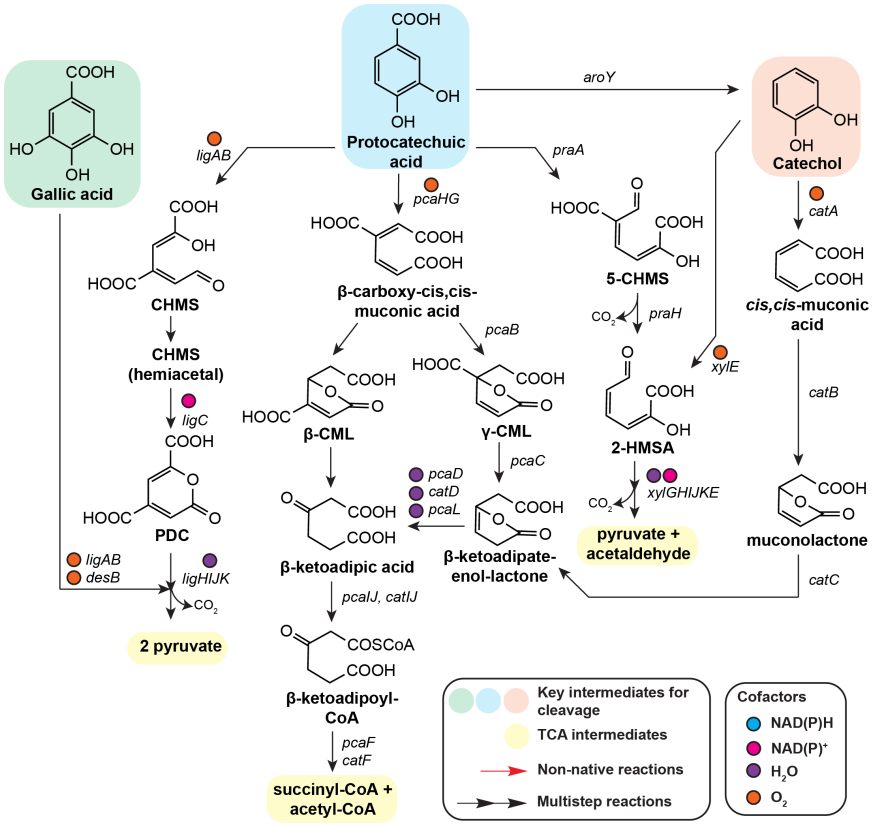


Figure 1.6. Cleavage of key intermediates into TCA intermediates.

Meta-cleavage pathways produce pyruvate, while the *ortho*-cleavage products result in succinyl-CoA and acetyl-CoA. Additionally, *meta*-cleavage pathways produce more CO_2 and generate NADH (Figure 1.7). From an engineering standpoint, the *meta*-pathways are ideal for pyruvate-based products where either total yield may not be as important or cofactor generation is important. *Ortho*-pathways are ideal for acetyl-CoA and muconate-derived products or for products that do not require a significant pool of available NADH.

The biochemistry of ring-cleavage pathways is typically well conserved across species, though there are gaps in the genome annotations of many eukaryotes as was seen with the funneling pathways [50]. Protocatechuic acid, gallic acid, and catechol are cleaved by dioxygenases. The *ortho*-cleavage pathways for protocate-

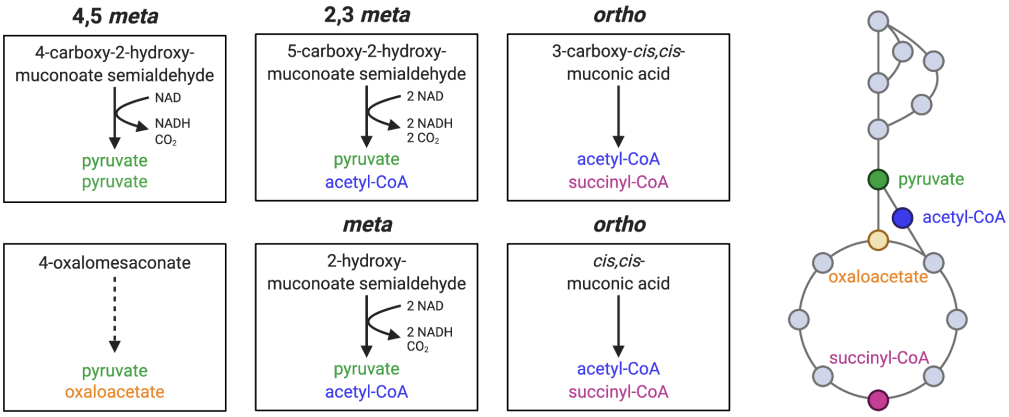


Figure 1.7. Summary of ring cleavage cofactors, by-products, and central metabolites produced.

chuic acid and catechol are initiated by protocatechuate 3,4-dioxygenase and catechol 1,2-dioxygenase. Both compounds are converted to a muconic acid, which are then eventually converted into β-ketoadipate. For this reason, the *ortho* pathways are also called the β-ketoadipate pathway. Looking closer at differences between eukaryotes and prokaryotes, there is a diverging point at β-carboxy-*cis,cis*-muconic acid. In bacteria, this cleavage product is converted to γ-carboxymuconolactone (γ-CLM). In eukaryotes, this product is converted to β-carboxymuconolactone (β-CML) [50, 74, 75]. Curiously, an alternative pathway altogether exists in *Rhodococcus jostii* and *Trichosporon cutaneum* WY2-2 such that protocatechuic acid is converted hydroxyquinol, which is then ring opened with a hydroxyquinol 1,2-dioxygenase [76, 77]. In addition, a decarboxylase was identified in *Klebsiella pneumoniae* that converts protocatechuic acid to catechol [59]. Intradiol dioxygenases were identified in fungal species, *Aspergillus niger* [78] and *A. nidulans* [79], but fungal annotation past this is scarce.

Meta-cleavage of protocatechuic acid can proceed by protocatechuate 2,3-

dioxygenase or by protocatechuate 4,5-dioxygenase. Cleavage by the former results in 5-carboxy-2-hydroxymuconoate-semialdhyde (5-CHMS). The gene responsible for this reaction in *Paenibacillus* sp. strain JJ-1b is *praA* [80]. The subsequent reactions take place through the *pra* operon [80, 81]. Cleavage at the 4,5-position results in 4-carboxy-2-hydroxymuconate-6-semialdehyde (CHMS) followed by converstion to 2-pyrone-4,6-dicarboxylate (PDC). The *lig* operon in *Sphingobium* sp. SYK-6 [82] and *pmd* operon in *Comamonas testosteroni* BR6020 [83] are well described for these reactions.

Extradiol cleavage of gallic acid is catalyzed by gallate dioxygenase and results in the keto form of 4-oxalomesaconate, proceeds through the protocatechuate 4,5-cleavage pathway at PDC. The bacterial genes *desB* and *galA* have been identified [84–86], but no homologous gene has been identified for yeast or fungi. Catechol is cleaved at the *meta* position by a catechol 2,3-dioxygenase. The resulting compound, 2-hydroxymuconic semialdehyde (2-HMSA), proceeds through protocatechuate 2,3-cleavage pathway.

1.3.4 Synthetic biology for enhanced lignin biotransformation

Most synthetic biology approaches to enhance lignin biotransformation include adding additional branches to funneling pathways to encompass a wider number of aromatic substrates and increasing flux through catabolism. Several studies describe a build-up of protocatechuate [59–61]. In *E. coli*, overexpression of AroY and the B-subunit of the pHBA decarboxylase, KpdB, from *Klebsiella pneumonia* eliminated this bottleneck measured during vanillin metabolism [59]. A similar approach was used in *P. putida* KT2440 using EcdBD, a KpdBD homolog from *Enterobacter cloa-*

cae [58]. In *A. baylyi* ADP1, removal of the BenM and CatM regulation of the *ben* operon allowed co-utilization of benzene and ferulic acid [87]. Guaiacol metabolism was added to *P. putida* KT2440 by fusing *catA*, a catechol 1,2-dioxygenase, to *gcoA* from *Amycolatopsis* sp. ATCC 39116 [88]. Addition of a cytochrome P450 monooxygenase and ferredoxin from *R. rhodochorus* J3 coupled to a ferrodoxin reductase from *Amycolatopsis* sp. ATCC 29116 facilitated guaiacol metabolism in *P. putida* EM42 [63]. GcoAB was also rationally engineered and subsequently expressed in *P. putida* KT2440 to allow syringol conversion without compromising guaiacol binding [73]. Similarly, *ligAB* from *Sphingobium* sp. SYK-6 was mutated to allow binding of 3-OMG for cleavage to PDC [89]. These studies emphasize how protein engineering can exand the range of substrates utilized without necessitating extra enzymes or pathways. Adaptive evolution is another method to improve aromatic tolerance and catabolism when rational engineering is not feasible due to lack of *a priori* knowledge. The *gcoA-catA* fusion was discovered after adaptive evolution of *A. baylyi* ADP1 [88]. An *E. coli* strain engineered with *pobA* from *Paenibacillus* sp. JJ-1B and either the *hca* operon from *A. baylyi* ADP1 or the *cou* pathway from *R. jostii* adaptively evolved in *p*-coumaric acid resulted in a up-regulation of *pobA* [90, 91]. An engineered *S. cerevisiae* strain evolved in a corn stover hydrolysate resulted in discovery that a knockout of a transcription factor, Yrr1p, is sufficient to upregulate ABC transporters, rRNA processing proteins, and ribosome biogenesis to increase tolerance to and metabolism of vanillin [92, 93]. Evolution studies of *R. opacus* PD630 identified a plethora of single nucleotide polymorphisms in redox, transport, and stress response genes that would otherwise not be. These works present a strong case for reverse engineering elusive and complex phenotypes through adaptive evolution, as well as enhancing our current understanding of the mechanisms utilized by microorganisms for tolerance to and metabolism of aromatic compounds.

1.4 Value-added products from lignin feedstocks

Cis,cis-muconic acid (referred to as muconic acid) is a precursor for bio-based nylon, and is a commonly targeted product from the *ortho*-cleavage of monoaromatic substrates. To facilitate muconic acid accumulation, downstream β-ketoadipate genes, such as *catBC* and *pcaHG* are typically knocked out from hosts such as *E. coli*, *P. putida*, and *Sphingobium* sp. SYK-6 [58–61, 94–97]. To date, the highest reported titer is by engineered *Corynebacterium glutamicum* MA-2 which produced 85 g/L of muconic acid from co-feeding catechol and glucose; however, there was a large drop in titer to 1.8 g/L when softwood lignin hydrolysate was used as feedstock [98]. Co-feeding glucose can facilitate high molar conversions of aromatic compounds to products since glucose can provide building blocks and energy carriers, such as acetyl-CoA and NADH, but aromatic pathways commonly suffer catabolite repression [60, 61, 97]. Deletion of a catabolite repression regulator, Crc, facilitated co-feeding with glucose and acetate [60, 97]. Engineered *P. putida* cells deficient in the Crc regulator increased molar yield of muconic acid derived from *p*-coumaric acid from 50% to 96.4% in the presence of glucose [97].

It is not uncommon to convert intradiol pathways to extradiol pathways to facilitate accumulation of *meta*-cleavage pathway intermediates, such as PDC, or to produce pyruvate-derived products [99–104]. Amongst these studies, the best strain produced 58 g/L of PDC and had a molar yield of 80.7% from pHBA in *P. putida* KT2440 by knocking out *pcaHG* and introducing *ligAB* from *Sphingobium* sp. SYK-6 [103]. This engineered *P. putida* strain outperformed PDC production from *Novosphigobium aromaticivorans*, a species that natively uses the *meta* pathway [105].

Not all value-added products stem directly from the ring-cleavage pathways. Some products can capitalize on the acetyl-CoA pools generated by the β-ketoadipate

pathway. *Acinetobacter baylyi* ADP1 produced 72 μg/L of 1-undecene using 110 mM of ferulic acid as the sole carbon source when coexpressing '*tesA* from *E. coli* and *undA* from *P. putida* [106]. *Pseudomonas putida* KT2440 fed 100 mM pHBA with constant glucose in a two-stage bioreactor produced 17 mM adipic acid with a 17.4% molar yield, when expressing *paaHG* from *E. coli* and *ter* from *Treponema denticola* [107]. *Pseudomonas putida* is a known aromatic metabolizer and robust producer of medium chain length polyhydroxyalkanoates (mcl-PHAs). Several studies attempted to produce PHAs from aromatic compounds [108, 109], but the best strain utilized a strong, constitutive promoter, P_{tac}, to drive expression of *phaG*, *alkK*, *phaC1*, and *phaC2*. The strong expression coupled to a fed-batch culturing of *p*-coumaric acid overcame biomass limitations and toxicity limits, while low-nitrogen conditions maintained high cellular PHA content, resulting in 953 mg/L mcl-PHAs with a 54.3% PHA content per DCW. The same strain produced 116 mg/L mcl-PHAs, a 17.7% PHA content per DCW, when utilizing corn stover hydrolysate [110]. *Rhodotorula toruloides* IFO 0880 was engineered to produce bisabolene using a *BIS* gene from *Abies grandis*. When using 0.5 wt% *p*-coumaric acid as the sole carbon source, the engineered strain produced approximately 75 mg/L bisabolene [111]. Oleaginous organisms, such as the *Rhodococcus* species, *Cutaneotrichosporon* species, and *Rhodosproridium* species are noted for their ability to uptake aromatic monomers, but most species lose their ability to accumulate high quantities of lipids [111–119].

1.5 The future of lignocellulosic biomass valorization

1.5.0.1 Discovering novel species and enzymes

A keen observer may have noticed most of the funneling and ring-cleavage pathways described were from bacterial species. There is a significant gap in knowledge regarding the genetics for yeast and fungal systems. A more thorough understanding of the mechanisms underlying these complex phenotypes will further metabolic engineering for enhanced lignin bioconversion and aid in discovery of novel organisms and enzymes with potential for lignin metabolism through data mining [37, 120–126].

1.5.0.2 Oleochemical production

As noted earlier, low-cost feedstocks, such as lignin, are well suited for production of low-cost or bulk products, such as oleochemicals (biofuels, biodiesels, plant oil replacements, etc). Oleaginous yeasts are appropriate solutions, as they accumulate at least 20% of their biomass as neutral lipids and have a naturally high flux through acetyl-CoA pools [3–5] (Figure . There are over 1,500 yeast species known, but only 100 are considered oleaginous due to their accumulation of at least 20% of their biomass as neutral lipids [127, 128]. Significant attention has been given to several oleaginous yeast, including *Yarrowia lipolytica* [129–132], *Lipomyces starkeyi* [133], and *Rhodosporidium toruloides* [134]. Substantial work to expand the genetic engineering tools available for these oleaginous yeasts has enabled significant metabolic engineering of these species [135–141]. Despite being established industrial hosts with significant prior work, these yeast species are not suitable for utilizing certain low-cost

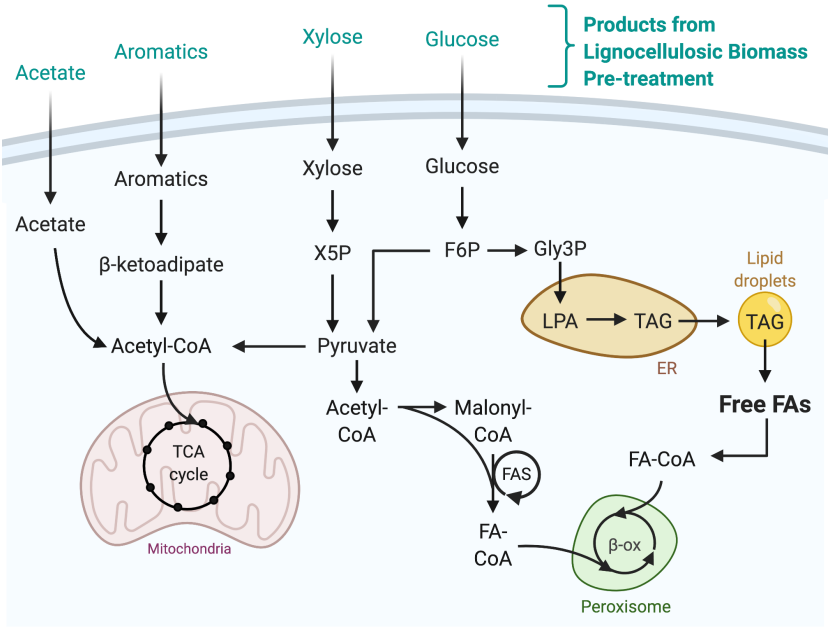


Figure 1.8. Schematic for fatty acid synthesis pathways in oleaginous yeasts from .

feedstocks, such as aromatic-rich lignin and phenolic wastewater streams.

To this end, *Cutaneotrichosporon oleaginosus* has gained attention as a non-model oleaginous yeast with particularly interesting properties [3–5, 142, 143]. It was first isolated in the 1950's in cheese plant floors and floor drains at Iowa State University, as researchers were hunting for an organism capable of metabolizing lactose [144]. Since it was first noted for metabolism of an alternate sugar substrate, most of the literature regarding this yeast focuses on its broad sugar metabolism [145–149]. The flexibility in its sugar metabolism supports its use for cellulosic and hemicellulosic biomass valorization. In fact, it has been grown in a number of industrial and agricultural wastewater [3, 142, 143]. Furthermore, *C. oleaginosus* often grows better in "dirty" feedstocks than synthetic, defined equivalents. This is likely due to this yeast's unique ability to metabolize a number of by-products associated with biomass hydrolysis, including toxic compounds such as 5-hydroxyfurfural, furfural, ammonia, and acetic acid [150–154]. Since these discoveries, a draft genome and annotation

were published and a metabolic model encompassing sugar-derived lipid production was developed [155, 156].

Preliminary searches through the genome hinted at the potential for *C. oleaginosus* to metabolize aromatic compounds. Related species, including *Cutaneotrichosporon guehoae* and *Cutaneotrichosporon cutaneum*, can metabolize aromatic compounds [117–119, 157–159]. The discovery of an organism that can tolerate the lignin compounds found in lignocellulosic biomass hydrolysates would facilitate use of biomass feedstocks without having to detoxify them and enable valorization of lignin itself. This dissertation focuses on discoveries of aromatic catabolism in *Cutaneotrichosporon oleaginosus* and its development as a potentially key microbe to convert lignin-derived aromatic compounds to oleochemicals.

1.6 Dissertation outline

The need to find an organism that can robustly convert lignin to value-added chemicals has resulted in extensive studies of microorganism both for identification of novel species and metabolic engineering of existing species that are genetically tractable. In effort to address the need for a robust metabolizer of lignin-derived feedstocks, the lacking yeast genome annotation underlying aromatic catabolism, and insufficient genetic tools developed for non-model, lignin-relevant organisms, this dissertation focuses on development of *Cutaneotrichosporon oleaginosus* as a potentially key microbe to convert lignin-derived aromatic compounds to oleochemicals.

Chapter 2 focuses on the screening of 36 oleaginous ascomycete and basidiomycete yeasts on six different model monoaromatic compounds: phenol, ferulic acid, *p*-coumaric acid, syringic acid, pHBA, and resoricnol. Glucose was included as a control. The *Rhodotorula*, *Rhodosporidium*, *Trichosporon*, and *Cutaneotrichosporon*

genera were revealed as strong metabolizers of a variety of aromatic substrates. *Cutaneotrichosporon oleaginosus* was further characterized for metabolism of phenol, *p*-coumarate, pHBA, and resorcinol and production of neutral lipids. While utilizing all compounds as sole carbon sources, *C. oleaginosus* remained oleaginous up to 50% on a per cell basis. This resulted in a publication in *Journal of Industrial Microbiology and Biotechnology* [160].

Chapter 3 focuses on characterization of *Cutaneotrichosporon oleaginosus* on three model lignin monomers — phenol, pHBA, and resorcinol — as sole carbon sources in high and low nitrogen sources to probe lipid accumulation and lipid profile. Cells cultured using fed-batch fermentation resulted in 69.9% lipid accumulation on a per cell basis — the second highest value reported for this yeast and one of the highest reported lipid accumulation of any wild-type oleaginous yeast utilizing model lignin monoaromatics as the sole carbon source. This resulted in a publication in *Microbial Cell Factories* [161].

Chapter 4 describes the growth and metabolism of *C. oleaginosus* while utilizing an alkaline pre-treated corn stover lignin. The major monomers released during pre-treatment were metabolized, but polymeric lignin is not broken down. It also addresses a large gap in the field regarding aromatic catabolism in yeast through discovery of the novel genes expressed in *C. oleaginosus* using a multi-omics dataset. This work is being prepared for publication.

Chapter 5 focuses on development of a genetic toolbox for *C. oleaginosus*. A small suite of strong, constitutive promoters and several steps towards an optimized transformation system are described.

Finally, conclusions and recommendations for future work are laid out in Chapter 6.

Bibliography

[1] Iskigor, F.; Becer, R. Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers.. *Poly Chem*, **2015**, 5, 447.

[2] Vanholme, R.; Demedts, B.; Morreel, K., *et al.* Lignin Biosynthesis and Structure. *Plant Physio*, **2010**, 153, 895–905.

[3] Yaguchi, A.; Rives, D.; Blenner, M. The New Kids on the Block: Emerging Oleaginous Yeast of Biotechnological Importance. *AIMS Microbiology*, **2017**, 3(2), 227–247.

[4] Yaguchi, A.; Spagnuolo, M.; Blenner, M. Engineering yeast for utilization of alternative feedstocks. *Curr Opin Biotechnol*, **2018**, 53, 122–129.

[5] Yaguchi, A.; Spagnuolo, M.; Blenner, M. Oleaginous yeast for biofuel and oleochemical production. *Curr Opin Biotechnol*, **2019**, 57, 73–81.

[6] Jin, J.; Ding, J.; Klett, A., *et al.* Carbon Fibers Derived from Fractionated–Solvated Lignin Precursors for Enhanced Mechanical Performance. *ACS Sustainable Chem Eng*, **2018**, 6(11), 14135–14142.

[7] Collins, M.; Nechifor, M.; Tanasa, F., *et al.* Valorization of lignin in polymer and composite systems for advanced engineering applications - A review. *Int. J. of Biol Macromol*, **2019**, 131, 828–849.

[8] Thakur, V.; Thakur, M.; Raghavan, P., *et al.* Progress in Gree Polymer Composites from Lignin for Multifunctional Applications: A Review. *ACS Sus Chem Eng*, **2014**, 2.

[9] Naseem, A.; Tabasum, S.; Zia, K., *et al.* Lignin-derivatives based polymers, blends and composites: A review. *Int J of Bio Macromol*, **2016**, 3, 296–313.

[10] Gosselink, R.JA.; Jong, E.; Guran, B., *et al.* Co-ordination network for lignin - standardisation, production and applications adapted to market requirements (EUROLIGNIN). *Ind Crop Prod*, **2004**, 20(2), 121–129.

[11] Thakur, V.; Thakur, M.; Raghavan, P., *et al.* Valorization of Biomass: Deriving More Value from Waste. *Science*, **2012**, 337, 695–699.

[12] Becker, J.; Whittmann, C. A field of dreams: Lignin valorization into chemicals, materials, fuels, and health-care products. *Biotech Adv*, **2019**, 37, 107360.

[13] Bugg, T.D.; Rahmanpour, R. Enzymatic conversion of lignin into renewable chemicals. *Curr Op in Chem Bio*, **2015**, 29, 10–17.

[14] Davis, R.; Tao, L.; Tan, E., *et al.* In *Process Design and Economics for the Conversion of Lignocellulosic Biomass to Hydrocarbons: Dilute-Acid and Enzymatic Deconstruction of Biomass to Sugars and Biological Conversion of Sugars to Hydrocarbons*. NREL, **2013**.

[15] Association, R.F. Ethanol Biorefinery Locations. Accessed 2020-11-17, 2020. https://ethanolrfa.org/biorefinery-locations/.

[16] Association, R.F. Annual World Fuel Ethanol Production (Mil. Gal.). Accessed 2020-11-17, 2020. https://ethanolrfa.org/statistics/annual-ethanol-production/.

[17] Klein-Marcuschamer, D.; Simmons, B.; Blanch, H. Techno-economic analysis of a lignocellulosic ethanol biorefinery with ionic liquid pre-treatment. *Biofuels, Bioproducts, and Biorefining*, **2011**, 5, 562–569.

[18] Bbosa, D.; Mba-Wright, M.; Brown, R. More than ethanol: a techno-economic analysis of a corn stover-ethanol biorefinery integrated with a hydrothermal liquefaction process to convert lignin into biochemicals. *Biofuels, Bioproducts, and Biorefining*, **2018**, 12, 497–509.

[19] Rijn, R.; Nieves, I.; Shanmugam, K., *et al.* Techno-Economic Evaluation of Cellulosic Ethanol Production Based on Pilot Biorefinery Data: a Case Study of Sweet Sorghum Bagasse Processed via L+SScF. *BioEnergy Res*, **2019**, 11, 414–425.

[20] Bozell, J.; Holladay, J.; Johnson, D., *et al.* In *Top Value Added Chemicals from Biomass*. PNNL, **2007**.

[21] Valdivia, M.; Galan, J.; Laffarga, J., *et al.* Biofuels 2020: Biorefineries based on lignocellulosic materials. *Microb Biotechnol*, **2016**, 9, 585–594.

[22] Ponnusamy, V.; Nguyen, D.; Dharmaraja, J., *et al.* A review on lignin structure, pretreatments, fermentation reactions and biorefinery potential. *Biores Tech*, **2019**, 271, 462–472.

[23] Mei, Q.; Shen, X.; Liu, H., *et al.* Selectively transform lignin into value-added chemicals. *Chinese Chem Letters*, **2018**, 125, 30183–30189.

[24] Lourenco, A.; Pereira, H. Compositional Variability of Lignin in Biomass. In *Lignin - Trends and Applications*, IntechOpen, **2018**, ???–???.

[25] Parthasarathi, R.; Romero, R.; A, e. Theoretical study of the remarkably diverse linkages in lignin. *European Journal of Lipid Science and Technology*, **2011**, 2, 2660–2666.

[26] Sette, M.; Wechselberger, R.; Crestini, C. Elucidation of lignin structure by quantitative 2D NMR. *Chemistry*, **2011**, 17(34), 9529–9535.

[27] Yang, H.; Yoo, C.; Meng, X., *et al.* Structural changes of lignins in natural Populus variants during different pretreatments. *Bioresour Technol*, **2020**, 295, 122240.

[28] Pelaez, F.; Martinez, M.; Martinez, A. Screening of 68 species of basidiomycetes for enzymes involved in lignin degradation. *Mycol Res*, **1995**, 99(1), 37–42.

[29] Knezevic, A.; Milovanovic, I.; Stajic, M., *et al.* Lignin degradation by selected fungal species. *Bioresour Tech*, **2013**, 138, 117–123.

[30] Varela, E.; Martinez, A.; Martinez, M. Southern blot screening for lignin peroxidase and aryl-alcohol oxidase genes in 30 fungal species. *J of Biotechnol*, **2000**, 83(3), 245–251.

[31] Arun, A.; Eyini, M. Comparative studies on lignin and polycyclic aromatic hydrocarbons degradation by basidiomycetes fungi. *Bioresour Tech*, **2011**, 102(17), 8063–8070.

[32] Levassseur, A.; Piumi, F.; Coutinho, P., *et al.* FOLy: An integrated database for the classification and functional annotation of fungal oxidoreductases potentially involved in the degradation of lignin and related aromatic compounds. *Fungal Genetics and Biol*, **2008**, 45, 638–645.

[33] Kameshwar, A.; Qin, W. Qualitative and Quantitative Methods for Isolation and Characterization of Lignin-Modifying Enzymes Secreted by Microorganisms. *Bioenergy Res*, **2017**, 10, 248–266.

[34] Abdelaziz, O.Y.; Brink, D.P.; Prothmann, J., *et al.* Biological valorization of low molecular weight lignin. *Biotechnol Adv*, **2016**, 34(8), 1318–1346.

[35] Li, C.; Chen, C.; Wu, X., *et al.* Recent advancement in lignin biorefinery: With special focus on enzymatic degradation and valorization. *Proc Natl Acad Sci U S A*, **2019**, 291, 121989.

[36] Li, X.; Zheng, Y. Biotransformation of lignin: Mechanisms, applications and future work. *Biotech Progress*, **2019**, e2922.

[37] Vob, H.; Heck, C.; Schallmey, M., *et al.* Database Mining for Novel Bacterial beta-Etherases, Glutathione-Dependent Lignin-Degrading Enzymes. *App Env Micro*, **2020**, 86, e02026.

[38] Wang, Y.; Liu, Q.; Yan, L., *et al.* A novel lignin degradation bacterial consortium for efficient pulping. *Bioresour Tech*, **2013**, 2013, 113–119.

[39] Ahmad, M.; Taylor, C.; Pink, D., *et al.* Development of novel assays for lignin degradation: comparative analysis of bacterial and fungal lignin degraders. *Mol Biosyst*, **2010**, 6, 815–821.

[40] Ramachandra, M.; Crawford, D.; Hertel, G. Characterization of an Extracellular Lignin Peroxidase of the Lignocellulolytic Actinomycete Streptomyces viridosporus. *Appl Environ Microb*, **1988**, 54(12), 3057–3063.

[41] Salvachúa, D.; Karp, E.M.; Nimlos, C.T., *et al.* Towards lignin consolidated bioprocessing: simultaneous lignin depolymerization and product generation by bacteria.. *Green Chem*, **2015**, 17, 4951–4967.

[42] Williams, P.; Murray, K. Metabolism of benzoate and the methylbenzoates by Pseudomonas putida (arvilla) mt-2: Evidence for the existence of a TOL plasmid. *J of Bacteriol*, **1974**, 120(1), 416–423.

[43] ??? Structure of Thermobifida fusca DyP-type peroxidase and activity towards Kraft lignin and lignin model compounds. *Archives of Biochem and Biophys*, **2016**, 594, 54–60.

[44] He, Y.; Li, X.; Ben, H., *et al.* Lipid production from dilute alkali corn stover lignin by Rhodococcus strains. *ACS Sus Chem Eng*, **2017**, 5, 2302–2311.

[45] Fisher, A.; Fong, S. Lignin biodegradation and industrial implications. *AIMS Bioeng*, **2014**, 1(2), 92–112.

[46] Tien, M.; Kirk, T. Lignin-degrading enzyme from Phanerochaete chrysosporium: Purification, characterization, and catalytic properties of a unique H2O2-requiring oxygenase. *PNAS*, **1984**, 81(8).

[47] Miki, K.; Renganathan, V.; Gold, M. Mechanism of beta-aryl ether dimeric lignin model compound oxidation by lignin peroxidase by Phanerochaete chrysosporium. *Biochem*, **1986**, 25, 4790–4796.

[48] Pollegioni, L.; Tonin, F.; Rosini, E. Lignin-degrading enzymes. *FEBS Journ*, **2015**, 282, 1190–1213.

[49] Orth, A.; Royse, D.; Tien, M. Ubiquity of Lignin-Degrading Peroxidases among Various Wood-Degrading Fungi. *App Env Microbiol*, **1993**, 59(12), 4017–4023.

[50] Lubbers, R.; Dilokpimol, A.; Visser, J., *et al.* A comparison between the homocyclic aromatic metabolic pathways from plant-derived compounds by bacteria and fungi. *Biotech Adv*, **2019**, 37, 107396.

[51] Smith, M.; Weaver, V.; Young, D., *et al.* Genes for Chlorogenate and Hydroxycinnamate Catabolism (hca) Are Linked to Functionally Related Genes in the dca-pca-qui-pob-hca Chromosomal Cluster of Acinetobacter sp. Strain ADP1. *App Env Microbio*, **2003**, 69, 524–532.

[52] Otani, H.; Lee, Y.; Casabon, I., *et al.* Characterization of p-Hydroxycinnamate Catabolism in a Soil Actinobacterium. *J of Bacteriol*, **2014**, 196(24), 4293–4303.

[53] Lubbers, R.; Dilokpimol, A.; Peng, M., *et al.* Discovery of Novel p-Hydroxybenzoate-m-hydroxylase, Protocatechuate 3,4 Ring-Cleavage Dioxygenase, and Hydroxyquinol 1,2 Ring-Cleavage Dioxygenase from the Filamentous Fungus Aspergillus niger. *ACS Sus Chem*, **2019**, 7, 19081–19089.

[54] Yoshida, T.; Inami, Y.; Matsui, T., *et al.* Regioselective carboxylation of catechol by 3,4-dihydroxybenzoate decarboxylase of Enterobacter cloacae. *P Biotehnol Lett*, **2010**, 32(5), 701–705.

[55] Grant, D.; Patel, J. The non-oxidative decarboxylation ofp-hydroxybenzoic acid, gentisic acid, protocatechuic acid and gallic acid byKlebsiella aerogenes (Aerobacter aerogenes). *Antonie van Leeuwenhoek*, **1969**, 35, 325–343.

[56] Anderson, J.J.; Dagley, S. Catabolism of tryptophan, anthranilate, and 2,3-dihydroxybenzoate in Trichosporon cutaneum. *J Bacteriol*, **1981**, 146(1), 291–297.

[57] Kamath, A.V.; Rao, N.A.; Vaidyanathan, C.S. Enzyme catalysed non-oxidative decarboxylation of aromatic acids. II. Identification of active site residues of 2,3-dihydroxybenzoic acid decarboxylase from Aspergillus niger. *Biochem Biophys Res Commun*, **1989**, 165, 20–26.

[58] Johnson, C.; Salvachua, D.; Khanna, P., *et al.* Enhancing muconic acid production from glucose and lignin-derived aromatic compounds via increased protocatechuate decarboxylase activity. *Met Eng Commun*, **2016**, 3, 111–119.

[59] Sonoki, T.; Morooka, M.; Sakamoto, K., *et al.* Enhancement of protocatechuate decarboxylase activity for the effective production of muconate from lignin-related aromatic compounds. *J Biotechnol*, **2014**, 192, 71–77.

[60] Vardon, D.R.; Franden, M.A.; Johnson, C.W., *et al.* Adipic acid production from lignin. *Energy Environ Sci*, **2015**, 8, 617–628.

[61] Salvachua, D.; Johnson, C.; Singer, C., *et al.* Bioprocess development for muconic acid production from aromatic compounds and lignin.. *Green Chem*, **2018**, 21, 5007–5019.

[62] Mallinson, S.; Machinova, M.; Silveira, R., *et al.* A promiscuous cytochrome P450 aromatic O-demethylase for lignin bioconversion. *Nature Comm*, **2018**, 9, 2487.

[63] Garcia-Hidalgo, J.; Ravi, K.; Kure, L., *et al.* Identification of the two-component guaiacol demethylase system from Rhodococcus rhodochrous and expression in Pseudomonas putida EM42 for guaiacol assimilation. *AMB Exp*, **2019**, 9, 34.

[64] Fetherolf, M.; Levy-Booth, D.; Navas, L., *et al.* Characterization of alkylguaiacol-degrading cytochromes P450 for the biocatalytic valorization of lignin. *PNAS*, **2020**, 117, 25771–25778.

[65] Venturi, V.; Zennaro, F.; Degrassi, G., *et al.* Genetics of ferulic acid bioconversion to protocatechuic acid in plant-growth-promoting Pseudornonas putida WCS358. *Microbiol*, **1998**, 144, 965–973.

[66] Overhage, J.; Priefert, H.; Steinbuchel, A. Biochemical and Genetic Analyses of Ferulic Acid Catabolism in Pseudomonas sp. Strain HR199. *App Env Microbiol*, **1999**, 65(11), 4837–4847.

[67] Plaggenborg, R.; Overhage, J.; Steinbuchel, A., *et al.* Functional analyses of genes involved in the metabolism of ferulic acid in Pseudomonas putida KT2440. *Appl Microbiol Biotechnol*, **2003**, 61, 528–535.

[68] Morawski, B.; Segura, A.; Ornston, L.N. Substrate Range and Genetic Analysis of Acinetobacter Vanillate Demethylase. *J Bacteriol*, **2000**, 182, 1383–1389.

[69] Priefert, H.; Rabenhorst, J.; Steinbuchel, A. Molecular characterization of genes of Pseudomonas sp. strain HR199 involved in bioconversion of vanillin to protocatechuate. *J Bacteriol*, **1997**, 179, 2595–2607.

[70] Buswell, J.A.; Ribbons, D.W. Vanillate O-demethylase from Pseudomonas species. *Lignin*, **1988**, 161, 294–301.

[71] Nishimura, M.; Nishimura, Y.; Abe, C., *et al.* Expression and substrate range of Streptomyces vanillate demethylase. *Biol Pharm Bull*, **2014**, 37(9), 1564–1568.

[72] Abe, T.; Masai, E.; Miyauchi, K., *et al.* A tetrahydrofolate-dependent O-demethylase, LigM, is crucial for catabolism of vanillate and syringate in Sphingomonas paucimobilis SYK-6. *J Bacteriol*, **2005**, 187(6), 2030–2037.

[73] Machovina, M.; Malinson, S.; Knott, B., *et al.* Enabling microbial syringol conversion through structure-guided protein engineering. *PNAS*, **2019**, 116, 13970–13976.

[74] Harwood, C.S.; Parales, R.E. The beta-ketoadipate pathway and the biology of self-identity. *Annu Rev Microbiol*, **1996**, 50, 553–590.

[75] Brink, D.; Ravi, K.; Liden, G., *et al.* Mapping the diversity of microbial lignin catabolism: experiences from the eLignin database. *Appl Microbiol Biotechnol*, **2019**, 103, 3979–4002.

[76] Spence, E.; Scott, H.; Dumond, L., *et al.* Characterisation of the hydroxyquinol degradation pathway in Rhodococcus jostii RHA1 and Agrobacterium sp.: an alternative pathway for degradation of protocatechuic acid and lignin degradation fragments. *Appl Environ Microbiol*, **1990**, 136, 1491–1495.

[77] Suzuki, K.; Itoh, M. Metabolism of p-Hydroxybenzoate via Hydroxyquinol by Trichosporon cutaneum WY2-2: Characterization of the Pathway using Superoxide Dismutase as a Stabilizer of Hydroxyquinol. *Plant Cell Physiology*, **1986**, 27(8), 1451–1460.

[78] Semana, P.; Powlowski, J. Four Aromatic Intradiol Ring Cleavage Dioxygenases from Aspergillus niger. *Appl Eviron Microbiol*, **2019**, 85, e01786.

[79] Martins, T.; Hartmann, D.; Planchon, S., *et al.* The old 3-oxoadipate pathway revisited: new insights in the catabolism of aromatics in the saprophytic fungus Aspergillus nidulans. *Fungal Genet Biol*, **2015**, 74, 32–44.

[80] Crawford, R.; Bromley, J.; Perkins-Olson, P. Catabolism of Protocatechuate by Bacillus macerans. *Appl Environ Microbiol*, **1979**, 37, 614–618.

[81] Kasai, D.; Fujinami, T.; Abe, T., *et al.* Uncovering the protocatechuate 2,3-cleavage pathway genes. *J Bacteriol*, **2009**, 191, 6758–6768.

[82] Hara, H.; Masai, E.; Miyauchi, K., *et al.* Characterization of the 4-Carboxy-4-Hydroxy-2-Oxoadipate Aldolase Gene and Operon Structure of the Protocatechuate 4,5-Cleavage Pathway Genes in Sphingomonas paucimobilis SYK-6. *J Bacteriol*, **2003**, 185, 41–50.

[83] Providenti, M.; Mampel, J.; MacSween, S., *et al.* Comamonas testosteroni BR6020 possesses a single genetic locus for extradiol cleavage of protocatechuate. *Microbiol*, **2001**, 147, 2157–2167.

[84] Kasai, D.; Masai, E.; Katayama, Y., *et al.* Degradation of 3-O-methylgallate in Sphingomonas paucimobilis SYK-6 by pathways involving protocatechuate 4,5-dioxygenase. *FEMS Micriobiol Lett*, **2005**, 274, 323–328.

[85] Notales, J.; Canales, A.; Jimenez-Barbero, J., *et al.* Molecular characterization of the gallate dioxygenase from Pseudomonas putida KT2440. The prototype of a new subgroup of extradiol dioxygenases. *J Biol Chem*, **2005**, 280, 35382–35390.

[86] Tack, B.; Chapman, P.; Dagley, S. Metabolism of Gallic Acid and Syringic Acid by Pseudomonas putida. *J Biol Chem*, **1972**, 247, 6438–6443.

[87] Biggs, B.W.; Bedore, S.R.; Arvay, E., *et al.* Development of a genetic toolset for the highly engineerable and metabolically versatile Acinetobacter baylyi ADP1. *Nucleic Acids Res*, **2020**.

[88] Tumen-Velasquez, M.; Johnson, C.W.; Ahmed, A., *et al.* Accelerating pathway evolution by increasing the gene dosage of chromosomal segments. *Proc Natl Acad Sci U S A*, **2018**, 115(27), 7105–7110.

[89] Barry, K.P.; Cohn, E.F.; Ngu, A., *et al.* Improving alternate lignin catabolite utilization of LigAB from Sphingobium sp. strain SYK-6 through site directed mutagenesis. *Process Biochemistry*, **2015**, 50(10), 1634–1639.

[90] Standaert, R.F.; Giannone, R.J.; Michener, J.K. Identification of parallel and divergent optimization solutions for homologous metabolic enzymes. *Metab Eng Commun*, **2018**, 6, 56–62.

[91] Close, D.M.; Cooper, C.J.; Wang, X., *et al.* Horizontal transfer of a pathway for coumarate catabolism unexpectedly inhibits purine nucleotide biosynthesis. *Mol Microbiol*, **2019**, 112(6), 1784–1797.

[92] Shen, Y.; Li, H.; Wang, X., *et al.* High vanillin tolerance of an evolved Saccharomyces cerevisiae strain owing to its enhanced vanillin reduction and antioxidative capacity. *J Ind Microbiol Biotechnol*, **2014**, 41(11), 1637–1645.

[93] Wang, X.; Liang, Z.; Hou, J., *et al.* The Absence of the Transcription Factor Yrr1p, Identified from Comparative Genome Profiling, Increased Vanillin Tolerance Due to Enhancements of ABC Transporters Expressing, rRNA Processing and Ribosome Biogenesis in Saccharomyces cerevisiae. *Front Microbiol*, **2017**, 8, 367.

[94] Wu, W.; Dutta, T.; Varman, A.M., *et al.* Lignin Valorization: Two Hybrid Biochemical Routes for the Conversion of Polymeric Lignin into Value-added Chemicals. *Sci Rep*, **2017**, 7(1), 8420.

[95] Sonoki, T.; Takahashi, K.; Sugita, H., *et al.* Glucose-free cis, cis-muconic acid production via new metabolic designs corresponding to the heterogeneity of lignin. *ACS Sustain Chem Eng*, **2017**, 6, 1256–1264.

[96] Kohlstedt, M.; Starck, S.; Barton, N., *et al.* From lignin to nylon: Cascaded chemical and biochemical conversion using metabolically engineered Pseudomonas putida. *Metab Eng*, **2018**, 47, 279–293.

[97] Johnson, C.W.; Abraham, P.E.; Linger, J.G., *et al.* Eliminating a global regulator of carbon catabolite repression enhances the conversion of aromatic lignin monomers to muconate in Pseudomonas putida KT2440. *Metabolic Engineering Communications*, **2017**, 5, 19–25.

[98] Becker, J.; Kuhl, M.; Kohlstedt, M., *et al.* Metabolic engineering of Corynebacterium glutamicum for the production of cis, cis-muconic acid from lignin. *Microb Cell Fact*, **2018**, 17(1), 115.

[99] Johnson, C.W.; Beckham, G.T. Aromatic catabolic pathway selection for optimal production of pyruvate and lactate from lignin. *Metab Eng*, **2015**, 28, 240–247.

[100] Mycroft, Z.; Gomis, M.; Mines, P., *et al.* Biocatalytic conversion of lignin to aromatic dicarboxylic acids in Rhodococcus jostii RHA1 by re-routing aromatic degradation pathways. *Green Chemistry*, **2015**, 17(15), 4974–4979.

[101] Qian, Y.; Otsuka, Y.; Sonoki, T., *et al.* Engineered microbial production of 2-pyrone-4,6-dicarboxylic acid from lignin residues for use as an industrial platform chemical. *Bioresources*, **2016**, 11(3), 6097–6109.

[102] Higuchi, Y.; Kato, R.; Tsubota, K., *et al.* Discovery of novel enzyme genes involved in the conversion of an arylglycerol-beta-aryl ether metabolite and their use in generating a metabolic pathway for lignin valorization. *Metab Eng*, **2019**, 55, 258–267.

[103] Johnson, C.W.; Salvachúa, D.; Rorrer, N.A., *et al.* Innovative Chemicals and Materials from Bacterial Aromatic Catabolic Pathways. *Joule*, **2019**, 3(6), 1523–1537.

[104] Suzuki, Y.; Okamura-Abe, Y.; Nakamura, M., *et al.* Development of the production of 2-pyrone-4,6-dicarboxylic acid from lignin extracts, which are industrially formed as by-products, as raw materials. *Journal of Bioscience and Bioengineering*, **2020**.

[105] Perez, J.M.; Kontur, W.S.; Alherech, M., *et al.* Funneling aromatic products of chemically depolymerized lignin into 2-pyrone-4-6-dicarboxylic acid with Novosphingobium aromaticivorans. *Green Chemistry*, **2019**, 21(6), 1340–1350.

[106] Luo, J.; Lehtinen, T.; Efimova, E., *et al.* Synthetic metabolic pathway for the production of 1-alkenes from lignin-derived molecules. *Microb Cell Fact*, **2019**, 18(1), 48.

[107] Niu, W.; Willett, H.; Mueller, J., *et al.* Direct biosynthesis of adipic acid from lignin-derived aromatics using engineered Pseudomonas putida KT2440. *Metab Eng*, **2020**.

[108] Lin, L.; Cheng, Y.; Pu, Y., *et al.* Systems biology-guided biodesign of consolidated lignin conversion. *Green Chemistry*, **2016**, 18(20), 5536–5547.

[109] Wang, X.; Lin, L.; Dong, J., *et al.* Simultaneous Improvements of Pseudomonas Cell Growth and Polyhydroxyalkanoate Production from a Lignin Derivative for Lignin-Consolidated Bioprocessing. *Appl Environ Microbiol*, **2018**, 84(18), e01469–89.

[110] Salvachua, D.; Rydzak, T.; Auwae, R., *et al.* Metabolic engineering of Pseudomonas putida for increased polyhydroxyalkanoate production from lignin. *Microb Biotechnol*, **2020**, 13(1), 290–298.

[111] Yaegashi, J.; Kirby, J.; Ito, M., *et al.* Rhodosporidium toruloides: a new platform organism for conversion of lignocellulose into terpene biofuels and bioproducts. *Biotech for Biofuels*, **2017**, 10(241).

[112] Kosa, M.; Ragauskas, A.J. Bioconversion of lignin model compounds with oleaginous Rhodococci. *Appl Microbiol Biotechnol*, **2012**, 93(2), 891–900.

[113] Henson, W.R.; Campbell, T.; DeLorenzo, D.M., *et al.* Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus. *Metab Eng*, **2018**, 49, 69–83.

[114] Wei, Z.; Zeng, G.; Huang, F., *et al.* Bioconversion of oxygen-pretreated Kraft lignin to microbial lipid with oleaginous Rhodococcus opacus DSM 1069. *Green Chem*, **2015**, 17, 2784–2789.

[115] Ahmad, M.; Roberts, J.; Haridman, E., *et al.* Identification of DypB from Rhodococcus jostii RHA1 as a lignin peroxidase. *Biochem*, **2011**, 50, 5096–5107.

[116] Chen, X.; Huang, C.; Xiong, L., *et al.* Microbial oil production from corncob acid hydrolysate by Trichosporon cutaneum. *Biotech Lett*, **2012**, 34, 1025–1028.

[117] Alexieva, Z.; Gerginova, M.; Zlateva, P., *et al.* Monitoring of aromatic pollutants biodegradation. *Biochemical Engineering Journal*, **2008**, 40(2), 233–240.

[118] Nogue, V.; Black, B.A.; Kruger, J.S., *et al.* Integrated diesel production from lignocellulosic sugars via oleaginous yeast. *Green Chem*, **2018**, 20, 4349–4365.

[119] Sitepu, I.; Jin, M.; Fernandez, J.E., *et al.* Identification of oleaginous yeast strains able to accumulate high intracellular lipids when cultivated in alkaline pretreated corn stover. *Appl Microbiol Biotech*, **2014**, 98, 7645–7657.

[120] Picart, P.; Muller, C.; Mottweiler, J., *et al.* From Gene Towards Selective Biomass Valorization: Bacterial beta-Etherases with Catalytic Activity on Lignin-Like Polymers. *Chem Sus Chem*, **2014**, 7, 3164–3171.

[121] Kumar, M.; Verma, S.; Gazara, R., *et al.* Genomic and proteomic analysis of lignin degrading and polyhydroxyalkanoate accumulating beta-proteobacterium Pandoraea sp. ISTKB. *Biotech for Biofuels*, **2018**, 11, 154.

[122] Kameshwar, A.; Qin, W. Recent Developments in Using Advanced Sequencing Technologies for the Genomic Studies of Lignin and Cellulose Degrading Microorganisms. *Int J Biol Sci*, **2016**, 12, 156–171.

[123] Goncalves, C.; Bruce, T.; Silva, C., *et al.* Bioprospecting Microbial Diversity for Lignin Valorization: Dry and Wet Screening Methods. *Front microbiol*, **2020**, 11.

[124] Riley, R.; Salamov, A.; Brown, D., *et al.* Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. *PNAS*, **2014**, 111(27), 9923–9928.

[125] Dilokpimol, A.; Makela, M.; Cerullo, G., *et al.* Fungal glucuronoyl esterases: Genome mining based enzyme discovery and biochemical characterization. *New Biotech*, **2018**, 40, 282–287.

[126] Pham, L.; Kim, Y. Discovery and characterization of new O-methyltransferase from the genome of the lignin-degrading fungus Phanerochaete chrysosporium for enhanced lignin degradation. *Enzyme and Microb Technol*, **2016**, 82, 66–73.

[127] Garay, L.; Sitepu, I.; Cajka, T., *et al.* Eighteen new oleaginous yeast species. *JIMB*, **2016**, 43, 887–900.

[128] Sitepu, I.; Garay, L.; Sestric, R., *et al.* Oleaginous yeasts for biodiesel: Current and future trends in biology and production. *Journal of Biotech Adv*, **2014**, 32, 1336–1360.

[129] Nicaud, J.M. Yarrowia lipolytica. *Yeast*, **2012**, 29(10), 409–418.

[130] Rodriguez, G.M.; Hussain, M.S.; Gambill, L., *et al.* Engineering xylose utilization in Yarrowia lipolytica by understanding its cryptic xylose pathway. *Biotechnol Biofuels*, **2016**, 9, 149.

[131] Qiao, K.; Abidi, S.HI.; Liu, H.J., *et al.* Engineering lipid overproduction in the oleaginous yeast Yarrowia lipolyticale. *Metabolic Engineering*, **2015**, 29, 56–65.

[132] Blazeck, J.; Hill, A.; Liu, L.Q., *et al.* Harnessing Yarrowia lipolytica lipogenesis to create a platform for lipid and biofuel production. *Nat Commun*, **2014**, 5.

[133] Wang, Q.; Yurkov, A.; Goker, M., *et al.* Phylogenetic classification of yeasts and related taxa within Pucciniomycotina. *Stud Mycol*, **2015**, 81, 149–189.

[134] Zhang, S.; Ito, M.; Skerker, J.M., *et al.* Metabolic engineering of the oleaginous yeast Rhodosporidium toruloides IFO0880 for lipid overproduction during high-density fermentation. *Appl Microbiol Biotechnol*, **2016**, 100(21), 9393–9405.

[135] Schwartz, C.; Shabbir-Hussain, M.; Frogue, K., *et al.* Standardized Markerless Gene Integration for Pathway Engineering in Yarrowia lipolytica. *ACS Synth Biol*, **2017**, 6(3), 402–409.

[136] Schwartz, C.M.; Hussain, M.S.; Blenner, M., *et al.* Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR-Cas9-Mediated Genome Editing in Yarrowia lipolytica. *Acs Synth Biol*, **2016**, 5(4), 356–359.

[137] Hussain, M.; Gambill, L.; Smith, S., *et al.* Engineering Promoter Architecture in Oleaginous Yeast Yarrowia lipolytica. *Acs Synth Biol*, **2016**, 5(3), 213–223.

[138] Hussain, M.; Rodriguez, G.; Gao, D., *et al.* Recent advances in bioengineering of the oleaginous yeast Yarrowia lipolytica. *AIMS Bioengineering*, **2016**, 3(4), 493–514.

[139] Blazeck, J.; Liu, L.Q.; Redden, H., *et al.* Tuning Gene Expression in Yarrowia lipolytica by a Hybrid Promoter Approach. *Appl Environ Microb*, **2011**, 77(22), 7905–7914.

[140] Blazeck, J.; Reed, B.; Garg, R., *et al.* Generalizing a hybrid synthetic promoter approach in Yarrowia lipolytica. *Appl Microbiol Biot*, **2013**, 97(7), 3037–3052.

[141] Dulermo, R.; Brunel, F.; Dulermo, T., *et al.* Using a vector pool containing variable-strength promoters to optimize protein production in Yarrowia lipolytica. *Microb Cell Fact*, **2017**, 16(1), 31.

[142] Sreeharsha, R.; Mohan, S. Obscure yet Promising Oleaginous Yeasts for Fuel and Chemical Production. *Trends in Biotech*, **2020**.

[143] Navarrete, C.; Jacobsen, I.; Martinez, J., *et al.* Cell Factories for Industrial Production Processes: Current Issues and Emerging Solutions. *Processes*, **2020**, 8, 768.

[144] Moon, N.; Hammond, E.; Glatz, B. Conversion of cheese whey and whey permeate to oil and single-cell protein. *J Dairy Sci*, 61, 1537–type = Journal Article.

[145] Hassan, M.; Blanc, P.; Pareilleux, A., *et al.* Production of cocoa butter equivalents from prickly-pear juice fermentation by an unsaturated fatty acid auxotroph of Cryptococcus curvatus grown in batch culture. *Process Biochem*, **1995**, 30, 629–634.

[146] Liang, Y.; Jarosz, K.; Wardlow, A., *et al.* Lipid Production by Cryptococcus curvatus on Hydrolysates Derived from Corn Fiber and Sweet Sorghum Bagasse Following Dilute Acid Pretreatment. *Appl Biochem and Biotechnol*, **2014**, 173, 2086–2098.

[147] Heredia, L.; Ratledge, C. Simultaneous utilization of glucose and xylose by Candida curvata D in continuous culture. *Biotech Lett*, **1988**, 10, 25–30.

[148] Görner, C.; Redai, V.; Bracharz, F., *et al.* Genetic engineering and production of modified fatty acids by the non-conventional oleaginous yeast Trichosporon oleaginosus ATCC 20509. *Green Chem.*, **2016**, 18(7), 2037–2046.

[149] Gong, Z.; Zhou, W.; Shen, H., *et al.* Co-utilization of corn stover hydrolysates and biodieselderived glycerol by Cryptococcus curvatus for lipid production. *Bioresour Technol*, **2016**, 219, 552–558.

[150] Gong, Z.W.; Shen, H.W.; Zhou, W.T., *et al.* Efficient conversion of acetate into lipids by the oleaginous yeast Cryptococcus curvatus. *Biotechnology for Biofuels*, **2015**, 8.

[151] Beligon, V.; Poughon, L.; Christophe, G., *et al.* Improvement and modeling of culture parameters to enhance biomass and lipid production by the oleaginous yeast Cryptococcus curvatus grown on acetate. *Bioresour Tech*, **2015**, 192, 582–591.

[152] Christophe, G.; Deo, J.; Kumar, V., *et al.* Production of oils from acetic acid by the oleaginous yeast Cryptococcus curvatus. *Appl Biochem Biotechnol*, **2012**, 165, 1270–1279.

[153] Yu, X.; Zheng, Y.; Dorgan, K.M., *et al.* Oil production by oleaginous yeasts using the hydrolysate from pretreatment of wheat straw with dilute sulfuric acid. *Bioresour Technol*, **2011**, 102(10), 6134–6140.

[154] Zheng, Y.; Chi, Z.; Ahring, B., *et al.* Oleaginous yeast Cryptococcus curvatus for biofuel production: Ammonia's effect. *Biomass and Bioen*, 37, 114–type = Journal Article.

[155] Kourist, R.; Bracharz, F.; Lorenzen, J., *et al.* Genomics and Transcriptomics Analyses of the Oil-Accumulating Basidiomycete Yeast Trichosporon oleaginosus: Insights into Substrate Utilization and Alternative Evolutionary Trajectories of Fungal Mating Systems. *MBio*, **2015**, 6(4), e00918.

[156] Pham, N.; Reijnders, M.; Suarez-Diez, M., *et al.* Genome-scale metabolic modelling underscores the potential of Cutaneotrichosporon oleaginosus ATCC 20509 as a cell factory for biofuel production. *Biotech for Biofuels*, **2020**.

[157] Anderson, J.J.; Dagley, S. Catabolism of Aromatic-Acids in Trichosporon-Cutaneum. *J Bacteriol*, **1980**, 141(2), 534–543.

[158] Gaal, A.; Neujahr, H.Y. Metabolism of Phenol and Resorcinol in Trichosporon-Cutaneum. *J Bacteriol*, **1979**, 137(1), 13–21.

[159] Gonzalez-Garcia, Y.; Hernandez, R.; Zhang, G.C., *et al.* Lipids accumulation in Rhodotorula glutinis and Cryptococcus curvatus growing on distillery wastewater as culture medium. *Environ Prog Sustain*, **2013**, 32(1), 69–74.

[160] Yaguchi, A.; Franaszek, N.; O'Neill, K., *et al.* Identification of oleaginous yeasts that metabolize aromatic compounds. *JIMB*, **2020**.

[161] Yaguchi, A.; Robinson, A.; Mihealsick, E., *et al.* Metabolism of aromatics by Trichosporon oleaginosus while remaining oleaginous. *Microb Cell Fact*, **2017**, 16(1), 206.

Chapter 2

Phenotypic screening of 36 oleaginous yeast for aromatic metabolism

2.1 Abstract

As mentioned in Chapter 1, the valorization of lignin is critical for the economic viability of the bioeconomy. Microbial metabolism is advantageous for handling the myriad of aromatic compounds resulting from lignin chemical or enzymatic depolymerization. Coupling aromatic metabolism to fatty acid biosynthesis makes possible the production of biofuels, oleochemicals, and other fine/bulk chemicals derived from lignin. To identify robust metabolizers of aromatic compounds, thirty-six oleaginous yeast species from the Phaff yeast collection were screened for growth on several compounds representing S-, G-, and H- type lignin. The analysis reported in this chapter

This chapter is based off **Yaguchi, A.**, et al. *JIMB*, **2020**, **47**(9-10) with permissions from Springer according to license number 4916761217613.

suggests that aromatic metabolism is largely segregated to the *Cutaenotrichosporon*, *Trichosporon*, and *Rhodotorula* genera. Each species tested within each genera has different properties with respect to the aromatics metabolized and the concentrations of aromatics tolerated. The combined analysis suggests that *Cutaneotrichosporon* yeast are the best suited to broad spectrum aromatic metabolism and support its development as a model system for aromatic metabolism in yeast.

2.2 Introduction

The biomass economy has largely focused on the sugars derived from the cellulose and hemicellulose fractions of lignocellulose; however, recent studies by the U.S. Department of Energy show that valorization of lignin is necessary to make biofuels economical [1–3]. A billion-ton biomass economy would produce approximately 225 million tons of lignin by-product [2, 4]. Proposed uses for this lignin are largely low-value, including polyurethane products [5], and phenolic adhesives [6], or are niche markets requiring significant lignin purification, such as carbon fibers [7–9]. Several barriers prevent the valorization of lignin, including its heterogenous composition and the resulting heterogeneity of its depolymerized components [10–13]. Another use for this lignin is to use it as a feedstock for bioproduction of fuels and other chemicals. While many catalytic processes require fairly pure feedstocks, microbial systems have sufficient funneling pathways to metabolize a wide variety of feedstocks [14–16]. Therefore, lignin depolymerization followed by microbial valorization of the small oligomers and monomers is increasingly studied [17–20]. The majority of microbial systems studied for their aromatic metabolism are from a limited number of genera, including the *Pseudomonas*, *Rhodococcus*, and *Sphingobium*, as most microorganisms are not capable of aromatic metabolism [14]. Synthetic biology and metabolic engi-

neering tools have been successfully applied to improve the metabolism of aromatics as well as the production of compounds such as muconic acid, vanillin, and polyhydroxyalkanoates [21–24].

Yeasts are far less studied for aromatic metabolism; however, they offer many advantages over bacteria, including resistance to phage infection and tolerance to extreme pH, high osmolarity, and solvents. Even more compelling is the eukaryotic cell physiology that enables the production and quality control of fungal enzymes for the modification and breakdown of polymeric lignin [25–27]. Approximately 100 of the over 1,600 known yeast are considered oleaginous due to their accumulation of greater than 20% of their biomass as neutral lipids, primarily triacylglycerides (TAGs) [28, 29]. These yeast have been increasingly studied for the sustainable production of biofuels and oleochemicals [30, 31]. Since the products of aromatic metabolism via the *ortho*-cleavage pathway, namely acetyl-CoA, are the precursors for fatty acid biosynthesis, identification of oleaginous yeasts able to metabolize aromatic compounds could increase product yields. Unfortunately, like most microbes, metabolism of aromatic compounds is an uncommon phenotype exhibited by yeasts.

Previous studies in bioremediation identified *Trichosporon cutaneum* (now *Cutaneotrichosporon cutaneum*) as an aromatic metabolizer [32–38], but the cells were not shown to retain oleaginous behavior when grown on aromatics. We previously reported *Cutaneotrichosporon oleaginosus*, a closely related yeast, grew on phenol, 4-hydroxybenzoic acid (pHBA) and resorcinol as sole carbon sources. This yeast was shown to accumulate almost 70% of dry weight as lipids using resorcinol as the sole carbon source [39]. Another recent study of the oleaginous red yeast, *Rhodosporidium toruloides* (now *Rhodotorula toruloides*) identified metabolism of *p*-coumaric acid, ferulic acid, vanillic acid, and pHBA as sole carbon sources, but did not describe resulting lipid accumulation. The same study also demonstrated *Cutaneotrichosporon*

guehoae's ability to metabolize pHBA as the sole carbon source [40]. Other reports of aromatic metabolism in yeast come from *Candida* sp. [41–43] and some species of heterobasidiomycetous yeasts, particularly those in the order Sporidiobolales [44].

In order to gain a wider appreciation of the prevalence of aromatic metabolism amongst oleaginous yeast, thirty-six oleaginous yeast were screened from the Phaff Yeast Culture Collection. These yeast were screened for growth on aromatics as sole carbon sources, including phenol, resorcinol, *p*-coumarate, pHBA, syringate, and ferulate. Most of the yeast that were able to utilize several aromatic compounds were from genera *Cutaneotrichosporon* and *Trichosporon* (class Tremellomycetes, order Trichosporonales) and genera *Rhodotorula* and *Rhodosporidiobolus* (class Microbotryomycetes, order Sporidiobolales). Overall, metabolism of S- and G-type lignin was rare and phenol toxicity limited its consumption. A deeper analysis revealed that *C. oleaginosus* accumulated approximately 40-50% of its dry weight as lipids using several aromatic compounds as sole carbon sources. These data support use of yeasts in the genus *Cutaneotrichosporon*, *Cutaneotrichosporon oleaginosus* in particular, for aromatic conversion to oleochemicals.

2.3 Materials and Methods

2.3.1 Materials

2.3.2 Cell Culture

Thirty-six oleaginous ascomycete and basidiomycete yeasts were obtained from the Phaff Yeast Culture Collection (Table 1) at the University of California Davis. These strains were selected based on results of the previous analysis of lipids accumulated under various conditions, including assimilation of several carbon compounds,

though not aromatics [28, 45, 46]. All yeasts were pre-cultured overnight in 2 mL YPD (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose). Cells were washed three times with YSC buffered with 50 mM phosphate buffer (pH 6.5), containing no carbon source. Cell washing entailed centrifuging cells at 1,100 × g for 4 min at 4 °C, decanting supernatant, and resuspending in 50 mM phosphate buffered YSC (pH 6.5) media. YSC was made according to manufacturer formulation (6.7 g/L DB Difco™ Yeast Nitrogen Base (YNB) without Amino Acids, 0.69 g/L CSM-Leu (Sunrise Scientific cat. #1005–010), 100 mg/L leucine) and then buffered with 0.5 M potassium phosphate buffer (pH 6.5) to a final concentration of 50 mM. Carbon sources included glucose, phenol, resorcinol, *p*-coumaric acid, *p*-hydroxybenzoic acid (pHBA), ferulic acid, and syringic acid. The pH of all media was adjusted to 6.5 using potassium hydroxide. Wells of a 48-well plate (Nunclon® cat. #150687) were filled with 250 μL of media and inoculated to a starting OD of 0.3. Plates were sealed with Breathe-Easy breathable film (Diveresified Biotech cat. #BEM-1). Plates were incubated in a VWR/Shel Lab 1565 T General Purpose Incubator on a VWR Micro Plate Shaker 120 V at 28 °C for 72 hr. Absorbance at 600 nm was measured every 6 hr using a Biotek® Epoch™ 2 microplate reader.

Selected strains were chosen for further analysis in flasks. As before, cells were pre-cultured overnight in 2 mL YPD. Cells were washed as before with YSC buffered with 50 mM phosphate buffer, pH 6.5, containing no carbon source. Washed cells were used to inoculate 15 mL of low nitrogen media (50 mM phosphate buffer, 1.7 g/L YNB without Amino Acids or Ammonium Sulfate, 0.69 g/L CSM-Leu, 100 mg/L leucine) to an initial OD of 0.3 in 50 mL baffled flasks. Ammonium sulfate was supplemented to achieve a C:N molar ratio of 200 across all carbon sources and carbon concentrations. Flasks were shaken at 250 rpm and 28 °C for 72 hr and monitored every 6 hr by measuring OD600 using a cuvette on a Thermo Scientific Nanodrop™

2000.

2.3.3 Dry Weight and Lipid Analysis

Cells were harvested for dry cell weight by washing 10 mL of cell culture with 20 mL of Milli-Q water three times and drying overnight at 40 °C under vacuum in aluminum pans. Dry cell weights were measured using an analytical balance. To identify and quantify lipids in cell biomass, extracted cellular lipids were transesterified to FAMEs as described previously [47] with minor modifications. Briefly, 1 mL of cell culture was harvested and centrifuged at 13,000 rpm for 3 min at 25 °C. Glyceryl triheptadecanoate was added as internal standard (100 μL of a 2 mg/mL stock in methanol) to the cell pellet before extraction. Lipids were transesterified to FAMEs with 500 μL of 0.5 N sodium methoxide followed by 30 min of vortexing at 2,000 rpm. The solution was neutralized with 40 μL sulfuric acid. FAMEs were extracted by adding 850 μL hexane followed by 20 min of vortexing at 2,000 rpm. The mixture was centrifuged for 1 min at 8,000 rpm, and 800 μL of the organic layer was collected for GC-FID analysis using an Agilent 7890B system.

2.3.4 Substrate Utilization

Aromatic substrate utilization was studied as previously reported [48]. Briefly, a Waters 600E high-performance liquid chromatography (HPLC) system was used in conjunction with a BioRad Fast Acid Analysis HPLC column and Waters 996 PDA detector. The running buffer was 10% v/v acetonitrile and 0.01 N H_2SO_4 in a 1:1 mixture, run at a flow rate 0.6 mL/min and at 65 °C. Wavelengths for various substrates were as follows: phenol at 270 nm, *p*-coumarate at 325 nm, resorcinol at 274 nm, and pHBA at 254 nm. Concentrations were calculated from standard curves

created for each carbon source in the appropriate medium.

2.3.5 Phylogenetic Analysis

The partial large ribosomal subunit (26S) sequence of *C. oleaginosus* was accessed from GenBank. The remaining sequences were generated using Phaff collection yeast strains. GenBank accession numbers for all species are listed in Fig. 2.1 and Table 2.1. Phylogenetic distances of the strains used in this study were based on a parsimony analysis on the 26S ribosomal rRNA sequences using an online server called Phylogeny.fr [49]. Briefly, alignment was performed by MUSCLE, curation with Gblocks, and the phylogenetic map was constructed using PhyML. All settings were left as default and Shimodaira-Hasegawa (SH)-like approximate likelihood ratio tests provided statistical tests for branch support. The output Newick file was input to an online map viewer called EvolView v2 [50]. Bootstrap scores and branch lengths were not shown for visibility.

2.4 Results

2.4.1 Aromatic screen to identify aromatic metabolizing yeast

Figure 2.1 summarizes the data of all thirty-six yeasts grown with phenol, resorcinol, *p*-coumarate, pHBA, syringate, ferulate, and glucose as sole carbon sources. The first seven columns of the figure categorize maximum A600 reached from cells grown with each carbon source while the far right column marks presence (patterend black boxes) or absence (open white boxes) of a draft genome on GenBank. Accession numbers of draft genomes and the 26S sequences used to generate the phylogenetic tree, as well as other organism information, are listed in Table 2.1. Where indicated

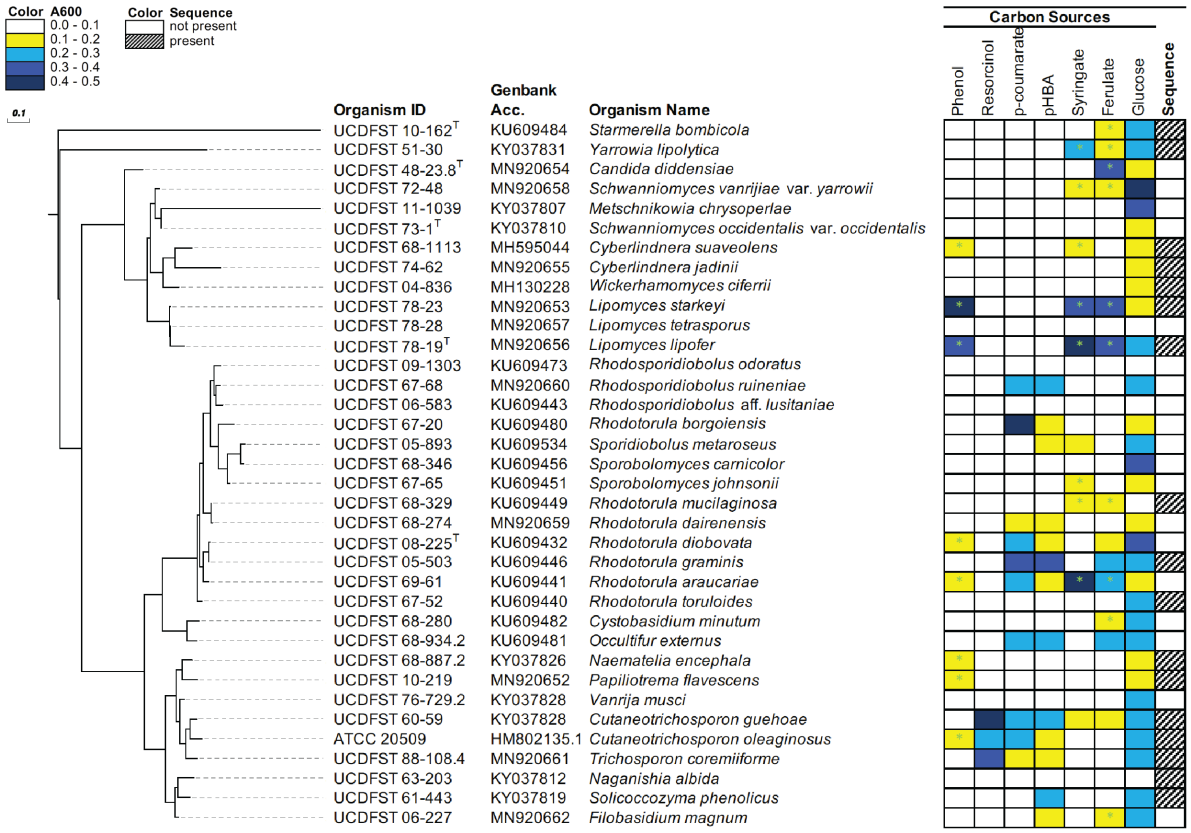


Figure 2.1. Phylogenetic tree sorting of yeast used in this chapter. Samples include Genbank Accession numbers. The first 7 columns of the heat map represent ranking of A600. The last column indicates presence (black dashed box) or absence (open white box) of a draft genome in Genbank.

in Table 2.1, ^{T} indicates type strain and - indicates data not available. Boxes with an asterick indicate maximum A600 values achieved by cells grown with 1 g/L carbon source, rather than 2 g/L.

Table 2.1. Yeast strains used in this chapter.

UCDFST strain ID	**Species**	**Genbank accession (Genome)**	**Assembly level**	**Genbank accession (26S)**
Phylum: Ascomycota				
48-23.8^{T}	*Candida diddensiae*	-	-	MN920654
74-62	*Cyberlindnera jadinii*	GCA_001245095.1	Scaffold	MN920655
68-1113	*Cyberlindnera suaveolens*	GCA_003709245.2	Scaffold	MH595044
78-19^{T}	*Lipomyces lipofer*	GCA_003705915.1	Scaffold	MN920656
78-23	*Lipomyces starkeyi*	GCA_001661325.1	Scaffold	MN920653
78-28	*Lipomyces tetrasporus*	-	-	MN920657
11-1039	*Metschnikowia chrysoperlae*	-	-	KY037807
72-48	*Schwanniomyces vanrijiae* var. *yarrowii*	-	-	MN920658
73-1^{T}	*Schwanniomyces occidentalis* var. *occidentalis*	-	-	KY037810
10-162^{T}	*Starmerella bombicola*	GCA_004124885.1	Contig	KU609484

Continued on next page

Table 2.1 – *Continued from previous page*

UCDFST strain ID	**Species**	**Genbank accession (Genome)**	**Assebmly level**	**Genbank accession (26S)**
04-836	*Wickerhamomyces ciferrii*	GCA_000313485.1	Contig	MH130228
51-30	*Yarrowia lipolytica*	GCA_000002525.1	Chromosome	KY037831
Phylum: Basidiomycota				
60-59	*Cutaneotrichosporon guehoae*	GCA_001600415.1	Scaffold	KY037828
-	*Cutaneotrichosporon oleaginosus*	GCA_001027345.1	Scaffold	HM802135.1
68-227	*Filobasidium magnum*	-	-	MN920662
68-887.2	*Naematelia encephala*	GCA_002105065.1	Contig	KY037826
63-203	*Naganishia albida*	GCA_001599735.1	Scaffold	KY037812
68-934.2	*Occultifur externus*	-	-	KU609481
10-219	*Papiliotrema flavescens*	GCA_000442785.1	Scaffold	MN920652
67-20	*Rhodotorula bogoriensis*	-	-	KU609480
69-61	*Rhodotorula araucariae*	-	-	KU609441
68-274	*Rhodotorula dairenensis*	-	-	MN920659

Continued on next page

Table 2.1 – *Continued from previous page*

UCDFST strain ID	**Species**	**Genbank accession (Genome)**	**Assebmly level**	**Genbank accession (26S)**
08-225^{T}	*Rhodotorula diobovata*	-	-	KU609432
05-503	*Rhodotorula graminis*	GCA_001329695.1	Scaffold	KU609446
68-329	*Rhodotorula mucilaginosa*	GCA_002806785.1	Scaffold	KU609449
67-52	*Rhodotorula toruloides*	GCA_000222205.2	Scaffold	KU609440
06-583	*Rhodosporidiobolus* aff. *lusitaniae*	-	-	KU609443
09-1303	*Rhodosporidiobolus odoratus*	-	-	KU609473
67-68	*Rhodosporidiobolus ruineniae*	-	-	MN920660
61-443	*Solicoccozyma phenolicus*	GCA_001600015.1	Scaffold	KY037819
05-893	*Sporidiobolus metaroseus*	-	-	KU609534
68-346	*Sporobolomyces carnicolor*	-	-	KU609456
67-65	*Sporobolomyces johnsonii*	-	-	KU609451
88.108.4	*Trichosporon coremiiforme*	GCA_001752605.1	Scaffold	MN920661
76.729.2	*Vanrija musci*	-	-	KY037828

Of the yeasts studied, 58% of the ascomycete yeasts were able to grow on aromatic carbon sources and 79% of the basidiomycete yeasts were able to grow. Importantly, none of the ascomycete yeasts were able to grow in concentrations greater than 1 g/L. Metabolism of H-lignin compounds was exclusively segregated to basidiomycetes. Ten yeast species grew in media containing 2 g/L *p*-coumaric acid. Most of the yeasts belonged to genera *Rhodotorula*, *Rhodosporidiobolus*, *Trichosporon*, or *Cutaneotrichosporon*. The only yeast not in these genera was *Occultifur externus*. Thirteen yeasts grew in 2 g/L pHBA. Eight of these species were in genera *Rhodotorula*, *Rhodosporidiobolous*, *Trichosporon*, and *Cutaneotrichosporon*. The four not in these were *Filobasidium magnum*, *Occultifur externus*, *Solicoccozyma phenolicus*, and *Sporidiobolus metaroseus*, the latter of which is classified in order Sporidiobolales [51].

None of the yeast grew in media containing 2 g/L phenol, which is known to be toxic to most microbes; however, during a subsequent experiment, eight yeast grew in media containing 1 g/L phenol: *Cutaneotrichosporon oleaginosus*, *Cyberlindnera suaveolens*, *Lipomyces lipofer*, *Lipomyces starkeyi*, *Naematelia encephala*, *Papiliotrema flavescens*, *Rhodotorula araucariae*, and *Rhodotorula diobovata*. Interestingly, results show phenol metabolism is more divergently found across ascomycetes and basidiomycetes than *p*-coumarate and pHBA. Metabolism of 2 g/L of the aromatic diol resorcinol, on the other hand, was very narrowly found in two closely related genera, *Trichosporon* and *Cutaneotrichosporon*.

Only two yeasts, *Cutaneotrichosporon guehoae* and *Sporidiobolus metaroseus*, were able to grow on 2 g/L syringate. When the syringate concentration was dropped to 1 g/L, eight more yeasts were able to grow, including *Yarrowia lipolytica*, *Schwanniomyces vanrijiae* var. *yarrowii*, *Cyberlindnera suaveolens*, *Lipomyces starkeyi*, *Lipomyces lipofer*, *Sporobolomyces johnsonii*, *Rhodotorula mucilaginosa*, and *Rhodotorula arau-*

cariae.

Only four yeasts, *Rhodotorula diobovata*, *Rhodotorula graminis*, *Occultifur externus*, and *Cutaneotrichosporon guehoae* were able to grow in media containing 2 g/L ferulate. Ten more yeasts were able to grow when ferulate concentration was lowered to 1 g/L, though many did not grow to an absorbance higher than 0.1. These yeast are *Starmarella bombicola*, *Yarrowia lipolytica*, *Candida diddensiae*, *Schwanniomyces vanrijiae* var. *yarrowii*, *Lipomyces starkeyi*, *Lipomyces lipofer*, *Rhodotorula mucilaginosa*, *Rhodotorula araucariae*, *Cystobasidium minutum*, and *Filobasidium magnum*. Unlike H- lignin monomer, the metabolizers of G- and S- lignin monomers were split roughly evenly between ascomycetes and basidiomycetes, but growth on compounds at 2 g/L was exclusive to basidiomycetes.

Collectively, aromatic metabolism is enriched among oleaginous basidiomycetes compared to oleaginous ascomycetes, even when the concentration of substrates was lowered to 1 g/L. Many of the *Rhodotorula*, *Trichosporon*, and *Cutaneotrichosporon* species were able to grow in at least half of the aromatic compounds, even when substrate concentrations of 2 g/L were used. A single species with promising characteristics was selected to connect aromatic catabolism to an oleaginous phenotype. The availability of draft genomes of all the listed *Trichosporon* and *Cutaneotrichosporon* species was an advantage over the *Rhodotorula* species. Furthermore, these results indicated that *Cutaneotrichosporon* could grow faster in several substrates compared to *Rhodotorula* (data not shown). To consider which yeast would be best to investigate further, we considered the genetic tools available for both yeasts. The genetic transformation system for *C. guehoae* seems limited to spheroplast fusion, whereas *C. oleaginosus* has Agrobacterium-mediated transformation and electroporation developed. All three methods result in random genomic integration. Auxotrophic strains and heterologous drug resistance genes exist for both species, but neither have stable

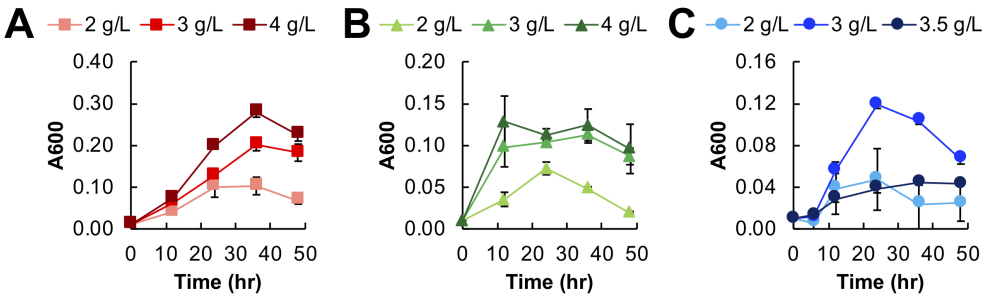


Figure 2.2. Growth charts of *C. oleaginosus* grown with increasing concentrations of A) pHBA (■), B) *p*-coumarate (▲), C) resorcinol (●).

plasmid expression [48, 52–56]. However, heterologous expression of several proteins for enhanced TAG production and novel lipid synthesis was performed in *C. oleaginosus*, demonstrating practical use of a more advanced genetic toolkit [52, 53]. As such, we continued with *C. oleaginosus* for the remainder of the experiments in this chapter.

2.4.2 *Cutaneotrichosporon oleaginosus* remains oleaginous when grown on several aromatics.

While all previous growth assays were conducted in media containing up to 2 g/L of substrate, higher concentrations were tested to determine the maximum concentration cells could tolerate before growth was inhibited (Fig. 3.2). Cells were grown in 48-well plates over 48 hr in phosphate-buffered YSC media. Cells were able to grow with no detriment in 4 g/L of both pHBA and *p*-coumarate. Higher concentrations of pHBA and *p*-coumarate could not be tested due to solubility limitations. Cells were able to grow in up to 3.5 g/L resorcinol; however, the cells showed an increased lag phase and decreased growth rate.

Therefore, in shake flask experiments, 4 g/L pHBA, 4 g/L *p*-coumarate, 3 g/L

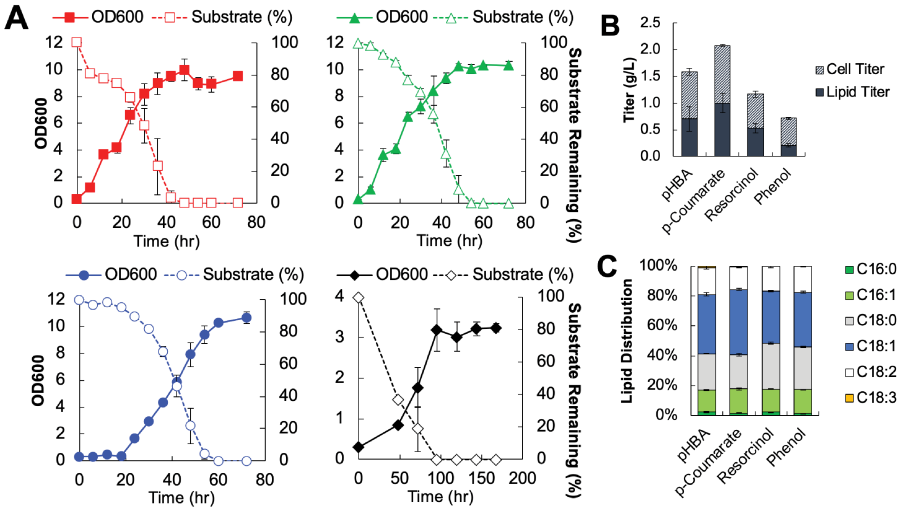


Figure 2.3. Flask growth of cells grown in low nitrogen media. A) OD600 (left axis, solid markers) and substrate utilization (pen markers, dashed line) over 72 hours for pHBA (■), *p*-coumarate (▲), resorcinol (●), and phenol (◆), B) biomass (dashed) and lipid titer (solid), C) lipid profile distribution.

resorcinol, and 1 g/L phenol were used in a phosphate-buffered YSC media, pH 6.5. To induce lipogenesis, a molar carbon to nitrogen (C:N) ratio of 200 was used for all substrates. All aromatic compounds were fully metabolized by 72 hr, as determined by HPLC–UV-vis (Fig. 2.3A). All media types and carbon concentrations reached the same final OD600 of about 10, except for cells grown with phenol, which reached OD600 just over 3.0. The highest biomass resulted from cells grown with 4 g/L *p*-coumarate, followed by cells grown with 4 g/L pHBA, 3 g/L resorcinol, and 1 g/L phenol, in that order (Fig. 2.3B; Table 2). Surprisingly, growth on *p*-coumarate resulted in 0.5 g/L more cells than cells grown on pHBA. Despite differences in final biomass titer achieved, all cells grown with each aromatic substrate remained highly oleaginous at nearly 50% lipid accumulation on a dry biomass basis, except for phenol-grown cells (Table 2.2). Cells grown on phenol reached 29.2% lipid on a biomass basis, still retaining oleaginous behavior.

The majority of the fatty acid was oleic acid for cells grown on each substrate.

Table 2.2. Measurement of dry cell weight, lipid titer, and percentage lipid accumulation in *C. oleaginosus* cells grown in flask studies.

Substrate	**Biomass titer (g/L)**	**Lipid Titer (g/L)**	**Lipid Accumulation (%)**
pHBA	1.58±0.06	0.70±0.23	44.9±16.2
p-coumarate	2.08±0.02	1.00±0.18	48.2±9.0
Resorcinol	1.17±0.06	0.53±0.09	45.6±8.9
Phenol	0.72±0.02	0.21±0.03	29.2±5.4

Table 2.3. Lipid profile of *C. oleaginosus* cells grown in flask studies.

Substrate	**C16:0**	**C16:1**	**C18:0**	**C18:1**	**C18:2**	**C18:3**
pHBA	2.3±0.5	14.7±0.3	24.4±0.2	39.7±1.2	17.9±1.0	1.1±0.1
p-coumarate	1.6±0.2	16.3±0.7	22.9±1.1	43.7±0.7	14.9±0.3	0.8±0.0
Resorcinol	2.3±0.1	15.4±0.1	30.8±0.7	34.9±0.4	16.6±1.0	0.0±0.0
Phenol	1.4±0.1	16.1±0.2	28.5±0.3	36.7±0.5	17.5±0.8	0.0±0.0

Both pHBA and *p*-coumaric acid resulted in a linoleic component of approximately 1% of the total fatty acid profile, but cells grown on resorcinol and phenol accumulated oleic acid rather than linoleic acid (Fig. 2.3C; Table 2.3).

2.5 Discussion

Overall, a higher percentage of basidiomycete yeasts than ascomycetes are able to grow in lignin-derived monoaromatic compounds as sole carbon sources. This is in agreeance with Sampaio, et al [57], who screened 227 yeast strains isolated from terrestrial and aquatic habitats on guaiacol, protocatechuic acid, vanillic acid, ferulic acid, and veratric acid as sole carbon sources. Not all of the yeasts screened were oleaginous yeasts, as they were in this chapter; however, a number of them fell similar clades examined in this study, such as *Candida*, *Cryptococcus*, *Rhodosporidium*, *Rhodotorula*, and *Trichosporon* species. Mills et al. screened 68 yeasts from six genera

isolated from sewage against 15 hydroxy derivatives of phenol and benzoic acid as sole carbon sources at 1 g/L and concluded no obvious relationship between utilization of the compounds and phylogentic classification [58]. However, only two ascomycete yeast species were tested, one of which was not oleaginous, and the rest were basidiomycetes of the *Candida*, *Cryptococcus*, *Rhodotorula*, and *Trichosporon* genera. We notice that the choice of carbon source varies the prevalence of ascomycete versus basidiomycete yeasts among successful growers, in agreeance with Sampaio's conclusions.

Of the yeasts capable of aromatic catabolism in this chapter, some had previously been reported by others to be capable of aromatic catabolism. A study of *Occultifur externus* sp. nov. reports carbon assimilation of a number of aromatic compounds, including 3-hydroxybenzic acid, pHBA, vanillyl alcohol, vanillic acid, ferulic acid, protocatechuate, *p*-coumaric acid, and caffeic acid, though a much lower concentrations of 0.1% (w/v) was used to avoid toxicity [59]. *Lipomyces starkeyi* ATCC 56304 has been shown to remove 62% of the phenolic content of diluted palm oil mill effluent, but cells only accumulated 21.3% lipid content [60]. In another study of olive oil mill effluent, *L. starkeyi* reduced phenolic content by approximately 50% and accumulated just under 30% of its biomass as lipids [61]. In both studies, the oil mill wastes also included sugars, so lipid content cannot be attributed to purely phenolic catabolism, nor do these works necessarily support catabolism of phenolics into central metabolites. While also not sole carbon sources, *L. starkeyi* NBRC10381 cells were shown to uptake vanillin and syringaldehyde present at concentrations of approximately 0.1 mM when mixed in a synthetic lignocellulosic hydrolysate blend. Because the synthetic hydrolysate had multiple carbon sources in it, lipid accumulation cannot be attributed to just the phenolics, but cells were highly oleaginous at 68% despite the presence of inhibitors [62].

An extensive study by Sánchez i Nogué et al. screened 31 oleaginous yeasts for their ability to grow on corn stover hydrolysate and determined *C. oleaginosus* ATCC 20509, *R. toruloides* DSM-4444, and *C. guehoae* UCDFST 60–59 were top performers. These three yeasts were screened for their growth utilizing 20 mM ferulic acid, vanillic acid, *p*-coumaric acid, and pHBA as sole carbon sources in YNB. *Cutaneotrichosporon oleaginosus* and *C. guehoae* were able to metabolize pHBA fully but did not grow when cultured with the other compounds. *Rhodotorula toruloides* was able to fully metabolize all four compounds [40]. These results agree with Yaegashi et al. who analyzed *R. toruloides* strain IFO 0880, the same strain used in the present study [63]. Differences between the studies may be attributed to media variations, as a potassium phosphate-buffered synthetic complete media was used in the present study, Sanchez i Nogué et al. used a potassium hydrogen phthalate buffered YNB media, and Yaegashi et al. used an unbuffered synthetic complete media. It is acknowledged that media choice may impact these results. YSC was chosen in this study due to its defined formulation and general use across a variety of yeast species. The ionization of the acidic substrates (syringate, ferulate, *p*-coumarate, and pHBA) motivated the use of a buffer. However, it is possible that some yeasts will grow better in media specifically formulated for its nutritional requirements. Due to the number of yeasts in the present study, the initial down-selection was intentionally based on a single media formulation. The results conclusively show which yeast appear to grow on the various aromatic substrates as a sole carbon source, but does not rule out the possibility that others could grow on aromatics if provided the right nutrients.

Some species formerly classified in the genus *Trichosporon* have been reclassified as *Cutaneotrichosporon* [51]. In this discussion, they will be referred to as *Cutaneotrichosporon*. *Cutaneotrichosporon cutaneum* ATCC 20271 is a closely related yeast to *C. oleaginosus* that is well described for tolerance and metabolism of

several lignocellulosic inhibitors, including phenolics. *Cutaneotrichosporon cutaneum* R57 was able to assimilate hydroxyquinone, resorcinol, phenol, 2,6-dinitrophenol, and *p*-cresol at 0.2 g/L and 3-nitrophenol and 3-chlorophenol at 0.1 g/L as sole carbon sources. Lipid titers were not provided, but degradation kinetics were described [54]. A strain isolated from a soil suspension, collected from olive tree fields irrigated by olive mill wastewater for three successive years, was able to grow in media containing phenol as the single carbon source up to 2 g/L, though lipid profiles are also not reported [64]. Phenolic aldehydes are particularly toxic as compared to the alcohol versions, but *C. cutaneum* suffered negligible growth defects when grown with 2.5 g/L syringaldehyde as the sole carbon source. 4-hydroxybenzaldehyde at 1.5 g/L resulted in significant growth defects and vanillin only partially inhibited cell growth at 2.0 g/L [65]. The lipid titers and profiles were again not reported. A separate study grew cells in batch cultures with the same three compounds and found only 4-hydroxybenzaldehyde resulted in oleaginous behavior even when concentrations reached 2 g/L. A 120 hr fed-batch fermentation with 4-hydroxybenzaldehyde resulted in 39% accumulation of biomass as lipids [66]. Lipid profiles were not provided. Mills et al. studied 68 different yeast species and six of the ten *C. cutaneum* species grew on the widest variety of substrates supplied as sole carbon sources at 1 g/L, including phenol, catechol, hydroquinone, hydroxyquinol, phloroglucinol, benzoic acid, salicylic acid, 3-hydroxybenzoic acid, 4-hydroxybenzoic acid, protocatechuic acid, gentisic acid, 2,4-dihydroxybenzoic acid, 2,6-dihydroxybenzoic acid, and 3,5-dihydroxybenzoic acid [58]. *Cutaneotrichosporon guehoae* CBS 8521 type strain assimilates phenol, hydroquinone, resorcinol, phloroglucinol, *m*-cresol, 4-methylcatechol, salicylate, 2,3-hydroxybenzoate, 4-hydroxyacetophenone, 3-hydroxycinnamate, and 4-hydroxy-cinnamate [67]. Lipids titers or accumulation are not reported. *Cutaneotrichosporon oleaginosus* ATCC 20509 was previously reported as being able to metabolize phenol, pHBA, and

resorcinol as sole carbon sources. A similar fed-batch study was performed, in which cells were grown with resorcinol over 156 hr and accumulated 69.5% of yeast cell biomass as lipids. As far as I am aware, this value remains the highest reported value for lipid accumulation derived a non-engineered microbe using aromatic compounds as sole carbon sources to date. Fifty percent of the accumulated lipids were C18:1 [48]. In this study, *C. oleaginosus* demonstrated oleaginous behavior with every aromatic compound under batch culturing conditions, although cells cultivated in phenol accumulated less intracellular lipid than in resorcinol, pHBA, and *p*-coumarate. A previous name for *C. oleaginosus* is *Cryptococcus curvatus*. *Cryptococcus psychrotolerans* IITRFD, now called *Vishniacozyma psychrotolerans*, was found to metabolize a range of aromatic hydrocarbons, such as phenol, anthracene, naphthalene, and pyrene, although it was in media containing 4% glucose. Because the latter three compounds are petroleum-based aromatic hydrocarbons, we will not discuss them and instead highlight their reports of phenol metabolism, since the authors demonstrate clear utilization of the phenol. Over the course of 144 hr, *C. psychrotolerans* was capable of accumulating 46.54% lipids per dry cell weight for a final titer of 6.4 g/L and productivity of 0.0444 g/L/h. Lipid accumulation, titer, and productivities are higher than that of the cells grown with glucose-only containing media. Lipid profiles show a near-even split across C16:0, C18:1, and C18:2 fatty acids, a similar profile to the glucose-only control [68].

Metabolism of lignocellulosic hydrolysates and tolerance and/or assimilation of lignin phenolic monomers has been studied in the *Rhodotorula* genera; this discussion focuses on assimilation rather than tolerance. *Rhodotorula* sp. ZM1, isolated from an acid mine drainage, was found to metabolize phenol in acidic conditions. While a lag phase was present, ZM1 cells could metabolize up to 1.3 g/L of phenol as the sole carbon source by 120 hr. Su et al. describe predicted *ortho*-cleavage pathways

based on gene predictions performed on their whole-genome sequencing and functional annotation. While many *Rhodotorula* strains are oleaginous, Su et al. did not examine lipid production [69]. Jarboui et al. found that a *R. mucilaginosa* strain isolated from olive mill wastewater was shown to grow in and fully metabolize protocatechuic acid, *p*-coumaric acid, and vanillic acid as sole carbon sources at 1 g/L, though lipid titers were not reported. It was noted this strain behaved differently from other reported *R. mucilaginosa* strains, which did not metabolize aromatics [70]. In another study, a different *R. mucilaginosa* strain, isolated from top soil, metabolized *p*-coumaric acid and ferulic acid present in a base-catalyzed depolymerized (BCD) lignin liquor at approximately 7 g/L and 0.5 g/L, respectively. Lipid titers were not reported [18]. Rodriguez et al. studied a *R. mucilaginosa* J31 and *R. toruloides* IFO0880 strains ability to metabolize pHBA, protocatechuic acid, vanillic acid, ferulic acid, *p*-coumaric acid, and benzoic acid as single carbon sources at a concentration of 1 g/L. The authors also show both cell lines completely metabolize protocatechuic acid and pHBA present in hydrolysates derived from two engineered lines of *Arabidopsis thaliana*. Generally, *R. mucilaginosa* grew to higher biomass than *R. toruloides*, except when grown with ferulic acid. No growth was seen for vanillin or 4-hydroxybenzaldehyde as sole carbon sources for either species [71]. Lipid titers were not reported, as the strains were engineered to produce bisaboline, similarly to Yaegashi et al. [63]. In the present study, *R. mucilaginosa* grew poorly in syringate and ferulate when the concentration was less than 1 g/L. This emphasizes a recurring concept: that different strains of the same species often have very different phenotypes such as metabolic capabilities.

Toxicity mechanisms are difficult to probe, but some studies observed an increase of unsaturated bonds in fatty acids in response to aromatic compounds. An increase of C18:1 was observed when *C. fermentans* CICC 1368 was grown with increasing concentrations of vanillyl alcohol in the presence of a glucose/xy-

lose blend, though the C18:1 increase was not as dramatic with hydroquinone or catechol [72]. *Rhodotorula kratochvilovae* (formerly classified as *Rhodosporidium kratochvilovae*) HIMPA1 exhibited an increase of monounsaturated fatty acids from 51.87 to 62.98% with increasing phenol concentrations, presumably due to the conversion of C18:0–C18:1 by a Δ-9 desaturase [73]. The same increase of C18:1 content of cells grown in resorcinol, pHBA, and phenol relative to cells grown in glucose was observed in *C. oleaginosus* cells [48]. Since *C. oleaginosus* accumulates high intracellular lipids when cultivated in phenolic conditions, and the lipids produced have a significant amount of unsaturated C18:1 fatty acids, this yeast may be particularly suited to overcome aromatic toxicity.

2.6 Conclusions

Significant effort has been made to engineer model organisms, such as *Saccharomyces cerevisiae* and *Escherichia coli*, for aromatic conversion to industrially relevant products [21–24, 74, 75]. Engineering new substrate metabolism is exceedingly difficult since this complex phenotype is directly tied to central metabolism. An increasingly popular approach is to identify microbes that inherently exhibit more of the desired properties and develop the tools to engineer them, as described in several recent reviews and opinions [39, 76, 77]. There is significant variation in metabolic capabilities between species, and among independently isolated strains of the same species; however, building biodiversity collections has become complicated due to habitat loss and the Nagoya Protocol [78, 79]. Therefore, it is important to preserve microorganisms in professionally managed culture collections such as the Phaff Yeast Culture Collection, to ensure that these microorganisms continue to be available for future studies [80, 81].

Although yeast genomes are being sequenced at a rapid pace, functional annotation is lagging. Genomes of bacteria, such as *Rhodococcus opacus* PD630 and *Pseudomonas putida*, have been more thoroughly annotated, but the low identity to eukaryotic genes yields few successful BLAST homology results. Therefore, functional data are required to even begin to explore aromatic metabolism in yeast. This chapter, which identified promising aromatic metabolizing yeasts, will aid in identifying the functional genes responsible for aromatic metabolism in yeast. The major findings of this study include identification of over a dozen potential aromatic metabolizing oleaginous yeast species and a better understanding of the phylogenetic distribution of aromatic metabolism pathway branches.

2.7 Acknowledgments

This work is adapted from Ref. 82 and is reprinted with permissions from Springer according to license number 4916761217613.

AY, NF, KN, and SL performed yeast growth experiments and lipid quantification. IS and KBM provided strains from the UC Davis Phaff collection and assisted in curating GenBank 26S sequences for each yeast. AY performed phylogenetic analysis. AY, KBM, and MB conceived experimental design and wrote and edited the manuscript.

This material is based upon work supported by a USDA Sun Grant (2014-38502-22598) and Creative Inquiry funds from Clemson University. AY has been supported by a GAANN Fellowship (P200A180076) from the US Department of Education. This material is based upon work supported by the US Department of Agriculture, National Institute of Food Agriculture under Award No. 2018-67009-27901. Any opinions, findings, and conclusions or recommendations expressed in this

publication are those of the author(s) and do not necessarily reflect the views of the US Department of Agriculture.

Bibliography

[1] Davis, R.; Tao, L.; Tan, E., *et al.* In *Process Design and Economics for the Conversion of Lignocellulosic Biomass to Hydrocarbons: Dilute-Acid and Enzymatic Deconstruction of Biomass to Sugars and Biological Conversion of Sugars to Hydrocarbons*. NREL, **2013**.

[2] Ibarra, M.; Polk, S.; K., B. In *Biomass for biopower : feedstock supply assessments.*. Department of Energy, **2012**.

[3] ??? In *Biorefinery Optimization Workshop Summary Report.*. Department of Energy BTO, **2016**.

[4] Holladay, J.; White, J.; Bozzel, J., *et al.* Top Value-Added Chemicals from Biomass. *PNNL-16983*, **2007**.

[5] Alinejad, M.; Henry, C.; Nikafshar, S., *et al.* Lignin-Based Polyurethanes: Opportunities for Bio-Based Foams, Elastomers, Coatings and Adhesives. *Polymers (Basel)*, **2019**, 11(7).

[6] Kalami, S.; Arefmanesh, M.; Master, E., *et al.* Replacing 100

[7] Jin, J.; Ding, J.; Klett, A., *et al.* Carbon Fibers Derived from Fractionated–Solvated Lignin Precursors for Enhanced Mechanical Performance. *ACS Sustainable Chem Eng*, **2018**, 6(11), 14135–14142.

[8] Klett, A.S.; Chappell, P.V.; Thies, M.C. Recovering ultraclean lignins of controlled molecular weight from Kraft black-liquor lignins. *Chemical Communications*, **2015**, 51(64), 12855–12858.

[9] Gosselink, R.JA.; Jong, E.; Guran, B., *et al.* Co-ordination network for lignin - standardisation, production and applications adapted to market requirements (EUROLIGNIN). *Ind Crop Prod*, **2004**, 20(2), 121–129.

[10] Ragauskas, A.J.; Beckham, G.T.; Biddy, M.J., *et al.* Lignin valorization: improving lignin processing in the biorefinery. *Science*, **2014**, 344(6185), 1246843.

[11] Xu, C.P.; Arancon, R.AD.; Labidi, J., *et al.* Lignin depolymerisation strategies: towards valuable chemicals and fuels. *Chemical Society Reviews*, **2014**, 43(22), 7485–7500.

[12] Wang, X.Y.; Rinaldi, R. Solvent Effects on the Hydrogenolysis of Diphenyl Ether with Raney Nickel and their Implications for the Conversion of Lignin. *Chemsuschem*, **2012**, 5(8), 1455–1466.

[13] Sturgeon, M.R.; O'Brien, M.H.; Ciesielski, P.N., *et al.* Lignin depolymerisation by nickel supported layered-double hydroxide catalysts. *Green Chemistry*, **2014**, 16(2), 824–835.

[14] Lubbers, R.; Dilokpimol, A.; Visser, J., *et al.* A comparison between the homocyclic aromatic metabolic pathways from plant-derived compounds by bacteria and fungi. *Biotech Adv*, **2019**, 37, 107396.

[15] Henson, W.R.; Campbell, T.; DeLorenzo, D.M., *et al.* Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus. *Metab Eng*, **2018**, 49, 69–83.

[16] Beckham, G.T.; Johnson, C.W.; Karp, E.M., *et al.* Opportunities and challenges in biological lignin valorization. *Curr Opin Biotechnol*, **2016**, 42, 40–53.

[17] Salvachúa, D.; Katahira, R.; Cleveland, N.S., *et al.* Lignin depolymerization by fungal secretomes and a microbial sink. *Green Chemistry*, **2016**, 18(22), 6046–6062.

[18] Rodriguez, A.; Salvachua, D.; Katahira, R., *et al.* Base-catalyzed depolymerization of solid lignin-rich streams enables microbial conversion. *ACS Sustain Chem Eng*, **2017**, 5, 8171–8180.

[19] Xie, S.; Sun, Q.; Pu, Y., *et al.* Advanced Chemical Design for Efficient Lignin Bioconversion. *ACS Sustainable Chemistry and Engineering*, **2017**, 5(3), 2215–2223.

[20] Liu, Z.-H.; Le, R.K.; Kosa, M., *et al.* Identifying and creating pathways to improve biological lignin valorization. *Renewable and Sustainable Energy Reviews*, **2019**, 105, 349–362.

[21] Salvachua, D.; Rydzak, T.; Auwae, R., *et al.* Metabolic engineering of Pseudomonas putida for increased polyhydroxyalkanoate production from lignin. *Microb Biotechnol*, **2019**, 290.

[22] Xu, Z.; Lei, P.; Zhai, R., *et al.* Recent advances in lignin valorization with bacterial cultures: microorganisms, metabolic pathways, and bio-products. *Biotechnol Biofuels*, **2019**, 12, 32.

[23] Vardon, D.R.; Franden, M.A.; Johnson, C.W., *et al.* Adipic acid production from lignin. *Energy Environ Sci*, **2015**, 8, 617–628.

[24] Sonoki, T.; Takahashi, K.; Sugita, H., *et al.* Glucose-free cis, cis-muconic acid production via new metabolic designs corresponding to the heterogeneity of lignin. *ACS Sustain Chem Eng*, **2017**, 6, 1256–1264.

[25] Jolivalt, C.; Madzak, C.; Brault, A., *et al.* Expression of laccase IIIb from the white-rot fungus Trametes versicolor in the yeast Yarrowia lipolytica for environmental applications. *Appl Microbiol Biotechnol*, **2005**, 66(4), 450–456.

[26] Li, Q.; Pei, J.; Zhao, L., *et al.* Overexpression and characterization of laccase from Trametes versicolor in Pichia pastoris. *Applied Biochemistry and Microbiology*, **2014**, 50(2), 140–147.

[27] Piscitelli, A.; Giardina, P.; Mazzoni, C., *et al.* Recombinant expression of Pleurotus ostreatus laccases in Kluyveromyces lactis and Saccharomyces cerevisiae. *Appl Microbiol Biotechnol*, **2005**, 69(4), 428–439.

[28] Garay, L.; Sitepu, I.; Cajka, T., *et al.* Eighteen new oleaginous yeast species. *JIMB*, **2016**, 43, 887–900.

[29] Sitepu, I.; Garay, L.; Sestric, R., *et al.* Oleaginous yeasts for biodiesel: Current and future trends in biology and production. *Journal of Biotech Adv*, **2014**, 32, 1336–1360.

[30] Papanikolaou, S.; Aggelis, G. Lipids of oleaginous yeasts. Part II: Technology and potential applications. *European Journal of Lipid Science and Technology*, **2011**, 113(8), 1052–1073.

[31] Probst, K.V.; Schulte, L.R.; Durrett, T.P., *et al.* Oleaginous yeast: a value-added platform for renewable oils. *Critical reviews in biotechnology*, **2016**, 36(5), 942–955.

[32] Alexieva, Z.; Gerginova, M.; Zlateva, P., *et al.* Monitoring of aromatic pollutants biodegradation. *Biochemical Engineering Journal*, **2008**, 40(2), 233–240.

[33] Anderson, J.J.; Dagley, S. Catabolism of Aromatic-Acids in Trichosporon-Cutaneum. *J Bacteriol*, **1980**, 141(2), 534–543.

[34] DeRito, C.M.; Madsen, E.L. Stable isotope probing reveals Trichosporon yeast to be active in situ in soil phenol metabolism. *ISME J*, **2009**, 3(4), 477–485.

[35] Gaal, A.; Neujahr, H.Y. Metabolism of Phenol and Resorcinol in Trichosporon-Cutaneum. *J Bacteriol*, **1979**, 137(1), 13–21.

[36] Powlowski, J.B.; Dagley, S. beta-ketoadipate Pathway in Trichosporon cutaneum Modified for Methyl-Substituted Metabolites. *Journal of Bacteriology*, **1985**, 163(3), 1126–1135.

[37] Shoda, M.; Udaka, S. Preferential Utilization of Phenol Rather than Glucose by Trichosporon cutaneum Possessing a Partially Constitutive Catechol 1,2-Oxygenase. *Applied and Environmental Microbiology*, **1980**, 39(6), 1129–1133.

[38] Suzuki, K.; Itoh, M. Metabolism of p-Hydroxybenzoate via Hydroxyquinol by Trichosporon cutaneum WY2-2: Characterization of the Pathway using Superoxide Dismutase as a Stabilizer of Hydroxyquinol. *Plant Cell Physiology*, **1986**, 27(8), 1451–1460.

[39] Yaguchi, A.; Rives, D.; Blenner, M. The New Kids on the Block: Emerging Oleaginous Yeast of Biotechnological Importance. *AIMS Microbiology*, **2017**, 3(2), 227–247.

[40] Nogue, V.; Black, B.A.; Kruger, J.S., *et al.* Integrated diesel production from lignocellulosic sugars via oleaginous yeast. *Green Chem*, **2018**, 20, 4349–4365.

[41] Fialová, A.; Boschke, E.; Bley, T. Rapid monitoring of the biodegradation of phenol-like compounds by the yeast Candida maltosa using BOD measurements. *International Biodeterioration and Biodegradation*, **2004**, 54(1), 69–76.

[42] Gerecova, G.; Nebohacova, M.; Zeman, I., *et al.* Metabolic gene clusters encoding the enzymes of two branches of the 3-oxoadipate pathway in the pathogenic yeast Candida albicans. *FEMS Yeast Res*, **2015**, 15(3).

[43] Krug, M.; Ziegler, H.; Straube, G. Degradation of Phenolic Compounds by the Yeast Candida tropicalis HP 15 I. Physiology of Growth and Substrate Utilization. *Journal of Basic Microbiology*, **1985**, 25(2), 103–110.

[44] Sampio, J.; Bauer, R.; Begerow, D., *et al.* Occultifur externus sp. nov., A New Species of Simple-pored Auricularioid Heterobasidiomycete from Plant Litter in Portugal. *Mycologia*, **1999**, 91(6), 1094–1101.

[45] Sitepu, I.; Sestric, R.; Ignatia, L., *et al.* Manipulation of culture conditions alters lipid content and fatty acid profiles of a wide variety of known and new oleaginous yeasts species. *Biores Technol*, **2013**, 144, 360–369.

[46] Sitepu, I.; Boundy-Mills, K. Carbon source utilization and inhibitor tolerance of 45 oleaginous yeast species. *JIMB*, **2014**, 41, 1061–1070.

[47] Rodriguez, G.M.; Hussain, M.S.; Gambill, L., *et al.* Engineering xylose utilization in Yarrowia lipolytica by understanding its cryptic xylose pathway. *Biotechnol Biofuels*, **2016**, 9, 149.

[48] Yaguchi, A.; Robinson, A.; Mihealsick, E., *et al.* Metabolism of aromatics by Trichosporon oleaginosus while remaining oleaginous. *Microb Cell Fact*, **2017**, 16(1), 206.

[49] Dereeper, A.; Guignon, V.; Blanc, G., *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nuc Acids Res*, **2008**, 36, W465–469.

[50] He, Z.; Zhang, H.; Gao, S., *et al.* Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nuc Acids Res*, **2016**, 44, W236–241.

[51] Wang, Q.; Yurkov, A.; Goker, M., *et al.* Phylogenetic classification of yeasts and related taxa within Pucciniomycotina. *Stud Mycol*, **2015**, 81, 149–189.

[52] Görner, C.; Redai, V.; Bracharz, F., *et al.* Genetic engineering and production of modified fatty acids by the non-conventional oleaginous yeast Trichosporon oleaginosus ATCC 20509. *Green Chem.*, **2016**, 18(7), 2037–2046.

[53] Koivuranta, K.; Castillo, S.; Jouhten, P., *et al.* Enhanced triacylglycerol production with genetically modified Trichosporon oleaginosus. *Front Microbiol*, **2**, 9, 1337.

[54] Gerginova, M.; Peneva, N.; Alexieva, Z. Investigation of feasibility of gene transfer in Trichosporon cutaneum yeast strain R57. *J BioSci Biotech*, **2014**, 2014, 95–99.

[55] Ochsner, U.; Glumoff, V.; Kalin, M., *et al.* Genetic transformation of auxotrophic mutants of the filamentous yeast Trichosporon cutaneum using homologous and heterologous marker genes. *Yeast*, **1991**, 7, 513–524.

[56] Sienko, M.; Stepien, P.; Paszewski, A. Generation of genetic recombinants in Trichosporon cutaneum by spontaneous segregation of protoplast fusants. *J Gen Microbiol*, **1992**, 138, 1409.

[57] Sampaio, J. Utilization of low molecular weight lignin-related aromatic compounds for the selective isolation of yeasts: Rhodotorula vanillica, a new basidiomycetous yeast species. *Syst Appl Microbiol*, **1995**, 17, 613–619.

[58] Mills, C.; Child, J.J.; T., S.JF. The utilization of aromatic compounds by yeasts. *Antonie Van Leeuwenhoek*, **1971**, 37, 281–287.

[59] Sampaio, J.P.; Bauer, R.; Begerow, D., *et al.* Occultifur externus sp. nov., A new species of simple-pored auricularioid heterobasidiomycete from plant litter in Portugal. *Mycologia*, **1999**, 91, 1094–1101.

[60] Islam, M.A.; Yousuf, A.; Karim, A., *et al.* Bioremediation of palm oil mill effluent and lipid production by Lipomyces starkeyi: a combined approach. *J Clean Prod*, **2018**, 172, 1779–1787.

[61] Yousuf, A.; Sannino, F.; Addorisio, V., *et al.* Microbial conversion of olive oil mill wastewaters into lipids suitable for biodiesel production. *J Agric Food Chem*, **2010**, 58, 8630–8635.

[62] Juanssilfero, A.B.; Kahar, P.; Amza, R.L., *et al.* Selection of oleaginous yeasts capable of high lipid accumulation during challenges from inhibitory chemical compounds. *Biochem Eng*, **2018**, 137, 182–191.

[63] Yaegashi, J.; Kirby, J.; Ito, M., *et al.* Rhodosporidium toruloides: a new platform organism for conversion of lignocellulose into terpene biofuels and bioproducts. *Biotech for Biofuels*, **2017**, 10(241).

[64] Chtourou, M.; Ammar, E.; Nasri, M., *et al.* Isolation of a yeast, Trichosporon cutaneum, able to use low molecular weight phenolic compounds: application to olive mill waste water treatment. *J Chem Technol Biotechnol*, **2004**, 79, 869–878.

[65] Wang, W.; Wei, H.; Knoshaug, E., *et al.* Fatty alcohol production in Lipomyces starkeyi and Yarrowia lipolytica. *Biotechnol Biofuels*, **2016**, 9.

[66] Hu, M.; Wang, J.; Gao, Q., *et al.* Converting lignin derived phenolic aldehydes into microbial lipid by Trichosporon cutaneum. *J Biochenol*, **2018**, 281, 81–86.

[67] Middelhoven, W.; Scorzeti, G.; Fell, J. Trichosporon guehoae sp.nov., an anamorphic basidiomycetous yeast. *Can J Microbiol*, **1999**, 45, 686–690.

[68] Deeba, F.; Pruthi, V.; Negi, Y.S. Aromatic hydrocarbon biodegradation activates neutral lipid biosynthesis in oleaginous yeast. *Bioresour Technol*, **2018**, 255, 273–280.

[69] Su, X.; Zhou, M.; Hu, P., *et al.* Whole-genome sequencing of an acidophilic Rhodotorula sp. ZM1 and its phenol-degrading capability under acidic conditions. *Chemosphere*, **2019**, 232, 76–86.

[70] Jarboui, R.; Baati, H.; Fetoui, F., *et al.* Yeast performance in wastewater treatment: case study of Rhodotorula mucilaginosa. *Environ Technol*, **2012**, 33, 951–960.

[71] Rodriguez, A.; Ersig, N.; Geiselman, G.M., *et al.* Conversion of depolymerized sugars and aromatics from engineered feedstocks by two oleaginous red yeasts. *Bioresour Technol*, **2019**, 286, 121365.

[72] Huang, C.; Chen, X.; Xiong, L., *et al.* Microbial oil production from corncob acid hydrolysate by oleaginous yeast Trichosporon coremiiforme. *Biomass Bioenerg*, **2013**, 49, 273–278.

[73] Patel, A.; Sartaj, K.; Arora, N., *et al.* Biodegradation of phenol via meta cleavage pathway triggers de novo TAG biosynthesis pathway in oleaginous yeast. *J Hazard Mater*, **2017**, 340, 47–56.

[74] Pyne, M.E.; Narcross, L.; Melgar, M., *et al.* An Engineered Aro1 protein degradation approach for increased cis, cis-muconic acid biosynthesis in Saccharomyces cerevisiae. *Appl Environ Microbiol*, **2018**.

[75] Xie, N.Z.; Liang, H.; Huang, R.B., *et al.* Biotechnological production of muconic acid: current status and future prospects. *Biotechnol Adv*, **2014**, 32(3), 615–622.

[76] Yaguchi, A.; Spagnuolo, M.; Blenner, M. Oleaginous yeast for biofuel and oleochemical production. *Curr Opin Biotechnol*, **2019**, 57, 73–81.

[77] Yaguchi, A.; Spagnuolo, M.; Blenner, M. Engineering yeast for utilization of alternative feedstocks. *Curr Opin Biotechnol*, **2018**, 53, 122–129.

[78] McCluskey, K.; Barker, K.; Barton, H., *et al.* The US Culture Collection Network responding to the requirements of the Nagoya Protocol on Access and Benefit Sharing. *mBio*, **2017**, 8, e00982–e1017.

[79] McCluskey, K.; Boundy-Mills, K.; Beattie, G.A. Complying with the nagoya protocol to the convention on biological diversity. *SIMB News*, **2018**, 68, 8–9.

[80] McCluskey, K. A Review of living collections with special emphasis on sustainability and its impact on research across multiple disciplines. *Biopreserv Biobank*, **2017**, 15, 20–30.

[81] Smith, D.; McCluskey, K.; Stackebrandt, E. Investment into the future of microbial resources: culture collection funding models and BRC business plans for biological resource centres. *SpringerPlus*, **2014**, 3, 1–12.

[82] Yaguchi, A.; Franaszek, N.; O'Neill, K., *et al.* Identification of oleaginous yeasts that metabolize aromatic compounds. *JIMB*, **2020**, 47, 801–813.

Chapter 3

Metabolism of model lignin model aromatics by *C. oleaginosus*

3.1 Abstract

The study in Chapter 2 showcases the strength of the *Cutaneotrichosporon* and *Trichosporon* genera in aromatic metabolism. The oleaginous yeast, *Cutaneotrichosporon oleaginosus*, has been extensively studied for its ability to metabolize many non-conventional feedstocks, including phenol-containing waste streams such as distillery wastewater. Given the desirability to valorize underutilized feedstocks such as the aromatic-rich biopolymer lignin, we investigated the ability of *C. oleaginosus* to tolerate and metabolize lignin-derived aromatic compounds. *Cutaneotrichosporon oleaginosus* can tolerate and metabolize model lignin monoaromatics and associated intermediates within putative funneling ortho-cleavage pathways. Growth rates and biomass yield were similar to glucose when grown in *p*-hydroxybenzoic acid (pHBA)

This chapter is based off **Yaguchi, A.**, et al. *MCF*, **2017**, **16**(1) in accordance to the Creative Commons CC BY license.

and resorcinol, but had an increased lag phase when grown in phenol. Oleaginous behavior was observed using resorcinol as a sole carbon source, and fed-batch feeding of this substrate resulted in lipid accumulation of 69.5% on a dry weight basis. This is the first yeast shown to be oleaginous over 50% while growing on aromatic substrates, and shows great promise as a model industrial microbe for biochemical and biofuel production from depolymerized lignin.

3.2 Introduction

Valorization of lignocellulosic biomass wastes is critical for the economic viability of the biomass economy [1]. Furthermore, valorization of wastewater, and food and agricultural wastes may enhance sustainability and provide additional economic benefits [1, 2]. Waste streams are often heterogeneous in nature and contain additives and by-products, such as phenolics, that are toxic to human health. A by-product or waste that has gained considerable attention recently is lignin. It is the second-most abundant biopolymer on Earth and the only renewable, readily-available biopolymer comprised of aromatics [1]. Lignin is a by-product of biomass used as a feedstock for biofuels and biochemical production, and is a prominent by-product of pulp and paper mills. Kraft lignin is typically burned for its heating value, and only 2% is recovered for nonfuel uses [3]. Rather than burning it, lignin could be utilized in biorefineries as a feedstock for microbial production of higher value products [2, 4]. In addition, the large quantities of aromatics in industrial wastewater effluents makes aromatic compounds a prime target for waste valorization [5].

Bacterial metabolism and growth on various phenolic compounds is well-characterized. Many of these aromatics are representative lignin hydrolysate compounds or common products of lignin depolymerization [6]. Significant work has

characterized phenolic metabolism in *Rhodococcus opacus* PD630 [7–9], *Acinetobacter baylyi* ADP1 [10, 11], and *Psuedomonas putida* [4, 12–15]. There are fewer examples of yeast that have been characterized to grow on phenolics, such as *Pichia holstii* [16], *Candida tropicalis* [17], and *Cutaneotrichosporon cutaneum* [18]. These yeasts have evolved to handle a wide variety of aromatic substrates, utilizing so-called funneling pathways to convert diverse molecular species into a small number of metabolites [19]. There are many pathways for aerobic aromatic metabolism, with the ortho- and meta-cleavage pathways being most common [19]. These pathways generate central metabolites, such as succinyl-CoA and acetyl-CoA [20–23]. Understanding the biochemical pathways of aromatic metabolizing organisms enables downstream engineering for high value products, such as oleochemicals for pharmaceutical, fuel, and specialty chemical applications.

Oleaginous microorganisms are a rational starting point for microbial production of oleochemicals, such as lipids for biofuels and omega-3 fatty acids. These microbes are characterized by their capacity to accumulate at least 20% of their mass as lipids. Significant attention has been given to several oleaginous yeast, including *Yarrowia lipolytica* [24–27], *Lipomyces starkeyi* [28], and *Rhodosporidium toruloides* [29]. Substantial work has been done to expand the genetic engineering tools available for these non-conventional oleaginous yeasts, enabling metabolic engineering of these species [30–36]. Despite being established industrial hosts with significant prior work, these yeast species are not suitable for utilizing certain low-cost feedstocks, such as aromatic-rich lignin and phenolic wastewater streams. Given the 50 million tons of lignin currently produced per year [3], there is a great need for microorganisms that are able to tolerate and even metabolize aromatic feedstocks. Of the oleaginous microorganisms that tolerate aromatic toxicity, bacteria do not achieve a high enough biomass and are prone to phage infection, and white-rot fungi grow

too slowly. Similarly, few oleaginous yeasts are known to metabolize aromatics. Furthermore, the oleaginous yeast shown to metabolize aromatics do not maintain high lipid accumulation under aromatic growth conditions.

This chapter addresses the narrow crossover between efficient aromatic metabolism, rapid growth kinetics, and high endogenous lipid accumulation by investigating *Cutaneotrichosporon oleaginosus*, a non-model, non-conventional yeast previously known as *Cryptococcus curvatus* and *Trichosporon oleaginosus*. A recent review summarizes the ability of *C. oleaginosus* to metabolize a number of non-conventional feedstocks and maintain oleaginicity [37]. In this study, we found *C. oleaginosus* tolerates several aromatics and metabolizes them when used as a sole carbon source. Simultaneous growth on mixtures of sugars and aromatics appeared diauxic; however, both substrates were completely consumed. Finally, we demonstrate fed-batch growth on aromatics results in over 69% of biomass accumulated as lipids in a simple shake flask. Due to its rapid growth rate on aromatics, and its significant lipid accumulation, we suggest *C. oleaginosus* has great potential as a model system for aromatic metabolizing oleaginous yeast.

3.3 Materials and Methods

3.3.1 Materials

All enzyme reagents were purchased from New England Biolabs (Ipswich, MA) unless otherwise stated. *Cutaneotrichosporon oleaginosus* was obtained from the American Type Culture Center (ATCC® 20509™).

3.3.2 Tolerance studies

Tolerance studies were performed in 48-well plates (Nunclon® 48-well plate) in a Biotek® Synergy™Mx multimode microplate reader. 2 mL YPD pre-cultures were grown overnight and used to inoculate 250 μL YPD (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose) containing various aromatic compounds described in Table 1 to an OD=0.3. Cells were grown over 48 hours in a 48 well plate with the lid on, fast shake speed, and at 28 °C. The plate reader scanned every hour at 600 nm. Studies were performed at least in triplicate. Hydroxyquinol and catechol resulted in media color changes that prevented use of the spectrophotometric measurement, so these samples were plated to assess CFUs.

3.3.3 Aromatic Growth Studies

Single and dual carbon source cultures were cultured in the same fashion. *Cutaneotrichosporon oleaginosus* (ATCC® 20509™) cells were grown in 2 mL YPD pre-cultures overnight. Cells were washed three times with new media, and inoculated to an initial OD600 of 0.3 in 50 mL baffled Erlenmeyer Corning® flasks containing 15 mL of either high nitrogen (TOHN) or low nitrogen (TOLN) media (modified from Kourist, et al [38]) with appropriate carbon source and concentration. Details on media composition are included in Table 3.1. Cell washing entailed centrifuging cells at 1,100 x g for 4 minutes at 4 °C, decanting supernatant, and re-suspending in destination media. Carbon sources included glucose, xylose, phenol, resorcinol, and *p*-hydroxybenzoic acid (pHBA). Optical density (OD) readings were measured on a Thermo Scientific NanoDrop™ 2000 at 600 nm and corrected by a factor of 17.725. All experiments were performed in triplicate.

Two-stage feeding and fed-batch cultures were started in the same manner.

Table 3.1. Media composition for high nitrogen (TOHN), low nitrogen (TOLN), and defined low nitrogen (DLN) conditions.

Component	**TOHN**	**TOLN**	**DLN**
Carbon source	X g/L	X g/L	3.0 g/L
Yeast extract	0.75 g/L	0.75 g/L	-
$(NH_4)_2SO_4$	4 g/L	1.2 mg/L	1.66 mg/L
$MgSO_4 \cdot 7\,H_2O$	1.5 g/L	1.5 g/L	1.5 g/L
KH_2PO_4	0.4 g/L	0.4 g/L	0.4 g/L
$CaCl_2 \cdot 2\,H_2O$	0.22 g/L	0.22 g/L	0.22 g/L
$ZnSO_4 \cdot 7\,H_2O$	0.55 g/L	0.55 g/L	0.55 g/L
$MnCl_2 \cdot 4\,H_2O$	24.2 μg/L	24.2 μg/L	24.2 μg/L
$CuSO_4 \cdot 5\,H_2O$	25 μg/L	25 μg/L	25 μg/L

After biomass accumulation in the two-stage feeding experiments, the whole cell culture was transferred to a 50 mL centrifuge tube, spun down at 4,000 rpm for 3 minutes at 4 °C, and re-suspended in 15 mL fresh TOLN or DLN media (Table 3.1). The entire 15 mL was transferred back to the 50 mL baffled flask and cultured for the 48 hour lipid accumulation phase. For fed-batch experiments, 1.5 mL of cell culture was removed after the initial biomass accumulating phase. Stock resorcinol (20 g/L in ddH_2O) was added to feed 2 g/L at a time. This removal of biomass and addition of stock resorcinol was repeated every 24 hours until the end of the experiment. All experiments were performed in triplicate.

3.3.4 Dry cell weight and lipid analysis

Cells were harvested for dry cell weight by washing 10 mL of cell culture with 20 mL of ddH_2O three times and drying overnight at 40 °C under vacuum in aluminum pans. Dry cell weights were measured using an analytical balance. To identify and quantify lipids in cell biomass, extracted cellular lipids were transesterified to FAMEs

as described previously with minor modifications. Briefly, 1 mL cell culture was harvested and spun down at 13,000 rpm for 3 minutes at 25 °C. 100 μL glyceryl triheptadecanoate at a concentration of 2 mg/mL methanol was added to the cell pellet as an internal standard. Lipids were transesterified to FAMEs with 500 μL of 0.5 N sodium methoxide followed by 30 min of vortexing at 2,000 rpm. The solution was neutralized with 40 μL sulfuric acid. FAMEs were extracted by adding 850 μL hexane followed by 20 min of vortexing at 2,000 rpm. The mixture was centrifuged for 1 min at 8,000 rpm, and 800 μL of the organic layer was collected for GC-FID analysis (Agilent 7890B) and quantification [25].

3.3.5 Substrate utilization

Aromatic substrate utilization was analyzed using a Waters 600E multisolvent delivery (Waters Corporation) high performance liquid chromatography (HPLC) system with a BioRad Fast Acid Analysis HPLC column with 10% v/v acetonitrile and 0.01 N H_2SO_4 in a 1:1 mixture as the eluent, a flow rate 0.6 mL/min, temperature of 65 °C, and with a Waters 996 PDA detector. Phenol was detected at 270 nm, resorcinol at 274 nm, and pHBA at 254 nm. Glucose and xylose were measured using the same column at 85 °C, 5 mM H_2SO_4 eluent, and a Waters 2414 refractometer. Concentrations were calculated from standard curves created for each carbon source in the appropriate medium.

3.4 Results

3.4.1 *Cutaneotrichosporon oleaginosus* tolerates model lignin-derived aromatics

Aromatic metabolizing microorganisms have a conserved pathways, known as funneling pathways, for cleavage of aromatic compounds. The putative ortho-cleavage pathways result in TCA intermediates succinyl-CoA and acetyl-CoA, which is shown in Figure 3.1 [4]. An initial BLAST search of the *C. oleaginosus* genome returns several putative enzymes from this ortho-cleavage pathway, suggesting it may be able to metabolize aromatics.

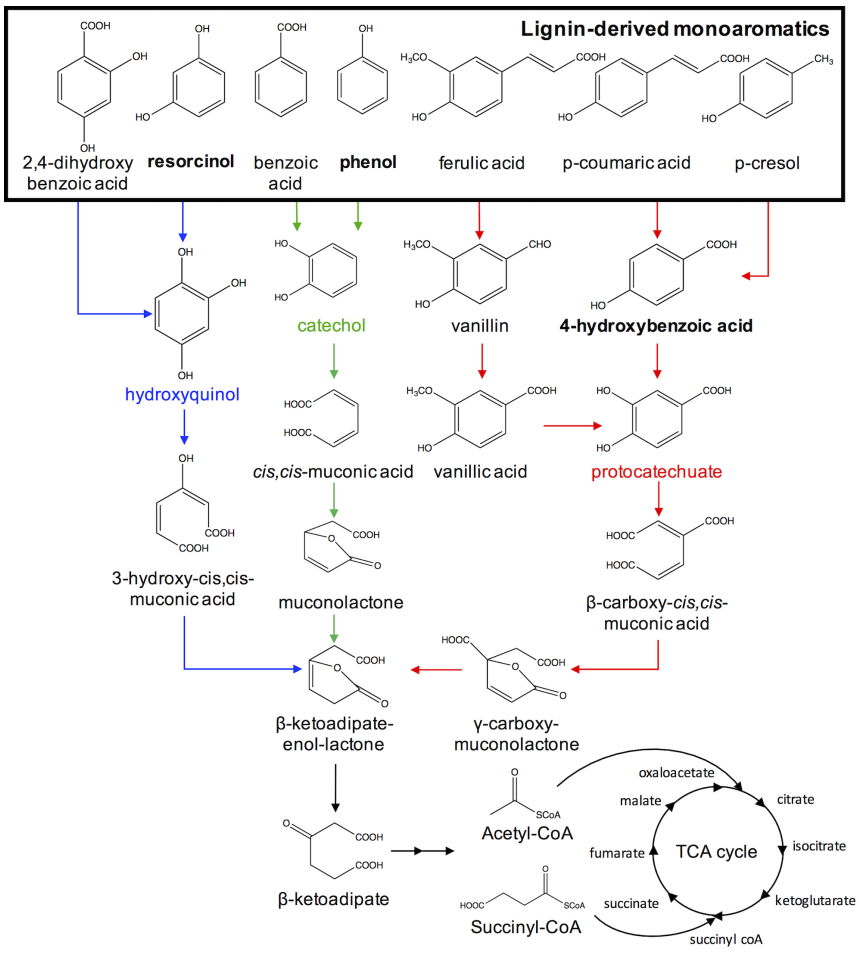


Figure 3.1. Funneling pathways for ortho-cleavage aromatic metabolism.

Given its potential for aromatic metabolism, and its known tolerance to several inhibitors, we reasoned that *C. oleaginosus* was likely to tolerate many aromatic compounds. Fifteen model aromatic compounds were screened to determine the minimum inhibitory concentration (MIC), or the concentration which completely inhibits cell growth (Table 3.2). Notably, protocatechuate had the highest MIC at 15.5 g/L for *C. oleaginosus*. The next highest MIC value was for hydroxyquinol at 7.5 g/L, followed by resorcinol at 5.5 g/L. Other compounds with MIC values over 1 g/L included: 2,4-dihydroxybenzoic acid (1.5 g/L), catechol (3.0 g/L), *p*-coumarate (1.4 g/L), *p*-hydroxybenzoic acid (4.0 g/L), and phenol (1.2 g/L).

Table 3.2. Table of minimum inhibitory concentration (MIC) values for *C. oleaginosus*.

Compound	**MIC (g/L)**
2,4-dihydroxybenoic acid	1.5
Benzoic acid	1.4
Catechol	3.0
Ferulic acid	0.6
Guaiacol	1.5
Hydroxyquinol	7.5
p-coumaric acid	1.4
p-cresol	0.8
pHBA	4.0
Phenol	1.2
Protocatechuate	15.5
Resorcinol	5.5
Syringic acid	2.0
Vanillic acid	1.6
Vanillin	0.2

3.4.2 *Cutaneotrichosporon oleaginosus* metabolizes model lignin-derived aromatics

The tolerance to a number of different compounds suggests *C. oleaginosus* may be a good candidate for aromatic metabolism. To determine if this yeast has a natural metabolism of aromatic compounds, we chose three aromatics for further analysis — phenol, resorcinol, and pHBA. Resorcinol was picked over 2,4-dihydroxybenzoic acid because it had a higher MIC. Phenol was picked over benzoic acid because of the broad interest in phenol-containing waste stream in industry and similar MIC. pHBA was picked over all other compounds upstream of protocatechuate because it had the highest MIC. Each aromatic serves as a representative compound for a branch of the funneling pathways shown in Figure 3.1. We picked compounds upstream in the pathway as a stringent test of metabolism. *Cutaneotrichosporon oleaginosus*

cells were grown in high nitrogen (TOHN) and low nitrogen (TOLN) media with 1 g/L carbon source (phenol, resorcinol, pHBA, or glucose). Growth curves shown in Figure 3.2 represent growth in TOHN and are compared to a no-carbon-added negative control, which accounts for growth attributed to the small amount of yeast extract in the media.

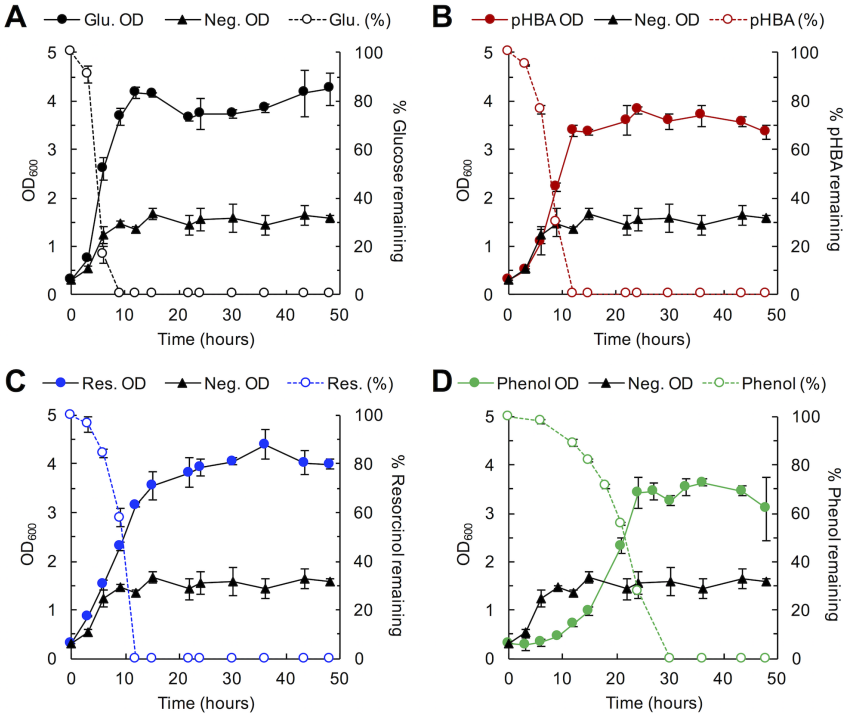


Figure 3.2. *Cutaneotrichosporon oleaginosus* can metabolize aromatic substrates.

All cells grown in media with an added carbon source reach similar biomass titers, indicating that aromatic compounds do not deleteriously affect total biomass accumulation. Cells inoculated into resorcinol and pHBA-containing media grow similarly to each other and somewhat slower to cells inoculated into glucose, with specific growth rates for glucose, resorcinol, and pHBA of 0.157± 0.016 hr^{-1}, 0.0965± 0.012 hr^{-1}, and 0.092± 0.001 hr^{-1}, respectively. Phenol induces a slower specific growth rate of 0.058± 0.002 hr^{-1} and has a longer lag phase, as cells did not enter exponential

phase until 12 hours after inoculation. The negative control has a growth rate of 0.101± 0.010 hr^{-1}. Substrate utilization data, as measured by HPLC-UV-vis, indicate that all the compounds were fully consumed by 9 hours for glucose, 12 hours for resorcinol and pHBA, and 30 hours for phenol. In the high nitrogen media, lipid accumulation was slightly lower in phenol (5.26± 0.09% w/w) as compared to resorcinol (7.21± 0.21% w/w), pHBA (7.55± 0.24% w/w), and glucose (6.72± 0.29% w/w) (Table 3.4).

Table 3.3. Measurement of dry cell weight, lipid titer, and percentage lipid accumulation in high nitrogen (TOHN) and low nitrogen (TOLN) media containing 1 g/L carbon.

Substrate	**N Conc. (g/L)**	**Biomass Titer(g/L)**	**Lipid Titer (mg/L)**	**Lipid Accumulation (%)**
Glucose	4	0.71±0.00	47.2±0.7	6.72±0.29
Resorcinol	4	0.70±0.06	50.4±3.3	7.21±0.21
pHBA	4	0.64±0.05	48.3±3.6	7.55±0.24
Phenol	4	0.80±0.01	42.2±0.8	5.26±0.09
Negative	4	0.22±0.03	26.5±4.9	12.22±1.20
Glucose	0.012	0.96±0.02	52.0±8.7	5.44±1.04
Resorcinol	0.012	0.66±0.01	17.3±2.3	4.67±1.24
pHBA	0.012	0.69±0.02	32.4±1.7	4.72±0.23
Phenol	0.012	0.80±0.01	49.2±2.8	6.13±0.25
Negative	0.012	0.28±0.00	30.7±7.8	6.11±0.69

GC-FID was used to determine the fatty acid composition of cellular lipids (Table 3.4). This study agrees with previous reports that this yeast accumulates high amounts of linoleic acid and α-linolenic acid when grown in glucose [39, 40]. However, cells grown with resorcinol, pHBA, and phenol show similar amounts of linoleic acid, an increase in oleic acid, and decrease in α-linolenic acid (Table 3.4). When cells are grown in low nitrogen glucose-containing media, lipid accumulation profiles show an increase in palmitic, palmitoleic, stearic, and oleic acid and a decrease in linoleic

Table 3.4. Fatty acid composition profile (%) for cells grown in high nitrogen (TOHN) and low nitrogen (TOLN) media containing 1 g/L carbon.

Substrate	**N Conc. (g/L)**	**C16:0**	**C16:1**	**C18:0**	**C18:1**	**C18:2**	**C18:3**
Glucose	4	8.9±1.2	14.2±0.4	6.6±0.4	19.7±0.4	41.4±2.1	9.2±0.1
Resorcinol	4	9.1±0.5	14.5±0.6	6.5±0.2	31.6±1.1	38.3±2.1	0.0±0.0
pHBA	4	10.1±0.7	14.2±0.4	7.6±0.7	30.8±1.3	36.0±0.6	1.4±1.9
Phenol	4	8.6±0.5	11.1±0.3	6.1±0.6	37.9±3.5	36.3±2.2	0.0±0.0
Glucose	0.012	12.5±0.4	17.4±0.5	8.0±0.3	26.2±0.3	29.3±0.4	6.7±0.1
Resorcinol	0.012	8.5±4.4	9.9±4.9	0.0±0.0	43.1±6.4	38.5±2.8	0.0±0.0
pHBA	0.012	18.0±0.5	13.4±0.4	4.0±2.9	31.9±1.7	32.6±1.4	0.0±0.0
Phenol	0.012	7.5±0.5	13.1±0.1	6.9±0.2	39.5±1.2	32.9±1.2	0.0±0.0

and α-linoleic acid. In low nitrogen aromatic containing media, the lipid profiles are similar to those measured in high nitrogen media.

Amongst the three aromatic compounds used in the metabolism studies, resorcinol could be used at a concentration of 3 g/L, the highest concentration with no change in the growth or health of the cells (data not shown). Therefore, we used this substrate in subsequent experiments to improve lipid accumulation from aromatics. Given the toxicity limit for aromatic substrates, we were unable to create nitrogen limited condition needed for lipid accumulation using high concentrations of carbon in growth experiments. Furthermore, a small amount of yeast extract was needed for *C. oleaginosus* growth, limiting the achievable carbon to nitrogen (C:N) ratio to 36:1, assuming 11.1% nitrogen in the BD technical bacto yeast extract (reported in the BD bionutrients technical manual). As a result, we used different feeding strategies to effectively increase the C:N ratio.

3.4.3 Co-utilization of carbon sources results in diauxic growth

Consolidated bioprocessing involving simultaneous biomass degradation and substrate utilization would contain sugars, such as glucose and xylose, and lignin-derived components; therefore, we measured growth in mixed carbon sources. *Cutaneotrichosporon oleaginosus* demonstrates diauxic growth when grown in low nitrogen media comprised of 1.5 g/L each of sugar (either glucose or xylose) and resorcinol (Figure 3.3). Substrate utilization data show preferential metabolism of the sugars by hour 12, followed by resorcinol; however, the slow consumption of resorcinol during the first 12 hours suggests some degree of simultaneous utilization. The consumption rate of resorcinol increases significantly once the sugars are fully consumed. The resulting growth after glucose or xylose is fully consumed has a long lag phase until at least hour 30, when cell growth continues. This observation is in contrast to *C. oleaginosus* growth in media containing both xylose and glucose, where glucose substrate was preferentially consumed, but the growth rates were similar between both carbon sources (Figure 3.4). Measurements of cell density and lipid accumulation show no significant difference in percent lipid accumulation between xylose+resorcinol media, xylose media, and resorcinol media; however, lipid accumulation percentage in glucose was lower than resorcinol and both were lower than mixed glucose+resorcinol media (Table 3.5). However, it should be noted that due to the use of a small amount of yeast extract, the C:N value of 36.1 is still somewhat high, discouraging significant lipid accumulation. Lipid profiles show consistent distribution of fatty acids across various carbon sources (Table 3.6).

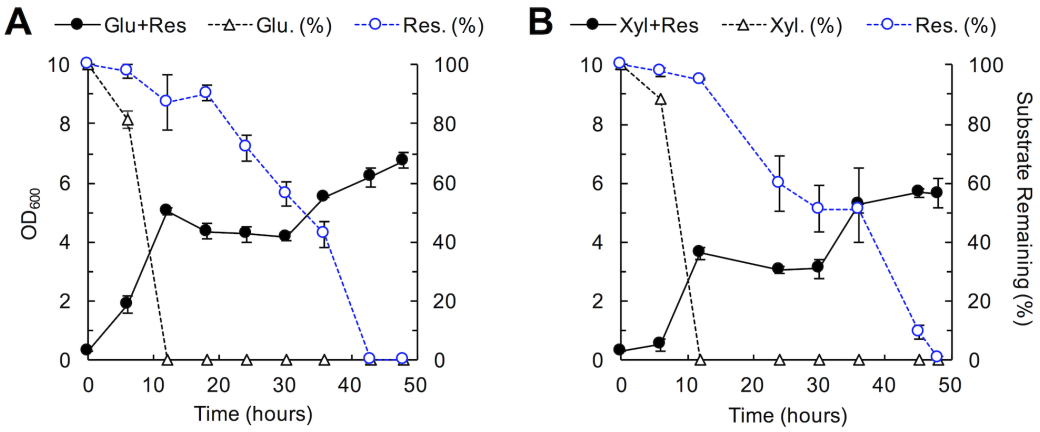


Figure 3.3. Diauxic growth of *C. oleaginosus* cells when cultured in dual-carbon media.

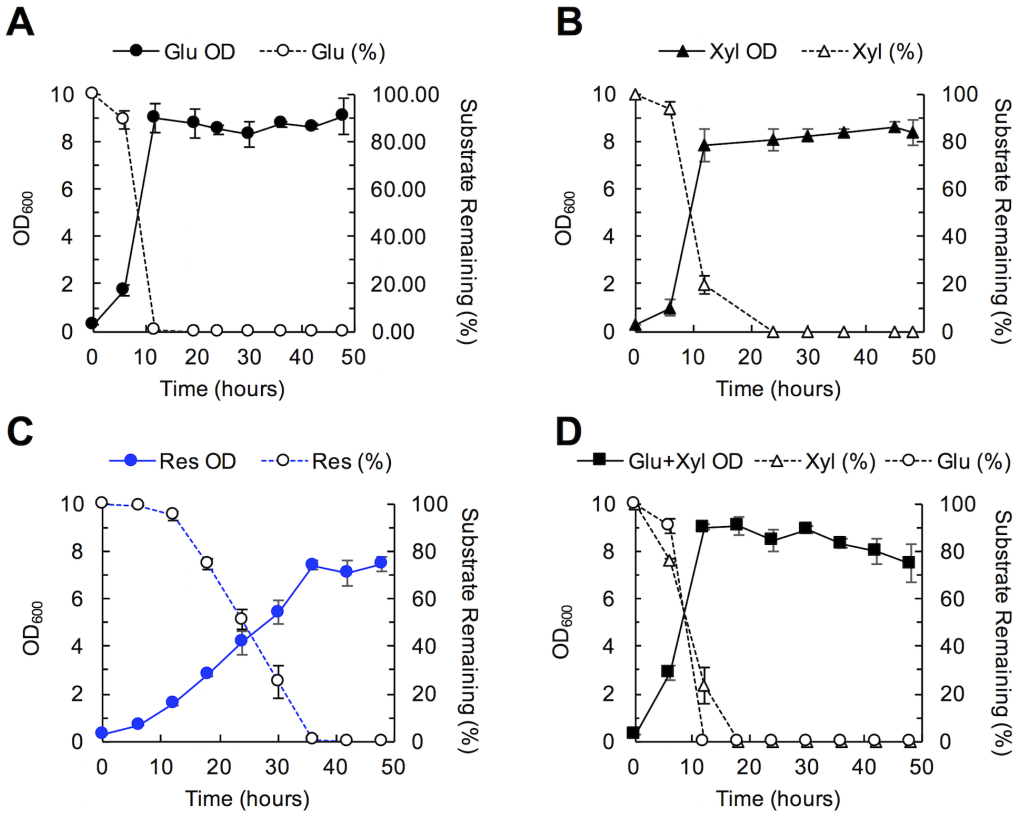


Figure 3.4. Growth curves and substrate consumption data for glucose, xylose, resorcinol, and glucose/xylose.

Table 3.5. Measurement of dry cell weight, lipid titer, and percentage lipid accumulation for cells grown in dual-carbon media and comparison to single carbon media.

Substrate	**C Conc. (g/L)**	**Biomass Titer(g/L)**	**Lipid Titer (mg/L)**	**Lipid Accumulation (%)**
Xylose	3.0	1.58±0.03	230.0±11.7	14.5±0.7
Glucose	3.0	1.19±0.02	97.5±0.6	8.2±0.2
Resorcinol	3.0	1.00±0.09	111.8±4.6	11.3±1.0
Xyl + Res	1.5 / 1.5	1.17±0.35	138.3±13.8	12.6±3.1
Glu + Res	1.5 / 1.5	0.72±0.03	100.6±10.5	13.9±1.7

Table 3.6. Fatty acid composition profile (%) for cells grown in dual-carbon source media.

Substrate	**C Conc. (g/L)**	**C16:0**	**C16:1**	**C18:0**	**C18:1**	**C18:2**	**C18:3**
Xylose	3.0	1.9±0.2	21.6±0.1	16.9±0.2	41.8±0.6	15.7±0.1	3.0±0.0
Glucose	3.0	3.2±0.6	21.2±0.8	14.5±0.8	41.8±1.2	15.7±0.6	2.7±0.1
Resorcinol	3.0	2.5±0.4	19.5±0.3	17.6±1.0	47.5±1.0	13.0±1.0	0.0±0.0
Xyl + Res	1.5 / 1.5	3.0±0.3	17.1±0.7	20.5±0.5	46.5±0.8	12.0±0.6	0.9±0.6
Glu + Res	1.5 / 1.5	3.4±0.3	17.6±0.6	18.5±0.4	46.4±1.3	12.8±0.2	1.4±0.1

3.4.4 *C. oleaginosus* remains oleaginous while metabolizing high concentration of aromatics

As noted earlier, the highest concentration of resorcinol that did not result in a growth rate defect was 3 g/L. While already able to metabolize a higher concentration of aromatic substrates than reported in the literature,37 we sought to determine if we could improve the amount of carbon delivered to the cells. The first attempt feed more resorcinol was through a two-stage feeding strategy where *C. oleaginosus* cells were grown for 30 hours in TOLN containing 3 g/L resorcinol to accumulate biomass. The time of 30 hours was chosen because the cells are in late exponential phase at this point (Figure 3.5). Cells were harvested and resuspended in fresh TOLN (C:N = 36.1) or with fresh defined low nitrogen (DLN) media (C:N = 361.66), both containing 3 g/L resorcinol and grown for another 48 hours . DLN media use was possible once cell biomass was obtained in the first stage, and the yeast extract could be omitted from the TOLN media. The DLN C:N ratio was altered to match that of previously established C:N ratios by eliminating yeast extract and supplementing nitrogen in the form of ammonium sulfate [38]. Two-stage feeding is required because DLN alone does not promote significant biomass accumulation (0.12±0.0 g/L cell density after 80 hours). The higher C:N ratio was better for lipid accumulation (Table 3.7). Switching into a lipid accumulating media results in a 60% improvement in lipid accumulation from 14.5±0.9% to 35.9±1.2% (Table 3.7) while maintaining a similar lipid composition profile similar to that of previous experiments (Additional File 1, Supplemental Table 2).

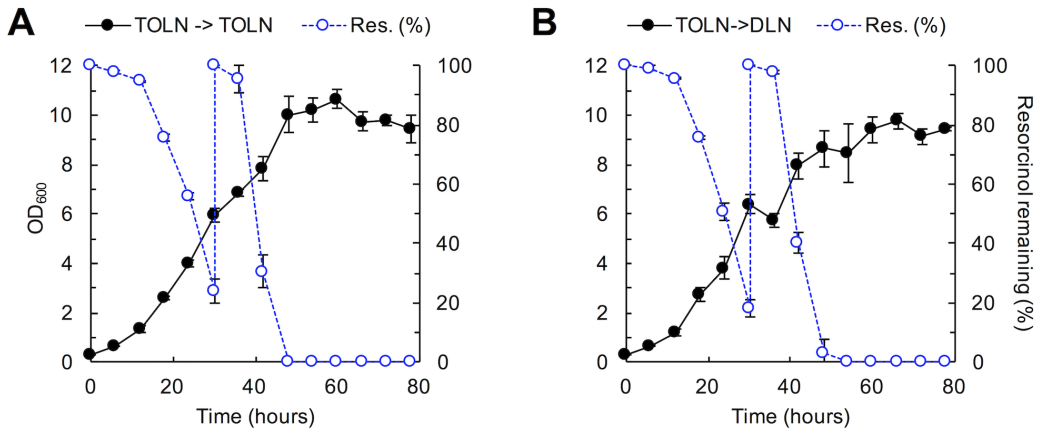


Figure 3.5. Two-stage feeding cultures demonstrate complete consumption of resorcinol and no metabolic limitations.

Table 3.7. Measurement of dry cell weight, lipid titer, and percentage lipid accumulation for cells grown in dual-carbon media and comparison to single carbon media.

Substrate	**Final C Conc. (g/L)**	**Final N Conc. (g/L)**	**Biomass Titer(g/L)**	**Lipid Titer (g/L)**	**Lipid Accumulation (%)**
TOLN→TOLN	6.0	0.00120	1.81±0.0	0.26±0.1	14.5±0.7
TOLN→DLN	6.0	0.00012	1.48±0.1	0.53±0.0	8.2±0.2
Fed batch	13.0	0.00120	2.36±0.2	1.64±0.2	11.3±1.0

3.4.5 *C. oleaginosus* accumulates greater than 69% of its biomass as lipids in a fed-batch shake flask using resorcinol as a sole carbon source

Cells for the fed-batch experiment were initially inoculated to an OD of 0.3 in TOLN containing 3 g/L of resorcinol and were grown for 36 hours. After this time, 2 g/L of resorcinol were fed every 24 hours while keeping the culture volume constant (Figure 3.6). In this manner, nutrients such as nitrogen, were depleted over time, and only carbon was replenished with every feeding. A total of 13 g/L of resorcinol was delivered over 160 hours, as compared to 3 g/L over 48 hours in batch experiments, resulting in a 73% improvement in total carbon fed. Fed-batch feeding results in 87.4% improvement in total lipid titer from 0.11±0.0 g/L (Table 3.5) to 1.64±0.2 g/L (Table 3.7). Lipid accumulation increased from 11.3±0.0% (Table 3.5) to 69.5±4.0% (Table 3.7), resulting in an 87.4% improvement. When compared to the two-stage feeding described earlier, lipid titer improved by 39.1% and lipid accumulation improved by 38.8%. The lipid profile was similar to those found in two-stage feeding experiments (Table 3.8).

Table 3.8. Fatty acid composition (%) for two-stage and fed-batch cultures.

Substrate	Final C Conc. (g/L)	C16:0	C16:1	C18:0	C18:1	C18:2	C18:3
TOLN→TOLN	6.0	2.1±0.1	19.9±0.3	13.1±0.3	46.7±0.4	17.1±0.3	1.0±0.1
TOLN→DLN	6.0	1.0±0.1	23.7±0.1	15.3±0.3	51.5±0.6	8.1±0.4	0.5±0.0
Fed batch	13.0	0.3±0.0	20.3±1.1	23.3±1.0	50.2±1.2	5.7±0.4	0.2±0.0

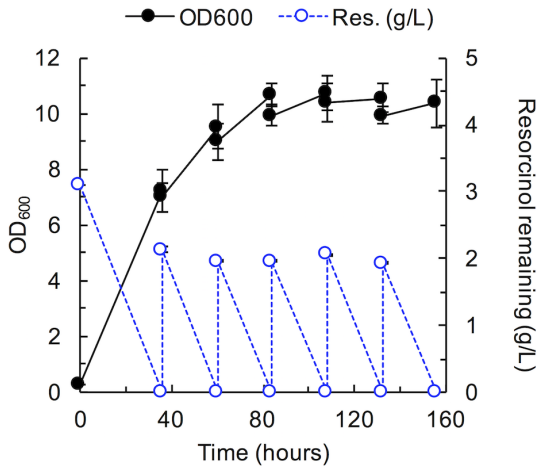


Figure 3.6. Fed-batch feeding strategy results in the highest literature value for lipid production derived from a phenolic substrate.

3.5 Discussion

Cutaneotrichosporon oleaginosus has been studied for its ability to accumulate a significant amount of lipids from a variety of feedstocks. This is the first report showing that *C. oleaginosus* is able to metabolize model lignin monoaromatic compounds when used as a sole carbon source. It is likewise the first report of *C. oleaginosus* metabolizing aromatics in mixed sugar and aromatic substrates. While other yeast have been shown to tolerate and metabolize aromatic compounds [41],

this is the first demonstration of yeast growing on aromatics and accumulating lipids greater than 20% on a dry weight basis. In fed-batch experiments, *C. oleaginosus* was able to accumulate over 69% of its dry cell weight as lipids using aromatic compound resorcinol as the sole carbon source, demonstrating its promise as a model organism for aromatic to oleochemical conversion.

To date, the highest reported lipid accumulation from *C. oleaginosus* is 73.4% on a dry weight basis using acetate as the sole carbon source [42]. This lipid accumulation compares favorably to the lipid accumulation of 69.5% on a dry weight basis from resorcinol demonstrated in this work. We were unable to find any report of lipid accumulation in a microbe grown on resorcinol as the sole carbon source. The literature has focused on bioremediation of xenobiotics and anthropomorphic aromatic substrates rather than lipid accumulation. Amongst oleaginous bacteria, *Rhodococcus opacus* strains DSM 1069 and PD630, accumulate up to 20% of dry weight as lipids when grown on pHBA and vanillic acid in optimized fed-batch reactors [8]. Phenol toxicity impacted the growth rate when reaching 0.3 g/L.9 A strain of *R. opacus* PD630 evolved for higher tolerance, increasing from 0.3 g/L up to 1.5 g/L achieving lipid accumulation of 11.7% on a dry weight basis [9]. The titer was not reported, but could not be higher than 1.5 g/L, the concentration of phenol used in the experiments, and theoretically must be much lower. By comparison, our highest lipid titer was 1.64$\pm$0.2 g/L and was achieved using a wild-type strain. *Cutaneotrichosporon oleaginosus* is able to accumulate significantly higher titers of lipids when grown on higher concentrations of preferred substrates such as glucose [39, 40, 43]; however, the toxicity of aromatic compounds limits the concentration of aromatics dosed at a single time. The toxicity limitation of aromatic substrates limits the carbon to nitrogen (C:N) ratio when growing initial cell biomass, which benefits significantly from a small amount of yeast extract. In one study, *C. oleaginosus* was previously shown

to achieve the highest percent accumulation and lipid titer at a C:N ratio of 99 [44], whereas another study used a C:N ratio of 360 [38]. These high C:N ratios are consistent with our observations of poor lipid accumulation using a C:N ratio of 36, but drastically greater lipid accumulation using a ratio of 360 in fed-batch experiments, emphasizing the role of the C:N ratio on lipid accumulation in *C. oleaginosus*.

In nature, lignin can be enzymatically depolymerized to a wide variety of aromatics. Similarly, catalytic depolymerization also results in a heterogonous mixture of aromatic compounds [41]. Funneling pathways overcome the inherent heterogeneity of lignin by converting key aromatic compounds to important intermediates for the TCA cycle [4, 20–23]. The β-ketoadipate pathway is well-conserved amongst aromatic metabolizing bacteria and yeast. The most common pathways are intra-diol (ortho) ring cleavage and extra-diol (meta) ring cleavage pathways. Based on our demonstration of resorcinol, pHBA, and phenol metabolism, *C. oleaginosus* appears to have each of the three major funneling pathways. The enzymatic activity of aromatic metabolism was previously studied *Cutaenotrichosporon cutaneum*. Metabolite analysis demonstrated ortho ring cleavage reaction products and the absence of meta ring cleavage products [45, 46]. Given both are from the same genus, it is likely that *C. oleaginosus* uses the ortho ring cleavage pathway, however, additional work is required to identify the precise mechanism of metabolism and the genes coding for aromatic metabolism genes.

Lignocellulosic hydrolysates contain sugars derived from the cellulose and hemicellulose content and aromatics from the lignin content; therefore, our demonstration that *C. oleaginosus* co-utilizes sugars (glucose or xylose) and resorcinol establishes the potential to utilize multiple components of lignocellulosic biomass. In addition, this finding suggests it may be unnecessary to remove phenolic content in recalcitrant feedstocks, as is typically required due aromatic toxicity and lack of

metabolism [47]. Related organisms show similar co-utilization. In fact, wild type *C. cutaneum* was shown to simultaneously uptake glucose and phenol, with the rate of glucose utilization being much higher than that of phenol utilization [48], consistent with our findings. The presence of glucose decreased phenol utilization and catechol-1,2 dioxygenase (an enzyme belonging to the catechol funneling pathway) activity by 60% and 75%, respectively.

We show *C. oleaginosus* is able to tolerate 15 different compounds to various degrees, with MICs as high as 15.5 g/L for protocatechuate and as low as 0.2 g/L for vanillin. Improving tolerance will be important for further development of industrially relevant strains. Aromatic compounds are often toxic to microbes due to their partitioning into the cell membrane. The mechanism for aromatic tolerance in bacteria is mostly attributed to redistribution of the highly dynamic lipid composition and the activation of aromatic efflux transporters [49–53]. *Pseudomonas putida* strains isolated from toluene-polluted sites responded to aromatic induced membrane stress by increasing the ratio of trans and cis fatty acid isomers, making the cell membrane more rigid. *Pseudomonas putida* DOT-T1E has been shown to tolerate and metabolize 17 g/L pHBA. When primed in low concentrations of pHBA, this strain can tolerate up to 24 g/L pHBA, also attributed to the cell membrane rigidity [54]. A more recent study showed the fatty acid composition of *R. opacus* strains grown on phenol and benzene also exhibited increased trans-fatty acids and increased the 10-methyl branched fatty acid content in the presence of aromatics [51]. While these specific membrane adaptations are unique to prokaryotes, eukaryotes can likewise alter membrane fluidity through alterations in composition and sterols [55].

Aside from altering membrane fluidity, strains of *P. putida* have been shown to utilize efflux pumps to remove toluene from the cell membrane [50, 53]. However, efflux pumps would be counter the desired aromatic metabolism. Recently, an evolved

strain of *R. opacus* PD630 showed increased importer activity led to faster growth on phenol, and higher tolerance to phenol compared to the wild type strain [9]. The recent demonstration of genetic engineering tools for *C. oleaginosus* [40] makes it possible to engineer increased tolerance through control of membrane fluidity and overexpression of rate limiting aromatic metabolism genes.

3.6 Conclusions

Cutaneotrichosporon oleaginosus is the first yeast shown to metabolize aromatics and remain highly oleaginous, accumulating nearly 70% of its biomass as lipids when grown on resorcinol in fed-batch. This yeast shows promise for utilizing aromatic-containing feedstocks, such as lignin-derived wastes and wastewater effluent, for microbial production of oleochemicals or aromatic-derived compounds due to its natural ability to tolerate and metabolize relatively high concentrations of aromatics. Improving tolerance to and metabolism of aromatics will be required for industrial applications. This is likely to require the development of additional metabolic engineering tools and a more complete understanding of *C. oleaginosus* genetics and metabolism. Nevertheless, this yeast is well positioned to become a model system for aromatics metabolism to lipids and oleochemicals.

3.7 Acknowledgments

This work is adapted from Ref. 56. According to the Creative Commons CC BY license, I am not required to obtain special permissions since this work is open access and I am a primary author.

This material is based upon work supported by U.S. Department of Agriculture

(USDA) Sun Grant (2014-38502-22598). This work was also supported by Creative Inquiry funds from Clemson University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the USDA or Clemson University.

Bibliography

[1] Thakur, V.; Thakur, M.; Raghavan, P., *et al.* Valorization of Biomass: Deriving More Value from Waste. *Science*, **2012**, 337, 695–699.

[2] Ragauskas, A.J.; Beckham, G.T.; Biddy, M.J., *et al.* Lignin valorization: improving lignin processing in the biorefinery. *Science*, **2014**, 344(6185), 1246843.

[3] Gosselink, R.JA.; Jong, E.; Guran, B., *et al.* Co-ordination network for lignin - standardisation, production and applications adapted to market requirements (EUROLIGNIN). *Ind Crop Prod*, **2004**, 20(2), 121–129.

[4] Linger, J.G.; Vardon, D.R.; Guarnieri, M.T., *et al.* Lignin valorization through integrated biological funneling and chemical catalysis. *Proc Natl Acad Sci U S A*, **2014**, 111(33), 12013–12018.

[5] Villegas, L.; Mashhadi, N.; Chen, M., *et al.* A Short Review of Techniques for Phenol Removal from Wastewater. *Current Pollution Reports*, **2016**, 2, 157–167.

[6] Wang, H.; Tucker, M.; Ji, Y. Recent Development in Chemical Depolymerization of Lignin: A Review. *Journal of Applied Chemistry*, **2013**, 2013, 1–9.

[7] Hollinshead, W.D.; Henson, W.R.; Abernathy, M., *et al.* Rapid metabolic analysis of Rhodococcus opacus PD630 via parallel C-13-metabolite fingerprinting. *Biotechnol Bioeng*, **2016**, 113(1), 91–100.

[8] Kosa, M.; Ragauskas, A.J. Bioconversion of lignin model compounds with oleaginous Rhodococci. *Appl Microbiol Biotechnol*, **2012**, 93(2), 891–900.

[9] Yoneda, A.; Henson, W.R.; Goldner, N.K., *et al.* Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630. *Nucleic Acids Res*, **2016**, 44(5), 2240–2254.

[10] Dal, S.; Trautwein, G.; Gerischer, U. Transcriptional organization of genes for protocatechuate and quinate degradation from Acinetobacter sp. strain ADP1. *Appl Environ Microbiol*, **2005**, 71(2), 1025–1034.

[11] Fischer, R.; Bleichrodt, F.S.; Gerischer, U.C. Aromatic degradative pathways in Acinetobacter baylyi underlie carbon catabolite repression. *Microbiology*, **2008**, 154(Pt 10), 3095–3103.

[12] Basu, A.; Apte, S.K.; Phale, P.S. Preferential utilization of aromatic compounds over glucose by Pseudomonas putida CSV86. *Appl Environ Microbiol*, **2006**, 72(3), 2226–2230.

[13] Jimenez, J.I.; Minambres, B.; Garcia, J.L., *et al.* Genomic analysis of the aromatic catabolic pathways from Pseudomonas putida KT2440. *Environ Microbiol*, **2002**, 4(12), 824–841.

[14] Kumar, A.; Kumar, S.; Kumar, S. Biodegradation kinetics of phenol and catechol using Pseudomonas putida MTCC 1194. *Biochemical Engineering Journal*, **2005**, 22(2), 151–159.

[15] Ravi, K.; Garcia-Hidalgo, J.; Gorwa-Grauslund, M.F., *et al.* Conversion of lignin model compounds by Pseudomonas putida KT2440 and isolates from compost. *Appl Microbiol Biotechnol*, **2017**, 101(12), 5059–5070.

[16] Sinigaglia, M.; Benedetto, N.A.; Bevilacqua, A., *et al.* Yeasts isolated from olive mill wastewaters from southern Italy: technological characterization and potential use for phenol removal. *Appl Microbiol Biot*, **2010**, 87(6), 2345–2354.

[17] Ahuatzi-Chacon, D.; Ordorica-Morales, G.; Ruiz-Ordaz, N., *et al.* Kinetic study of phenol hydroxylase and catechol 1,2-dioxygenase biosynthesis by Candida tropicalis cells grown on different phenolic substrates. *World J Microb Biot*, **2004**, 20(7), 695–702.

[18] Alexieva, Z.; Gerginova, M.; Zlateva, P., *et al.* Monitoring of aromatic pollutants biodegradation. *Biochemical Engineering Journal*, **2008**, 40(2), 233–240.

[19] Harwood, C.S.; Parales, R.E. The beta-ketoadipate pathway and the biology of self-identity. *Annu Rev Microbiol*, **1996**, 50, 553–590.

[20] Bugg, T.D.; Ahmad, M.; Hardiman, E.M., *et al.* Pathways for degradation of lignin in bacteria and fungi. *Nat Prod Rep*, **2011**, 28(12), 1883–1896.

[21] Pieper, D.H. Aerobic degradation of polychlorinated biphenyls. *Appl Microbiol Biotechnol*, **2005**, 67(2), 170–191.

[22] Schlomann, M. Evolution of chlorocatechol catabolic pathways. Conclusions to be drawn from comparisons of lactone hydrolases. *Biodegradation*, **1994**, 5(3-4), 301–321.

[23] Zaki, S. Detection of meta - and ortho -cleavage dioxygenases in bacterial phenol-degraders. *Journal of Applied Sciences and Environmental Management*, **2006**, 10(3), 75–81.

[24] Nicaud, J.M. Yarrowia lipolytica. *Yeast*, **2012**, 29(10), 409–418.

[25] Rodriguez, G.M.; Hussain, M.S.; Gambill, L., *et al.* Engineering xylose utilization in Yarrowia lipolytica by understanding its cryptic xylose pathway. *Biotechnol Biofuels*, **2016**, 9, 149.

[26] Qiao, K.; Abidi, S.HI.; Liu, H.J., *et al.* Engineering lipid overproduction in the oleaginous yeast Yarrowia lipolyticale. *Metabolic Engineering*, **2015**, 29, 56–65.

[27] Blazeck, J.; Hill, A.; Liu, L.Q., *et al.* Harnessing Yarrowia lipolytica lipogenesis to create a platform for lipid and biofuel production. *Nat Commun*, **2014**, 5.

[28] Wang, Q.; Yurkov, A.; Goker, M., *et al.* Phylogenetic classification of yeasts and related taxa within Pucciniomycotina. *Stud Mycol*, **2015**, 81, 149–189.

[29] Zhang, S.; Ito, M.; Skerker, J.M., *et al.* Metabolic engineering of the oleaginous yeast Rhodosporidium toruloides IFO0880 for lipid overproduction during high-density fermentation. *Appl Microbiol Biotechnol*, **2016**, 100(21), 9393–9405.

[30] Schwartz, C.; Shabbir-Hussain, M.; Frogue, K., *et al.* Standardized Markerless Gene Integration for Pathway Engineering in Yarrowia lipolytica. *ACS Synth Biol*, **2017**, 6(3), 402–409.

[31] Schwartz, C.M.; Hussain, M.S.; Blenner, M., *et al.* Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR-Cas9-Mediated Genome Editing in Yarrowia lipolytica. *Acs Synth Biol*, **2016**, 5(4), 356–359.

[32] Hussain, M.; Gambill, L.; Smith, S., *et al.* Engineering Promoter Architecture in Oleaginous Yeast Yarrowia lipolytica. *Acs Synth Biol*, **2016**, 5(3), 213–223.

[33] Hussain, M.; Rodriguez, G.; Gao, D., *et al.* Recent advances in bioengineering of the oleaginous yeast Yarrowia lipolytica. *AIMS Bioengineering*, **2016**, 3(4), 493–514.

[34] Blazeck, J.; Liu, L.Q.; Redden, H., *et al.* Tuning Gene Expression in Yarrowia lipolytica by a Hybrid Promoter Approach. *Appl Environ Microb*, **2011**, 77(22), 7905–7914.

[35] Blazeck, J.; Reed, B.; Garg, R., *et al.* Generalizing a hybrid synthetic promoter approach in Yarrowia lipolytica. *Appl Microbiol Biot*, **2013**, 97(7), 3037–3052.

[36] Dulermo, R.; Brunel, F.; Dulermo, T., *et al.* Using a vector pool containing variable-strength promoters to optimize protein production in Yarrowia lipolytica. *Microb Cell Fact*, **2017**, 16(1), 31.

[37] Yaguchi, A.; Rives, D.; Blenner, M. The New Kids on the Block: Emerging Oleaginous Yeast of Biotechnological Importance. *AIMS Microbiology*, **2017**, 3(2), 227–247.

[38] Kourist, R.; Bracharz, F.; Lorenzen, J., *et al.* Genomics and Transcriptomics Analyses of the Oil-Accumulating Basidiomycete Yeast Trichosporon oleaginosus: Insights into Substrate Utilization and Alternative Evolutionary Trajectories of Fungal Mating Systems. *MBio*, **2015**, 6(4), e00918.

[39] Evans, C.T.; Ratledge, C. A Comparison of the Oleaginous Yeast, Candida-Curvata, Grown on Different Carbon-Sources in Continuous and Batch Culture. *Lipids*, **1983**, 18(9), 623–629.

[40] Görner, C.; Redai, V.; Bracharz, F., *et al.* Genetic engineering and production of modified fatty acids by the non-conventional oleaginous yeast Trichosporon oleaginosus ATCC 20509. *Green Chem.*, **2016**, 18(7), 2037–2046.

[41] Abdelaziz, O.Y.; Brink, D.P.; Prothmann, J., *et al.* Biological valorization of low molecular weight lignin. *Biotechnol Adv*, **2016**, 34(8), 1318–1346.

[42] Gong, Z.W.; Shen, H.W.; Zhou, W.T., *et al.* Efficient conversion of acetate into lipids by the oleaginous yeast Cryptococcus curvatus. *Biotechnology for Biofuels*, **2015**, 8.

[43] Meesters, P.AEP.; Huijberts, G.NM.; Eggink, G. High cell density cultivation of the lipid accumulating yeast Cryptococcus curvatus using glycerol as a carbon source. *Appl Microbiol Biot*, **1996**, 45(5), 575–579.

[44] Park, W.S.; Murphy, P.A.; Glatz, B.A. Lipid-Metabolism and Cell Composition of the Oleaginous Yeast Apiotrichum-Curvatum Grown at Different Carbon to Nitrogen Ratios. *Can J Microbiol*, **1990**, 36(5), 318–326.

[45] Anderson, J.J.; Dagley, S. Catabolism of Aromatic-Acids in Trichosporon-Cutaneum. *J Bacteriol*, **1980**, 141(2), 534–543.

[46] Gaal, A.; Neujahr, H.Y. Metabolism of Phenol and Resorcinol in Trichosporon-Cutaneum. *J Bacteriol*, **1979**, 137(1), 13–21.

[47] Gonzalez-Garcia, Y.; Hernandez, R.; Zhang, G.C., *et al.* Lipids accumulation in Rhodotorula glutinis and Cryptococcus curvatus growing on distillery wastewater as culture medium. *Environ Prog Sustain*, **2013**, 32(1), 69–74.

[48] Spanning, A.; Neujahr, H.Y. The Effect of Glucose on Enzyme-Activities and Phenol Utilization in Trichosporon-Cutaneum Grown in Continuous Culture. *J Gen Microbiol*, **1990**, 136, 1491–1495.

[49] Georgopo, S.; Zafirato, C.; Georgiad, E. Membrane Functions and Tolerance to Aromatic Hydrocarbon Fungitoxicants in Conidia of Fusarium Solani. *Physiol Plantarum*, **1967**, 20(2), 373.

[50] Huertas, M.J.; Duque, E.; Molina, L., *et al.* Tolerance to sudden organic solvent shocks by soil bacteria anal characterization of Pseudomonas putida strains isolated from toluene polluted sites. *Environ Sci Technol*, **2000**, 34(16), 3395–3400.

[51] Tsitko, I.V.; Zaitsev, G.M.; Lobanok, A.G., *et al.* Effect of aromatic compounds on cellular fatty acid composition of Rhodococcus opacus. *Appl Environ Microb*, **1999**, 65(2), 853–855.

[52] Sikkema, J.; Debont, J.AM.; Poolman, B. Interactions of Cyclic Hydrocarbons with Biological-Membranes. *J Biol Chem*, **1994**, 269(11), 8022–8028.

[53] Udaondo, Z.; Molina, L.; Daniels, C., *et al.* Metabolic potential of the organic-solvent tolerant Pseudomonas putida DOT-T1E deduced from its annotated genome. *Microb Biotechnol*, **2013**, 6(5), 598–611.

[54] Ramos-Gonzalez, M.I.; Godoy, P.; Alaminos, M., *et al.* Physiological characterization of Pseudomonas putida DOT-T1E tolerance to p-hydroxybenzoate. *Appl Environ Microb*, **2001**, 67(9), 4338–4341.

[55] Beney, L.; Gervais, P. Influence of the fluidity of the membrane on the response of microorganisms to environmental stresses. *Appl Microbiol Biot*, **2001**, 57(1-2), 34–42.

[56] Yaguchi, A.; Robinson, A.; Mihealsick, E., *et al.* Metabolism of aromatics by Trichosporon oleaginosus while remaining oleaginous. *Microb Cell Fact*, **2017**, 16(1), 206.

Chapter 4

Genomic and transcriptomic analysis of aromatic metabolism in *C. oleaginosus*

4.1 Abstract

As mentioned in **??**, lignin valorization is key for economic feasibility and success of lignocellulosic biorefineries. Lignin pre-treatments, such as alkaline hydrolysis, break down polymeric lignin into lower molecular weight compounds. The resulting hydrolysate is a mixture of compounds that contains byproducts, such as sugars and acetate, alongside the aromatic monomers. Previous chapters have covered the uptake and metabolism of model mono-aromatic compounds in defined media by *Cutaneotrichosporon oleaginosus*; however, utilization of aromatic monomers alone does not indicate a cell will survive in a lignin hydrolysate. The work discussed in this chapter describes the *C. oleaginosus* growth in and metabolism of an alkaline pre-treated corn stover lignin hydrolysate in various conditions. The yeast showed

robust metabolism of *p*-coumaric acid and slow metabolism of ferulic acid. There was no depolymerization of high molecular weight lignin nor was peroxidase or laccase activity detected in the supernatant. The slow ferulate uptake and catabolism suggests a metabolic bottleneck that could be alleviated through genetic engineering. Unfortunately, yeast genomes are notoriously understudied for these functions, so this chapter discusses a multi-omic approach used to identify putative genes acting funneling, ring-cleavage, and conversion of aromatic monomers to TCA intermediates. The discoveries made about the genes involved in aromatic catabolism in *C. oleaginosus* contributes to knowledge of yeast genetics for aromatic conversion as a whole, facilitates identification of other aromatic catabolizing yeasts through genotypic searches, and provides targets for genome engineering in *C. oleaginosus*. Collectively, the data support *C. oleaginosus*' ability to convert aromatic monomers in a lignin hydrolysate to lipids and expands our understanding of aromatic catabolism in yeast systems.

4.2 Introduction

Lignocellulosic biomass is the most abundance source of bio-polymers on Earth and a renewable source of bio-based fuels and chemicals. The majority of biomass is cellulose and hemicellulose – a source of hexose and pentose sugars. Lignin comprises 15-30% of biomass and is a highly branched, heterogeneous, and recalcitrant polymer. It is the most energy dense of all the lignocellulosic biopolymers and the most abundant renewable source of aromatic compounds [1–5]. Despite this, approximately 98% of lignin is burned for its process heat value, and most work towards valorizing biomass has focused on the polysaccharide fractions [4, 6–8]. Increased development of lignocellulosic biorefineries will generate lignin at rates surpassing its fuel demand, and technoeconomic studies have shown lignin utilization is critical for

profitability [6, 9–13]. Therefore, pursuits to identify and develop microbial strains capable of tolerating and metabolizing lignin-derived compounds into value-added products are rising.

Cutaneotrichosporon oleaginosus is an oleaginous yeast well described for its tolerance and metabolism of a wide number of sugars and toxic byproducts associated with lignocellulosic biomass pretreatment, such as furfural, acetate, and 5-hydroxymethylfurfural (5-HMF) [14–16]. Recently, it has proven to be adept at tolerating lignin-derived aromatic monomers and converting them to neutral lipids in quantities up to 70% of dry cell weight [17, 18]. This makes this yeast one of the unusual organisms that can utilize toxic byproducts and all three main depolymerization products of lignocellulosic biomass.

The heterogeneity of lignin results in mixtures of aromatic compounds of varying composition. Biological systems typically utilize funneling pathways to convert many aromatic monomers into a few common intermediates, which are then ring-opened and converted to central metabolites. These pathways are genetically well described for bacterial and fungal systems, but are explored in yeast only on a biochemical basis [19–21]. The genetic understanding of how yeast metabolize aromatic compounds remains largely unknown. Understanding the mechanisms for catabolism will facilitate rational strain engineering and develop this strain as a model for lignin conversion to oleochemicals.

In this chapter, *C. oleaginosus* was grown in an alkaline pretreated corn stover lignin (APL). It not only demonstrated uptake of aromatic monomers, *p*-coumaric acid and ferulic acid, but also established oleaginous behavior in nitrogen-limited conditions. While cells were not able to depolymerize high molecular weight lignin, they retained a high growth rate even in completely un-supplemented hydrolysate. The bottleneck observed during ferulic acid catabolism could be alleviated by rational

genetic engineering, but our ability to identify targets to pursue is greatly limited by our understanding of aromatic catabolism in yeast on a genetic basis. To address this, we pursued a multi-omic approach to identify genes involved in converting each substrate to central metabolites. Ultimately, we improved the genome annotation by over 20% in comparison to the previously published genome annotation [22]. We analyzed the transcriptome and proteome of cells grown in five aromatic monomers, APL, and glucose to identify genes active in funneling, ring-cleavage, and β-ketoadipate pathways that are unique or common across aromatic substrates. Together, these results enhance our knowledge of aromatic catabolism in yeast species and provide a critical step towards establishing *C. oleaginosus* as a model strain for lignin-derived production of biochemicals.

4.3 Materials and Methods

4.3.1 Lignin utilization and characterization

4.3.1.1 Growth on lignin and sample collection

Alkaline pre-treated lignin (APL) was prepared from corn stover biomass as previously described [23]. Lignin was titrated to a pH of 7.0 using 50% H_2SO_4. Insoluble lignin was separated by centrifugation at 10,000 x g for 10 minutes at room temperature. Soluble lignin was filter sterilized through a 0.2 μm PES bottle-top vacuum filter. Lignin hydrolysate was prepared immediately before use. APL was supplemented with 50 mM potassium phosphate buffer, 5 g/L ammonium sulfate, both, or neither. 50 mL of media was prepared in a 250 mL non-baffled flask. A single negative (non-inoculated) control for each condition was analyzed alongside inoculated cultures that were run in duplicate. Cultures were inoculated to an OD_{600}

of 0.3 and grown at 28 °C for 126 hours at 250 rpm.

Every 6 hours, 1.5 mL of culture was sterilely removed and centrifuged at maximum speed for 2 minutes. The supernatant was filtered with a 0.22 μm PES syringe filter to remove particulates. 0.6 mL of the filtered supernatant was saved at -20 °C for HPLC analysis. 0.4 mL of the filtered supernatant was saved at -20 °C for enzyme assays. 50 μL was used to measure pH of the supernatant. The cell pellet was resuspended with 1.5 mL of sterile MilliQ water and centrifuged as before. The cell pellet was resuspended to measure OD_{600} using a NanoDrop™ 2000.

To characterize the lignin profile and composition, the entire culture supernatant was collected at the end of the culturing period. Cells were separated from the aqueous media via centrifugation at 10,000 rpm for 10 minutes at 4 °C. The supernatant was lyophilized overnight in a 50 mL conical centrifuge tube. The dried powder was used in the subsequent analyses.

4.3.1.2 Lignin utilization and characterization

Analysis of aromatics by LC-MS/MS. LC-MS/MS was performed as described previously [24]. 14.5 mg of lyophilized supernatant was reconstituted in 1 mL methanol. 25 μL of sample was injected to an Agilent 1100 LC system, separated with a YMC C30 Carotenoid column (0.3 μm, 4.6 x 150 mm), and measured using a diode array detector (DAD) at 210 nm. The mobile phase consisted of (A) water modified with ammonium hydroxide (pH 7), and (B) 9:1 acetonitrile and water modified with ammonium hydroxide (pH 7). A flow rate of 0.7 mL min^{-1} and temperature of 30 °C was used throughout the run. A six-step gradient was applied to each run as follows: 0-3 min, 0% B; at 16 min, 7% B; at 21 min, 8.5% B; at 34 min, 10% B; at 46 min, 25% B; and held at 30% B at 51-56 min, for a total run time of 65 min including equilibrium. The HPLC-DAD system was connected with an Agilent Ion Trap SL MS with an

in-line electrospray ionization (ESI). The ESI nebulizer gas was set to 60 psi, with dry gas flow of 11 L min^{-1} held at 350 °C. MS and MS/MS tuned parameters are as follows: smart parameter setting with target mass set to 165 Da, compound stability 70%, trap drive 50%, capillary at 3500 V, fragmentation amplitude of 0.75 V with a 30 to 200% ramped voltage implemented for 50 msec, and an isolation width of 2 m/z (He collision gas). MS scans and precursor isolation-fragmentation scans were performed across the range of 40-350 Da in negative- and positive-ion alternating mode. Monomers measured include vanillic acid, *p*-coumaric acid, *p*-hydroxybenzoic acid, and ferulic acid. Standards were prepared in the range of 1 – 100 μg/mL and run at the same conditions as the samples. Samples were diluted accordingly to maintain detector response within the linear range of the calibration curves (R^2 value of ≥99.5%).

2D-HSQC NMR. Sample preparation for two-dimensional nuclear magnetic resonance spectroscopy (2D NMR) generally followed Mansfield et al. [25]. Briefly, 50 mgs of each sample was dissolved in $DMSO_{d6}$ and sonicated for several hours. Heteronuclear Single Quantum Coherence (HSQC) NMR spectra were acquired at 25 °C on a Bruker Avance III 600MHz spectrometer at 14.7 T using a room temperature broadband probe. Spectra were acquired with 1024 points and a SW of 12 ppm in the F2 (^{1}H) dimension and 256 points and SW of 220 ppm in the F1 (^{13}C) dimension. Peaks were assigned according to literature [25, 26]. All spectral processing was done in Bruker Topspin 3.6.1 with spectral processing parameters from Mansfield, et al. [25]. Error in integration is ± 1%.

Gel permeation chromatography (GPC). Approximately 30 mg of the lyophilized supernatant collected at the end of the culturing period was acetylated in a mixture

of pyridine (0.5 mL) and acetic anhydride (0.5 mL) at 40 °C for 24 hr with stirring. The reaction was terminated by addition of methanol (0.2 mL). The acetylation solvents were then evaporated from the samples at 40 °C under a stream of nitrogen gas, and the samples were dried at 40 °C overnight under vacuum. The dried, acetylated samples were suspended in tetrahydrofuran (THF, Baker HPLC grade), and then filtered through 0.45 μm nylon membrane syringe filters. GPC analysis was performed using an Agilent HPLC with three GPC columns (Polymer Laboratories, 300 $\times$ 7.5 mm) packed with polystyrene-divinyl benzene copolymer gel (10 μm beads) having nominal pore diameters of 10^4, 10^3, and 50 Å. An injection volume of 25 μL was used and the eluent was THF set to a flow rate of 1.0 mL/min. The GPC was attached to a diode array detector (DAD) measuring absorbance at 260 nm (band width 80 nm). Retention time was converted into molecular weight (MW) by applying a calibration curve established using polystyrene standards.

Lignin compositional analysis. Lignin compositional analysis was performed according to Sluiter, et al and NREL Laboratory Analytical Procedures (LAPs) [27]. Structural carbohydrates, lignin content, and protein content were measured according to the procedure in NREL LAP/TP-51-42618. Briefly, a two-step acid hydrolysis fractionates lignin into acid-soluble and acid-insoluble fractions. The acid-insoluble material accounts for ash and protein content, which is determined by gravimetric analysis. The acid-soluble lignin quantified by UV-vis spectroscopy. Polymeric carbohydrates are hydrolyzed into soluble, monomeric forms and are measured by HPLC-RID using a Biorad Aminex HPX-87H column. Acetyl content is measured by HPLC-RID with a Biorad Aminex HPX-87H column as well. Detailed methods are listed in Appendix E.

4.3.1.3 Ligninolytic enzyme assays

Ligninolytic enzyme assays were performed in 96 well plates using filtered cell culture supernatant which were collected as described earlier. Positive controls were filtered cell culture supernatant from an in-house grown *Trametes versicolor* culture. Enzyme assays were performed as described by Salvachúa, et al. [24]. Activity was measured in the culture supernatants by the oxidation of 5 mM 2,6-dimethoxyphenol (2,6-DMP) to dimeric cerulignone (ε_{496} = 55,000 M^{-1} cm^{-1}), 5 mM ABTS to its cation radical (ε_{418} = 36,000 M^{-1} cm^{-1}), or 5 mM veratryl alcohol to veratryl aldehyde (ε_{310} = 9,300 M^{-1} cm^{-1}) in 0.1 mM sodium tartrate buffer at pH 4. To measure peroxidase activity, 0.1 mM of freshly prepared H_2O_2 was included to catalyze the reaction. The peroxidase assays were performed in the presence and absence of Mn^{2+} (0.1 mM $MnSO_4$). Absorbance from peroxidase activity was corrected with that caused by laccase activity. Measurements were carried out at room temperature. One unit of activity (1 U) is defined as the amount of enzyme releasing 1 μmol of product per minute under the defined reaction conditions.

4.3.2 RNA extraction and transcriptomic analysis

4.3.2.1 Cell culturing

Cutaneotrichosporon oleaginosus cells were grown in a YPD (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose) culture overnight. Cells were washed three times with sterile milliQ water by centrifuging cells at 6,000 x g for 2 min, decanting supernatant, and resuspending in sterile milliQ water. TOCM media was made according to Table 4.1. 15 mL of media was prepared in 50 mL baffled flasks. Washed cells were inoculated to an OD600 of 0.3 and flasks were left to shake at 250 rpm at 28 °C.

Table 4.1. Media composition for TOCM media.

Component	Concentration
Carbon source	1.0 g/L
Yeast extract	1.5 g/L
$MgSO_4 \cdot 7\,H_2O$	0.4 g/L
KH_2PO_4	2 g/L
$MnSO_4 \cdot 4\,H_2O$	0.003 g/L
$CuSO_4 \cdot 5\,H_2O$	0.0001 g/L

4.3.2.2 RNA extraction and RNAseq library preparation

TOCM samples grown on 1 g/L model aromatic compounds or glucose were harvested at mid-exponential phase and normalized to 1 mL harvested per OD_{600} of 10. To harvest, cells were centrifuged at 8,000 x g for 2 minutes at room temperature. Cells were washed twice with sterile water by decanting supernatant and centrifugation at the same conditions. Cells were snap-frozen in liquid nitrogen and stored at -80 °C until extraction.

The first set of RNA extractions consisted of glucose, pHBA, phenol, and resorcinol. Biological replicates (n=2) were extracted using the E.Z.N.A. Yeast RNA kit from Omega Biotek, with the following modification: upon addition of glass beads, cells were vortexed using a Mini-BeadBeater-16 (Biospec Products). Cells were homogenized for 8 cycles, at 1.5 min per cycle, with 2 min breaks between each cycle. Homogenization was performed in a cold room. On-column DNase digestion was performed for all samples. RNA was eluted in molecular grade water, flash frozen in liquid nitrogen, and stored at -80 °C. Library preparation and Illumina sequencing was performed by the Clemson University Genomics Institute (CUGI). Raw reads were provided for DGE analysis.

The second set of RNA extractions consisted of *p*-coumaric acid, ferulic acid, and APL samples with their own glucose control. Biological replicates (n=4) were extracted using Lucigen's MasterPure™ Yeast RNA Purification Kit. The optional DNAse treatment was performed to eliminate DNA contamination. RNA was eluted in molecular grade water, flash frozen in liquid nitrogen, and stored at -80 °C. Library preparation was performed by Dr. Adam Guss' lab at Oak Ridge National Laboratory (ORNL). Briefly, RNA concentration was quantified using a Nanodrop 1000 instrument (ThermoScientific, Waltham, MA) and RNA quality was verified by obtaining RNA Integrity Numbers (RINs) using an RNA 6000 Nanochip on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Ribosomal RNA was depleted from total RNA samples using a RiboZero rRNA Removal Kit for Yeast (Illumina-Epicentre Inc. San Diego, CA) according to manufacturer's instructions. The depleted sample was purified on a RNA Clean & Concentrator-5 (Zymo Research, Irvine, CA, USA) following the manufacturer's protocol, and then the depleted material was quantified using a Nanodrop 1000 and visualized on an Agilent 2100 Bioanalyzer instrument with a RNA 6000 Nanochip (Agilent Technologies, Santa Clara, CA). RNA depleted of ribosomal RNA was used as input material to synthesize cDNA libraries using a ScriptSeq v2 RNA-Seq Library Preparation Kit (Illumina-Epicentre, San Diego, CA, USA) according to manufacturer's instructions and TruSeq compatible barcodes. The final RNA Seq libraries were quantified with a Qubit Fluorometer (Invitrogen, CA) and library quality was assessed on a Agilent Bioanalyzer using a DNA 7500 chip (Agilent, CA). The final barcoded libraries were pooled and paired end sequencing (PE150) was conducted on an Illumina NovaSeq 6000. Sequencing and de-multiplexing were completed as a service provided by Vanderbilt University Medical Center (VUMC VANTAGE Vanderbilt Technologies for Advanced Genomics (Nashville, TN)). Raw reads were provided for DGE analysis.

4.3.2.3 RNAseq pipeline and functional annotation

Raw paired end reads for the phenol, pHBA, resorcinol, and glucose datasets were obtained from the Clemson University Genome Institute (CUGI) and raw paired end reads for the *p*-coumaric acid, ferulic acid, and APL datasets were obtained from Vanderbilt University Medical Campus (VUMC). Illumina adapters and low quality reads were removed using Trimmomatic v.0.30 and v.0.38 [28], respectively, using a sliding window of 4:15. Low quality reads were those below 36 bp and with a leading and trailing threshold of 3. The trimmed reads for the phenol, pHBA, resorcinol, and glucose datasets were input to Necklace [29] using the JGI genome for *Cutaneotrichosporon oleaginosus* IBC0246 and *Yarrowia lipolytica* CLIB122 provided on NCBI served as reference genomes [22, 30]. The program creates a *de novo* assembly using Trinity, aligns the reads to it using HISAT2 v. 2.0.5 [31] and generates count tables using featureCounts v. 1.6.3 [32]. The count tables were upper quartile normalized using in-house Python scripts (see Appendix C) and then analyzed for differential gene expression using EdgeR [33]. Differential expression was denoted as logFC $\geq |1|$ and $p_{adj} \leq 0.05$. In-house Python scripts were used to generate hierarchical clustering, PCA plots, and volcano plots (see Appendix C). DGE analysis was also manually curated. The *de novo*, reference-guided assembly was annotated with Trinotate [34], KEGG Annotation Analysis Software (KAAS) [35], and TopGO [36]. Annotation completeness was determined by BUSCO [37].

Transcriptomic datasets were verified with reverse transcription quantitative PCR (RT-qPCR) using Quanta's qScript One-Step qRT-PCR Kit (cat no. 95057-200). Gene-specific primers were designed and tested to ensure efficiency between 90-110% before use (Table 4.2). The annotated β-actin gene (CC85DRAFT_288370) served as the reference gene. All experiments were performed in biological replicates

as described above in 96 well plates (Bio-Rad cat no. HSP9601) using the CFX Connect Real-Time PCR Detection System (Bio-Rad).

Table 4.2. Table of primers used for qPCR analysis.

Primer	**Sequence (5' → 3')**
150900_F	GGCACCACCATCGTCCGCCA
150900_R	CGGCTTGCCATCGGCGTCGT
240134_F	TCAAATCCGAGCAGGTGTGCC
240134_R	GGCGTGAAGCGGGAGACGAA
311359_F	ACGGAGGGCAAGCCGATCCCGA
311359_R	TGTCCGAGTAGAGCCGCCCGC
271894_F	ACCGTCCACAAGATGGCGACCG
271894_R	GCACAGGAGGCCGAGGTGGCG
282431_F	CGCCATCTCGGGCATCACAAACC
282431_R	GTGCGTCTGAGCGTGCGGGA
274445_F	TGTGGCTTGGCGGGCTCACT
274445_R	CGAACGATGGGTGGATGCTCTCTCC
287490_F	GCGTCTTCATTGCTGGCGA
287490_R	AGCCGAGGTTGTAGGTGTCCA
288370_F	GCCGACACTGGCCTCACCGGG
288370_R	TCGCACCAGCGACCTTGGCG

4.3.3 Protein extraction and proteomic analysis

Samples were grown in the same manner discussed above for RNA extraction. For *p*-coumaric acid, ferulic acid, and APL samples, the same cultures were used for transcriptomics and proteomics, but phenol, pHBA, resorcinol, and glucose samples, separate cell pellets were obtained. Proteomic samples were submitted to the Multi-User Analytical Lab (MUAL) & Metabolomics Core at Clemson University as cell pellets. Samples stayed at -80 °C until extraction. Samples were lysed in 500 μL of lysis buffer (5% SDS buffer containing 100 mM triethylammonium bicarbonate, adjusted to a pH of 7.55 with H_3PO_4 containing 20 mM TCEP reducing agent and 5% SDS) and sonicated at 25% amplitude for 50 seconds. Sonicated cells were incubated

at 95 °C for 10 minutes. 50 μL of 400 mM IAA alkylating agent was added and incubated in the dark for 30 minutes. Protein concentration was estimated using Pierce® BCA Protein Assay Kit - Reducing Agent Compatible. Cell lysate was loaded to a filter-aided spin column to an equivalent of 120 μg of protein. The samples were centrifuged and washed 8 times with binding buffer (90% aqueous methanol containing 100 mM TEAB adjusted to pH of 7.1). Trypsin and ProteaseMax (0.2%) was added at a tyrpsin:protein ratio of 1:20 and incubated in a 37 °C water bath overnight. Tryptic digests were eluted with 50 mM ABC, 0.1% formic acid, and 50% acetonitrile. Eluates were dried under vacuum and reconstituted in 0.1% formic acid.

A total of 2 μg of protein was injected onto an UltiMate™ 3000 RSLCnano (ThermoFisher) and separated on an EASY-Spray PepMap RSLC C18 column (2 μm, 100A, 75 μm x 50 cm. Samples were injected with a loading buffer of 3% aqueous acetonitrile. The mobile phase consisted of water + 0.1% aqueous formic acid (A) and 80% acetonitrile + 0.1% formic acid (B) and was run with a flow rate of 0.3 μL/min and 35 °C throughout the run. A four-step linear gradient of 4% buffer B for 100 minutes, 4-28% in 20 minutes, 28-50% B in 5 minutes, 50-95% B in 20 minutes, followed by equilibration with 4% B was applied for each run. The nanoLC system connected with an Orbitrap Fusion™ Tribrid™ Mass Spectrometer (ThermoFisher). The Orbitrap was run using a positive polarity, positive nanospray ionization (NSI) of 2200 V, RF lens value of 60%, and ion transfer tube temperature of 275 °C. For the first MS1 spectra, the Orbitrap detector was set to a resolution of 50,000, AGC target value of 400,000, and mass range of 375-1500 m/z. For the second MS2 spectra, the AGC target was set to 10,000 and the scan range was automatically adjusted. Fragmented ions were generated using a collision energy of 35% and isolated through a quadrupole with an activation time of 10 ms and Q of 0.25. Data was analyzed by MUAL using ProteomeDiscover software (ThermoFisher) and through manual

curation. Plots were generated using an in-house Python script (see Appendix D).

4.3.4 Yeast lysate activity assays

C. oleaginosus cells were grown in 2 mL YPD overnight. Cells were washed with sterile water and used to inoculate 15 mL of TOCM media containing 1 g/L phenol, pHBA, resorcinol, or glucose to an initial OD of 0.3. Cells were grown to mid-exponential phase and then harvested by centrifugation at 8,000 x g for 2 minutes at room temperature. Cell pellets were resuspended in 50 mM Tris-HCl buffer at pH 7.5 and lysed by bead milling in a Mini-BeadBeater-16 (Biospec Products). Cells were homogenized for 8 cycles, at 1.5 min per cycle, with 2 min breaks between each cycle. Homogenization was performed in a cold room. Cells were centrifuged to remove cellular debris. Supernatant was transferred to a fresh 1.7 mL tube and stored at 4 °C until use. Clarified yeast lysates were not stored longer than 2 weeks. Protein concentration was measured on Nanodrop 2000.

Catechol cleavage assays Catechol cleavage assays were based on [38]. To measure intradiol cleavage, the following were added to a quartz cuvette: 2 mL 50 mM Tris-HCl (pH 8.0), 0.7 mL distilled water, 0.1 mL 100 mM 2-mercaptoethanol, and 0.1 mL cell-free extract. The contents were mixed by inversion. 0.1 mL 1 mL catechol was added and mixed again then immediately measured on a Nanodrop 2000. Absorbance was measured at 260 nm over 5 mins at room temperature. To measure extradiol cleavage, the following were added to a quartz cuvette: 2 mL 50 mM Tris-HCl (pH 7.5), 0.6 mL distilled water, and 0.2 mL cell extract. The contents were mixed by inversion. Then 0.2 mL 100 mM catechol was added and mixed again, then immediately measured on a Nanodrop 2000. Absorbance was measured at 375 nm over 5 mins at room temperature.

Protocatechuate cleavage assay Protocatechuic acid (PCA) cleavage assays were based on [39]. To measure intradiol cleavage, 50 mM Tris-HCl (pH 8.5), cell extract, and 160 μM PCA were mixed to a total volume of 0.3 mL in a quartz cuvette. Intradiol cleavage was measured at 290 nm and extradiol cleavage was measured at 410 nm over 5 mins at room temperature in a Nanodrop 2000.

Resorcinol hydroxylation assay Resorcinol hydroxylase assays were based on [40]. To measure resorcinol hydroxylation, mix 2.9 mL of 20 mM phosphate buffer (pH 6.5), 40 μL of 25 mM NADPH, 20 μL of 25 mM resorcinol, and 5 μL of cell lysate. Mix contents by inversion and measured on a Nanodrop 2000. Absorbance was measured at 340 nm over 5 mins at room temperature.

Hydroxyquinol cleavage assay Hydroxyquinol cleavage assays were based on [41]. To measure hydroxyquinol cleavage, the following were mixed to a total of 1 mL in a quartz cuvette: 10 mM phosphate buffer (pH 7.2), 0.2 mM hydroxyquinol, 0.5 mM $FeSO_4$, and cell extract. Intradiol cleavage was measured at 243 nm over 5 mins at room temperature in a Nanodrop 2000.

4.3.5 Functional genomics

4.3.5.1 *In vitro* translation

The RNA extracted above from cells grown in phenol, resorcinol, pHBA, and glucose were pooled and used as template for cDNA synthesis using the quanta cDNA Flex Kit by QuantaBio. Up to 1 μg of RNA was added to a PCR tube with 2 μL of each random primer and oligo dT. The solution was vortexed to mix, centrifuged for 10 seconds in a tabletop centrifuge, incubated at 65 °C for 5 mins, then snap chilled on ice. Flex reaction mix and reverse transcriptase was added to each tube,

then mixed and centrifuged as before. The reaction was incubated at 25 °C for 10 mins, 42 °C for 90 mins, 85 °for 5 mins, then cooled to 4 °C. cDNA was purified using Qiagen's PCR Purification kit and eluted into 30 μL of 10 mM Tris-HCl, pH 8.5. Concentration was measured on a Nanodrop 2000.

All four cDNA libraries were pooled together and used in PCR reactions designed to add the necessary T7 promoter, RBS, and terminator for the PURExpress® In Vitro Protein Synthesis Kit from NEB. Primers and genes amplified are listed in Table 4.3. Of the listed genes, CC85DRAFT_311359, 282345, 326858, and 150900 are catechol 1,2-dioxygenases; the rest were phenol dioxygenases. 2 μL cDNA template, Q5 Hot Start 2X Master Mix from NEB, and gene specific primers were used in each PCR reaction. Genes were all amplified at 65 °C for 1 min 15 sec and visualized on an agarose gel.

The Q5 PCR amplified DNA was used as template in the PURExpress® kit. 250 ng of template DNA was added to each reaction. The *in vitro* translation was performed according to manufacturer recommendation. Translation occurred at 37 °C for 3 hours in an incubator and the reaction was stopped by placing the reaction on ice. The reaction was kept at 4 °C until use. Proteins were stored no longer than 2 weeks. To test whether proteins required cofactors to fold correctly, phenol hydroxylases were tested for translation in presence of $FADH_2$ and NAD(P)H at final concentrations of 10 μM each, and the catechol dioxygenase was translated in the presence of $FeSO_4$ or $MnSO_4$ at a final concentration of 200 μM.

Table 4.3. Table of primers used to amplify genes for *in vitro* translation for functional genomics.

Primer	**Sequence (5' → 3')**
Phenol hydroxylases	

Continued on next page

Table 4.3 – *Continued from previous page*

Primer	**Sequence (5' → 3')**
240134_F	GCGAATTAATACGACTCACTATAGGGCTTA AGTATAAGGAGGAAAAAATCATATGATGAC CAAGGAGATTGAGACCTACT
240134_R	AAACCCCTCCGTTTAGAGAGGGGTTATGCT AGGCTCAGCCTACACTGCCGCAGCCACAG
287490_F	GCGAATTAATACGACTCACTATAGGGCTTA AGTATAAGGAGGAAAAAATCATATGATGAC CAAGTACTCCGAGTC
287490_R	AAACCCCTCCGTTTAGAGAGGGGTTATGCT AGGCTCAGCTTAGACAAAGACCTTGGTCC
284368_F	GCGAATTAATACGACTCACTATAGGGCTTA AGTATAAGGAGGAAAAAATCATATGATGCC AGCTAAGCAGGCGCCT
284368_R	AAACCCCTCCGTTTAGAGAGGGGTTATGCT AGGCTCAGCCTAGTACTGCCGGCCGAGCGT
316289_F	GCGAATTAATACGACTCACTATAGGGCTTA AGTATAAGGAGGAAAAAATCATATGATGCC AGCATTCCAGGACGG
316289_R	AAACCCCTCCGTTTAGAGAGGGGTTATGCT AGGCTCAGCTCACACCTTGCGGCCCAGCA
287222_F	GCGAATTAATACGACTCACTATAGGGCTTA AGTATAAGGAGGAAAAAATCATATGATGTC TCCTAGCACCAACCC
287222_R	AAACCCCTCCGTTTAGAGAGGGGTTATGCT AGGCTCAGCTTATGCGCTGAGAGCGGCGT
Catechol 1,2-dioxygenases	
287490_F	GCGAATTAATACGACTCACTATAGGGCTTA AGTATAAGGAGGAAAAAATCATATGATGTC CAGCTTCGACTTCTC
287490_R	AAACCCCTCCGTTTAGAGAGGGGTTATGCT AGGCTCAGCTTACGCCTTCTGGAGGTGGA
284368_F	GCGAATTAATACGACTCACTATAGGGCTTA AGTATAAGGAGGAAAAAATCATATGATGCC CAAGGTCTCCGACGT
284368_R	AAACCCCTCCGTTTAGAGAGGGGTTATGCT AGGCTCAGCCTAAGCCTTGACCTCTGCCTT
316289_F	GCGAATTAATACGACTCACTATAGGGCTTA AGTATAAGGAGGAAAAAATCATATGATGGT CGAGGCGGACATTTCGACCGAG

Continued on next page

Table 4.3 – *Continued from previous page*

Primer	**Sequence (5' → 3')**
316289_R	AAACCCCTCCGTTTAGAGAGGGGTTATGCT AGGCTCAGCctaagccgccgccgGCAGCC
287222_F	GCGAATTAATACGACTCACTATAGGGCTTA AGTATAAGGAGGAAAAAATCATATGATGCT CTCCCTCACCGTCCTCCTT
287222_R	AAACCCCTCCGTTTAGAGAGGGGTTATGCT AGGCTCAGCCTACATTCGCGCGAAGCGAAG

Translations were analyzed using SDS-PAGE. Raw, unpurified translation reactions were boiled at 100 °C for 10 min with 2X Laemmli Sample Buffer (2.5 μL sample, 2.5 μL water, 5 μL loading dye). All 10 μL of sample were run on an Invitrogen NuPAGE 4-12% Bis-Tris gel using 1X MOPS as a running buffer for 40 mins at 200 V. Ladder was 10 μL of ThermoFisher PageRuler Prestained Ladder. After electrophoresis, gels were rinsed with diH_2O for 3 mins to remove buffer. Gels were stained using the Pierce™ Power Stainer by ThermoFisher. One filter pad was soaked with Power Stain Solution (cat. 1902206) and one filter pad was soaked with Destained Solution (cat. 190227). The sandwich was assembled on the Power Stainer such that the stain filter pad was on top, the destain filter pad was on bottom, and the gel was between the two.

PURExpress proteins may have been purified to remove the translation system proteins, which were His-tagged. To do so, proteins were purified using Ni-NTA resin columns from G-Biosciences (cat. 786-943) according to manufacturer directions. *Note: After protein is loaded to the Ni-NTA resin, collect the flow-through fraction, as the translation system is His-tagged and bound to the column.* The protein elution (flow-through after loading) was buffer exchanged using an Amicon Ultra-0.5 Centrifugal Filter Unit with a 3K nomial molecular weight limit (NMWL) (cat. UFC500324). The buffer used for exchange was Buffer A, described in section 4.3.5.3.

4.3.5.2 *E. coli* expression

Genes were cloned into pET21b(+) for expression in *E. coli* BL21(DE3) or NEB T7SHuffle Express cells. Cells were grown in LB+Amp overnight, washed with sterile water, and then used to inoculate 250 μL of autoinduction media containing ampicillin in 48 well plates (Nunclon® cat. #150687) sealed with Breathe-Easy breathable film (Diveresified Biotech cat. #BEM-1). When required, cells were supplemented with $FeSO_4$, $MnSO_4$, or $FeCl_3$ at a final concentration of 200 μM. To make autoinduction media, mix 25 mL of 20X P, 10 mL of 50X 5052, 100 μL of 1000X Trace Metals, 1 mL of 1 M $MgSO_4$, 10 μL of 20 mM Na_2SeO_3, 12.5 mL 1 M sodium succinate, and 15 mL 50% glycerol (v/v). Recipe for 20X P, 50X 5052, and 1000X Trace Metals are listed in Table 4.4. *Note: Na_2SeO_3 is highly toxic.* In the case of non-induced conditions, remove lactose from the AI media formulation. In the case of IPTG induction, remove lactose from the AI media and when cells reach mid-exponential phase, induce with 0.1 mM IPTG. Cells were incubated until late stationary phase and then harvested by transferring cells out of the 48-well plate to a 1.7 mL tube. Cells were centrifuged, the supernatant was decanted, and cells were stored at -20 °C. In all cases, mCherry served as a control for both growth conditions and cell lysis conditions.

Table 4.4. Composition of the autoinduction media used for protein in expression from *E. coli* for enzyme activity assays.

Chemical	**Conc.**
20X P	
Na_2HPO_4	1 M
KH_2PO_4	1 M
$(NH_4)_2SO_4$	0.5 M
50X 5052	
Glycerol	25% (v/v)

Continued on next page

Table 4.4 – *Continued from previous page*

Chemical	**Conc.**
Glucose	2.5% (w/v)
α-lactose monohydrate	10% (w/v)
1000X Trace Metals in 60 mM HCl	
$FeCl_3$	50 mM
$CaCl_2$	20 mM
$MnCl_2$	10 mM
$ZnSO_4$	10 mM
$CoCl_2$	2 mM
$CuCl_2$	2 mM
$NiCl_2$	2 mM
Na_2MoO_4	2 mM
H_3BO_3	2 mM

Both soluble and insoluble protein fractions were evaluated for protein production. 1 mL of cells were freeze-thawed (freeze 15 min at -80 °C, thaw at 37 °for 10 min) for several cycles until it was seen that mCherry was released from the cell pellet. In this case, when cells were spun down at max speed for 5 minutes at room temperature, the cell debris pellet remained white whereas the soluble fraction turned magenta. The cell debris (insoluble fraction) was resuspended by vortexing in 1 mL of Tris-HCl, pH 7 containing 1% SDS. The insoluble fraction was then boiled for 20 mins at 95 °C. 10 μL of this was mixed with 10 μL of 2X SDS Loading Dye and then boiled again at 95 °C for 10 mins. To make 2X SDS Loading Buffer, mix 2.5 mL 1 M Tris-HCl (pH 7), 0.5 mL ddH_2O, 1 g SDS, 0.8 mL 0.1% bromophenol blue, 4 mL glycerol, 2 mL BME, and adjust volume to 20 mL with ddH_2O). Store as 1 mL aliquots at 4 °C. All 20 μL of prepared samples were run at 180 V for 40 minutes on a 4-20% Mini-PROTEAN® TGX™Precast Protein Gels (cat. 4561096) using a Tris/Glycine/SDS running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3). Gels were stained using Genscript's eStain L1 Staining Solution (M00548-1), destained with milliQ water overnight, and imaged with a Biorad ChemiDoc XRS+

System.

To generate soluble fraction proteins for enzymatic assays, cells were harvested at late-stationary phase as above and cell lysates were generated by freeze-thaw cycles (freeze 15 min at -80 °C, thaw at 37 °for 10 min). Cell debris was removed by centrifugation at max speed for 5 minutes at room temperature. Clarified cell lysates were pipetted into a fresh 1.7 mL tube and stored at 4 °C until use. Cell lysates were stored no longer than 2 weeks.

4.3.5.3 Phenol hydroxylase and catechol dioxygenase activity assays

Enzymatic assays for phenol hydroxylase and catechol dioxygenase are described in Table 4.5. Buffer A was comprised of 50 mM sodium phosphate and 100 mM sodium chloride at pH 7.5. NADPH, NADH, and $FADH_2$ stocks were made in 20 mM Tris-HCl, H 7.5. $FADH_2$ was stored at -20 °C, while NADPH and NADH were stored at 4 °C. All substrates were made in Buffer A and stored at 4 °C. For the phenol hydroxylase, substrates tested included phenol, pHBA, and resorcinol. For the catechol dioxygenase, substrates tested included catechol, protocatechuic acid, and hydroxyquinol, and cofactor tested included $FeSO_4$ and $MnSO_4$. Negative controls were composed of identical conditions, except Buffer A replacing the volume of protein. The source of protein was raw PURExpress translation reaction, purified PURExpress translation reaction, and clarified *E. coli* lysate.

To test enzyme activity, all components of the reaction were mixed on ice in a Corning™ UV-transparent 96-well plate (cat. 07-200-623) in the following order: Buffer A, cofactor, $FADH_2$, substrate, and NAD(P)H. The protein was loaded to each well last and immediately before measuring. The plate was set to mix by shaking in the plate reader. Phenol hydroxylase activity was monitored by NADH consumption at 340 nm and NADPH consumption at 320 nm. Intradiol ring cleavage of catechol

was monitored at 260 nm and extradiol ring cleavage was monitored at 350 nm. Intradiol cleavage of protocatechuate was monitored at 290 nm and extradiol cleavage was monitored at 410 nm. Intradiol cleavage of hydroxyquinol was monitored at 243 nm.

Table 4.5. Composition of enzymatic assays for phenol monooxygenase and catechol dioxygenase activity.

Component	**Monooxygenase Assay**	**Dioxygenase Assay**
Buffer A	167.5 μL	172.5 μL
Protein	2.5 μL	2.5 μL
$FADH_2$ (400 mM)	5 μL	-
NAD(P)H (8 mM)	5 μL	-
Substrate (10 mM)	20 μL	20 μL
Cofactor (8 mM)	-	5 μL

4.4 Results

4.4.1 Growth in alkaline pre-treated corn stover lignin

4.4.1.1 Growth and monoaromatic uptake

The starting composition of the alkaline pre-treated corn stover lignin hydrolysate (APL) is listed in Table 4.6. Cells were grown in APL with and without the presence of a 50 mM phosphate buffer as well as with and without nitrogen supplementation. Cells grew best when the media was buffered, but accumulated more lipids in nitrogen-depleted media (Figure 4.1A-B, Table 4.7). As carbon to nitrogen ratio controls lipogenesis, this was expected. The lipid profile did not change across the different culturing conditions, with a majority of the lipids being oleic acid (Table 4.8). In addition, the lipid profile did not change across time, as Figure 4.2 shows

the lipid profile at 24, 60, and 126 hours. This data was collected every 12 hours and subsequently analyzed, but only these three time points are shown for clarity. The major aromatic monomers released during lignin pre-treatment were *p*-coumaric acid and ferulic acid. Both compounds and catabolic intermediates — vanillic acid and pHBA — were detected through HPLC (Figure 4.1C). In nitrogen supplemented cultures, *p*-coumaric acid was depleted within 24 hours with and without a buffer. In cultures lacking ammonium sulfate, the unbuffered culture showed complete uptake of *p*-coumaric acid by hour 102, but the buffered culture did not show full uptake even by the end of the culturing period. None of the cultures showed accumulation of pHBA, an intermediate of the *p*-coumarate metabolism, so it is unlikely there is a pathway bottleneck. None of the cultures showed complete uptake of ferulic acid, but the best uptake was measured in cell cultures grown in unbuffered media supplemented with ammonium sulfate. All culture conditions showed an accumulation of vanillic acid, an intermediate of ferulic acid metabolism, suggesting a pathway bottleneck. The negative controls, identical culture preparation except for cell inoculation, did not show the same decrease in *p*-coumaric acid or ferulic acid. GPC supports the uptake of monomeric aromatic compounds, by the decrease of the low molecular weight peak in the cultures containing cells (Figure 4.5). Cultures not containing cells did not show the same decrease of low molecular weight monomers. The supernatants from the cell cultures supplemented with ammonium sulfate were analyzed with ^{1}H-^{13}C HSQC 2D-NMR. The resulting spectra support the HPLC data that the cells utilized the p-coumaric acid (Figure 4.3, pink box). The changes to S and G lignin motifs as compared to the negative controls suggest structural alterations to the lignin may be occurring, but the precise changes remain obscure.

Table 4.6. Composition of APL hydrolysate derived from corn stover.

Compound	**Conc. (g/L)**
p-coumaric acid	1.87
Ferulic acid	0.49
Glucose	0.50
Xylose	2.57
Arabinose	1.37
Galactose	0.35
Lactic acid	0.25
Acetic acid	3.41

Table 4.7. Biomass titer, lipid titer, and lipid accumulation from cultures grown in alkaline pre-treated lignin. Error are standard deviation of biological replicates (n=2).

Supplement	**Biomass titer (g/L)**	**Lipid titer (g/L)**	**Lipid accum. (%)**
+Buff,+AS	1.077 ± 0.113	0.074 ± 0.011	9.17 ± 2.69
+Buff,-AS	1.122 ± 0.073	0.206 ± 0.014	20.84 ± 2.10
-Buff,+AS	0.853 ± 0.024	0.044 ± 0.003	7.91 ± 1.47
-Buff,-AS	1.152 ± 0.012	0.242 ± 0.019	31.77 ± 4.49

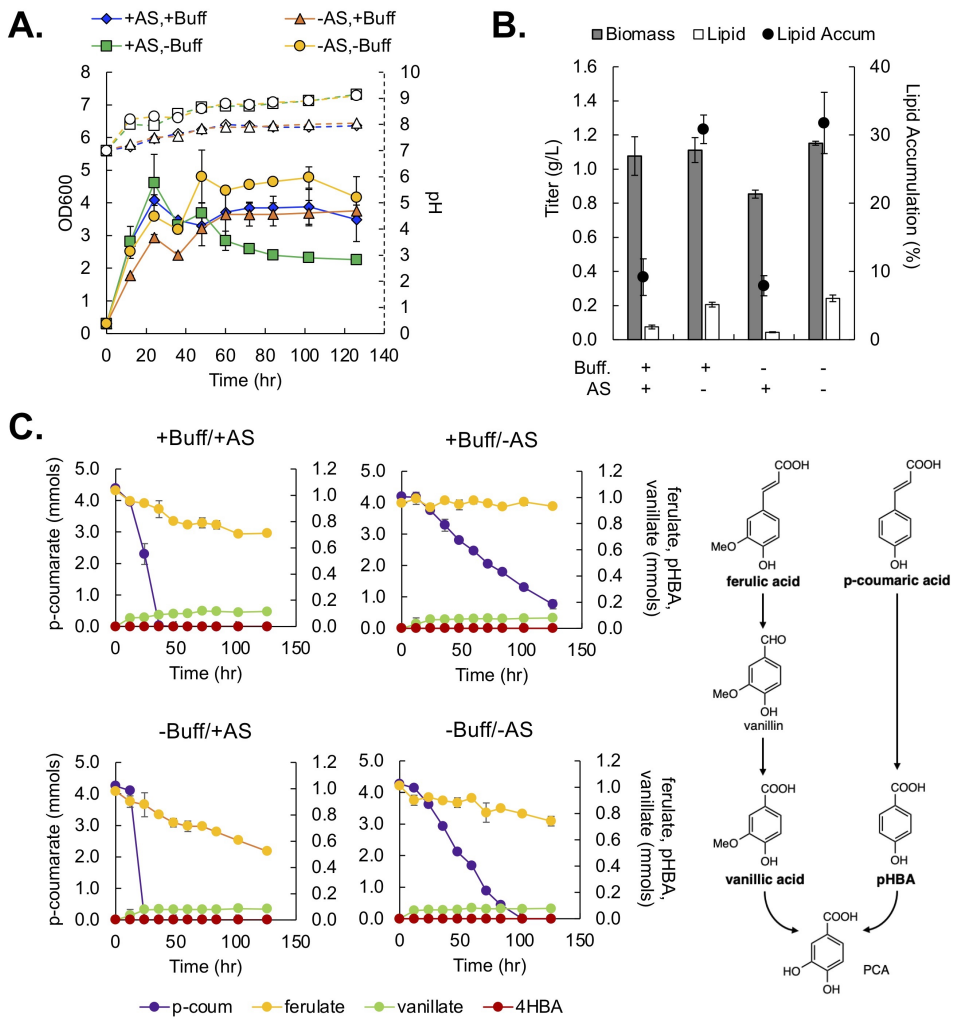


Figure 4.1. A. Growth (solid lines and filled markers, left-hand axis) and pH (dashed lines and open markers, right axis) over time for cultures containing buffer and $(NH_4)_2SO_4$ (blue diamond), buffer and no $(NH_4)_2SO_4$ (orange triangle), no buffer with $(NH_4)_2SO_4$ (green square), and no buffer and no $(NH_4)_2SO_4$ (yellow circle); B. biomass (gray bar, left axis), lipid titers (white bar, left axis), and lipid accumulation (black circle, right axis); and C. monoaromatic concentration of *p*-coumarate (purple, left axis), ferulate (yellow, right axis), vanillate (green, right axis), and pHBA (red, right axis) for *C. oleaginosus* cells grown in APL. The metabolic schematic included motivate the substrates chosen for HPLC detection, as pHBA is an intermediate of *p*-coumarate metabolism and vanillic acid is an intermediate of ferulate metabolism. The data in A-C are the mean and error bars are standard deviation of biological replicates (n=2). Lines are used only for visual aid.

Table 4.8. Fatty acid composition profile (%) of cells grown in alkaline pre-treated lignin. Errors are the standard deviation of biological replicates (n=2).

Supplement	**C16:0**	**C16:1**	**C18:0**	**C18:1**	**C18:2**	**C18:3**
+Buff,+AS	5.1±0.6	17.2±0.3	5.9±0.1	42.7±1.5	29.0±0.7	0.0±0.0
+Buff,-AS	1.9±0.1	19.8±0.1	14.3±0.1	45.4±0.0	17.3±0.3	1.4±0.0
Buff,+AS	8.9±0.5	18.5±0.9	5.9±0.6	40.3±0.3	26.5±	0.0±0.0
-Buff,-AS	1.9±0.3	19.4±0.3	15.7±0.3	45.2±0.4	16.4±0.5	1.3±0.0

The culture supernatant was harvested for lignin compositional analysis (Table 4.9). The cultures with buffer show a higher ash content, as is to be expected as the buffer added salt content. The buffered samples also show a decrease in lignin, glucose, xylose, and arabinose content. The samples supplemented with ammonium sulfate show higher protein content, which is expected since nitrogen is required for robust protein synthesis. The method for protein quantification relies on detection of nitrogen, so presence of any unconsumed ammonium sulfate in the media may have skewed this number. Across the culture conditions, there is a significant depletion of acetate in the replicates inoculated with cells versus the negative controls. This indicates cell growth was supported by acetate consumption.

4.4.1.2 Depolymerization of polymeric lignin

The polymeric lignin quantity and profile was evaluated from the supernatant at the end of the culturing period using Klason lignin analysis, GPC, and 2D-NMR. According to the GPC spectra, there was little to no depolymerization of the higher

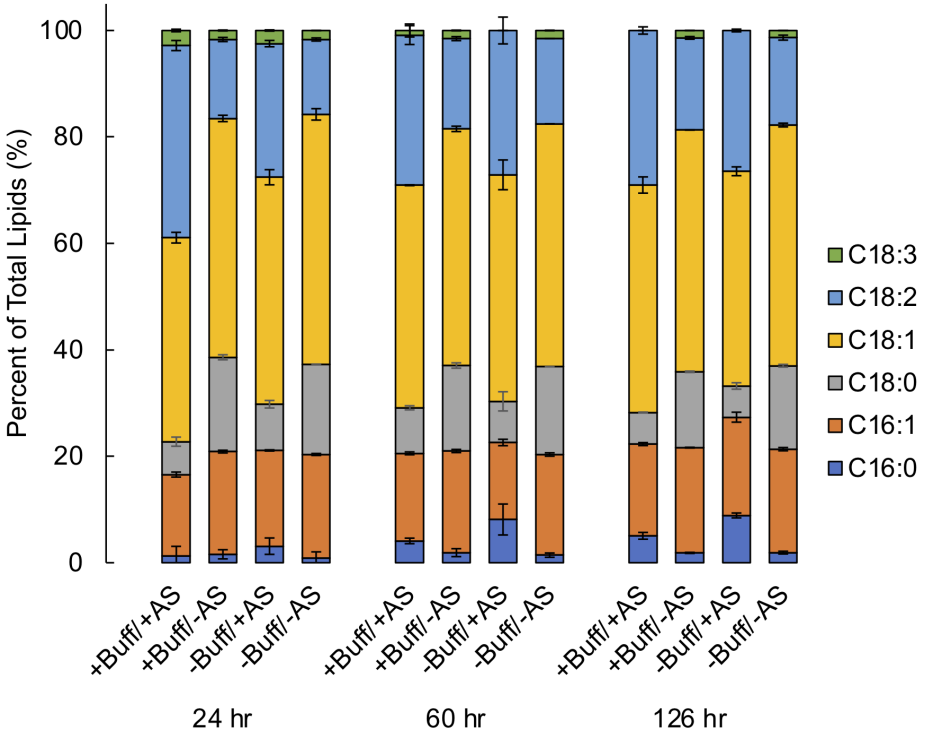


Figure 4.2. Lipid distribution profile of *C. oleaginosus* cells grown in supplemented and unsupplemented APL at 24, 60, and 126 hour time points. Standard deviation are error of the mean (n=2).

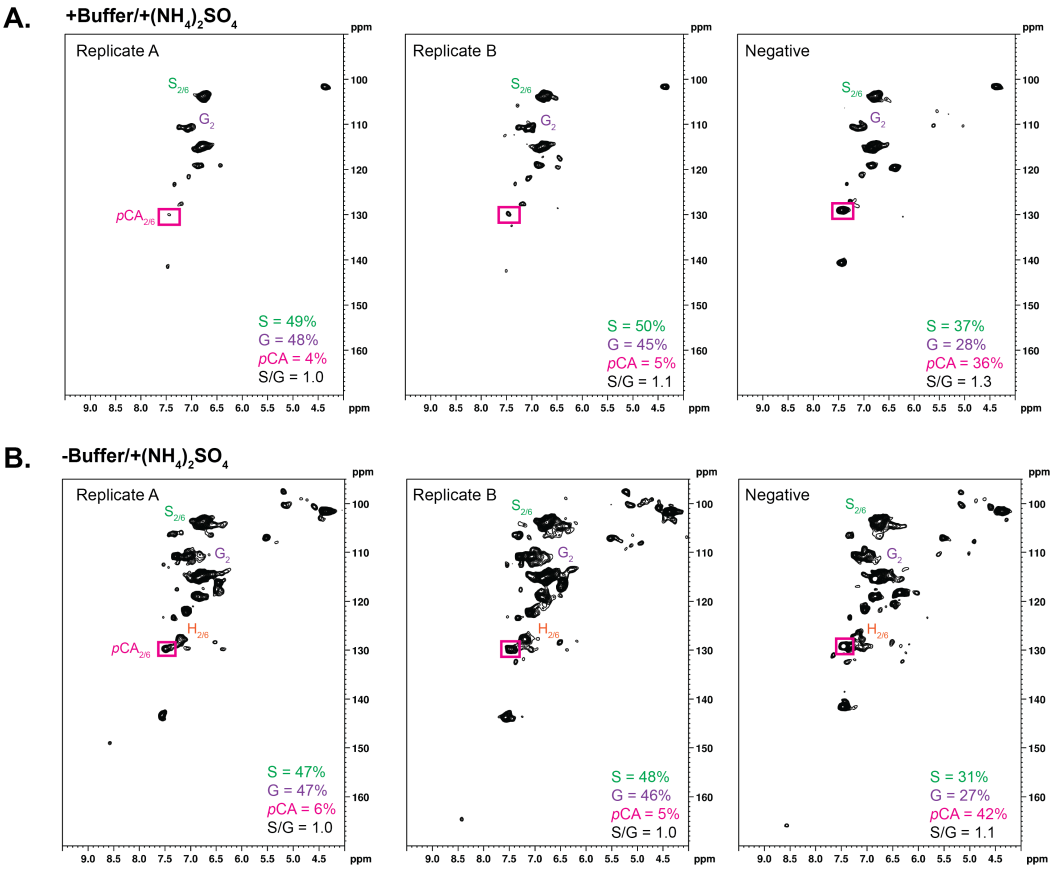


Figure 4.3. Aromatic and aldehyde regions of ^{1}H-^{13}C HSQC 2D-NMR of lignin cultures containing ammonium sulfate. The pink box indicates *p*-coumaric acid. In panel A and B, Replicate A and B indicate biological replicates and the negative indicates the no-cell control flask treated in the same condition as biological replicates, but not inoculated with cells.

Table 4.9. Compositional analysis of the culture supernatant after the culturing period.

Sample (Buff/A.S.)	**Structural components (% of total profile)**							
	Ash	Protein	Lignin	Glucan	Xylan	Galactan	Arabinan	Acetyl
+ / +, rep. a	68.95	7.96	14.07	0.75	3.30	1.02	1.53	0.10
+ / +, rep. b	68.69	7.71	14.55	0.63	3.26	1.04	1.52	0.12
+ / +, negative	65.99	6.65	14.91	0.66	3.29	0.95	1.41	1.56
+ / -, rep. a	69.77	2.71	15.47	0.68	3.49	1.06	1.60	0.14
+ / -, rep. b	69.14	2.71	15.29	0.70	3.64	1.08	1.62	0.10
+ / -, negative	66.07	2.71	15.66	0.69	3.53	1.00	1.51	1.93
- / +, rep. a	50.36	12.97	25.43	0.99	5.80	1.64	2.62	0.14
- / +, rep. b	50.18	12.77	24.00	1.14	5.71	1.52	2.62	0.16
- / +, negative	46.98	15.55	26.35	1.23	5.63	1.51	2.47	1.97
- / -, rep. a	49.50	4.39	26.23	1.14	6.16	1.76	2.87	0.36
- / -, rep. b	49.37	4.58	26.62	1.18	6.28	1.85	2.91	0.25
- / -, negative	48.76	4.46	27.23	1.21	5.97	1.70	2.68	4.62

molecular weight lignin across all culture conditions (Figure 4.5). This was confirmed by the ^{1}H-^{13}C HSQC 2D-NMR (Fig. 4.4). The aliphatic regions of the 2D-NMR analyzes the β-O-4 linkages (blue box) and aliphatic/side chain (green box) content in the polymeric and oligomeric lignin (Figure 4.4). The DMSO solvent peak is indicated by a red box across all figure panels. Methoxy groups are boxed in green and β-O-4 linkages are shown with a cyan box. There is an overall decrease in the total aliphatic/side content in the buffered replicates A and B as compared to the control. There is no discernable change in the β-O-4 linkages, indicating the lignin is not being depolymerized (Figure 4.4A). The non-buffered samples show a similar lack of depolymerization (Figure 4.4B). It should be noted that not all material solubilized and settled at the bottom of the NMR tube upon addition to the DMSO, perhaps due to salts possibly from the supernatant. This makes it difficult to definitively say whether the entire lignin profile is represented and it diminishes sample com-

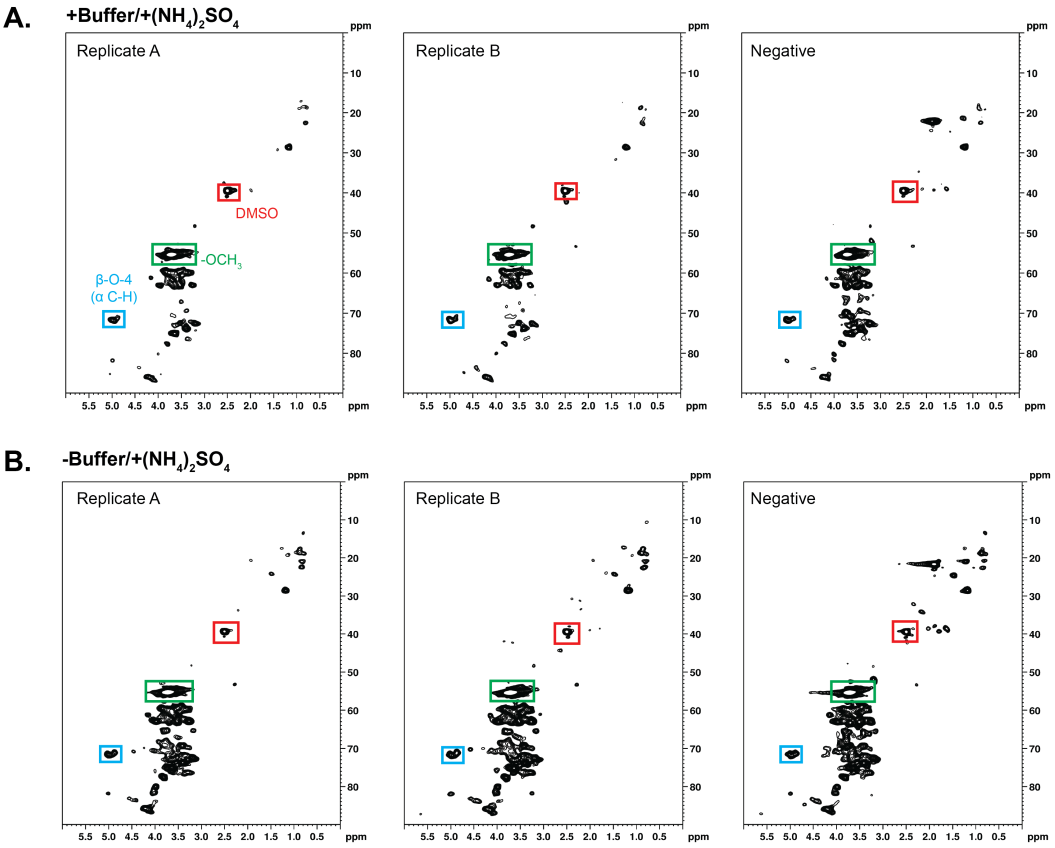


Figure 4.4. Aliphatic regions of ^{1}H-^{13}C HSQC 2D-NMR of lignin cultures containing ammonium sulfate. In panel A and B, Replicate A and B indicate biological replicates and the negative indicates the control flask treated in the same condition as biological replicates, but not inoculated with cells. The red box indicates DMSO solvent peak, the cyan box indicates β-O-4 bonds (α C-H), and the green box indicates methoxy peaks.

parability. As most ligninolytic enzymes are secreted, enzyme activity for laccases, Mn^{2+}-dependent peroxidases (MnP and DyP), and Mn^{2+}-independent peroxidases (DyP) were probed at 12 hour, 24 hour, and 72 hour timepoints using ABTS, 2,6-DMP, and veratryl alcohol as substrates. The activity for all enzyme types tested across all substrates was negligible (data not shown), which is consistent with a lack of depolymerization of the high molecular weight lignin.

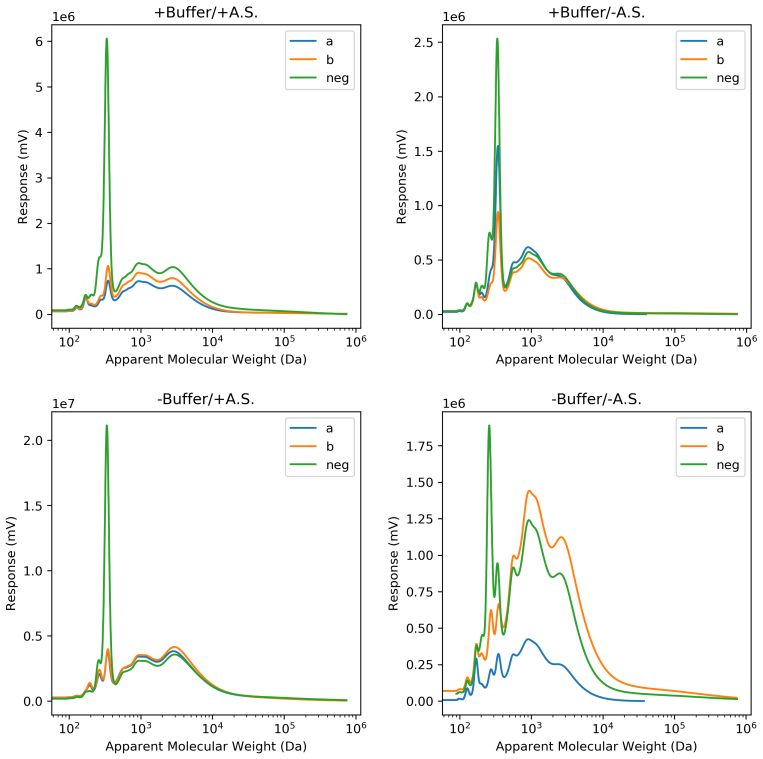


Figure 4.5. Gel permeation chromatography (GPC) of lignin supernatant after incubation with yeast cells. In all panels, the blue and orange lines indicate biological replicates ("a" and "b") and the green line represents the no cell control flask ("neg") treated in the same condition as biological replicates but not inoculated with cells.

4.4.2 Biochemical determination of pathways used by *C. oleaginosus* for aromatic catabolism

Aromatic cleavage can proceed through intradiol and extradiol pathways. It is clear *C. oleaginosus* is capable of catabolizing aromatic monomers, but it is unclear whether the aromatic rings are cleaved in an *ortho* or *meta* manner. To probe this, clarified yeast lysates were analyzed for their cleavage products for catechol, protocatechuic acid, and hydroxyquinol as well as for hydroxylation of resorcinol (Figure 4.6). The data indicate the yeast cells grown in the presence of phenol have the strongest activity to cleave catechol in a *ortho* manner. Cells grown with resorcinol and pHBA show slight activity towards catechol, and cells grown with glucose do not show activity to cleave catechol. No conditions resulted in catechol being cleaved at a *meta* position. No conditions resulted in significant activity of protocatechuic acid cleavage at either *ortho* or *meta* positions. It could be that pHBA is not consumed at a rate that can be observed within 5 minutes. Resorcinol hydroxylation was measured through depletion of NADPH at 340 nm. It is unsurprising that all yeast lysates showed depletion, as there are many enzymes in a cell that utilize NADPH. Thus, we measured if hydroxyquinol was able to be cleaved in an *ortho* position by measuring accumulation of the *ortho*-cleavage product, maleylacetate. All the cells showed accumulation of maleylacetate, which is also unsurprising, given how hydroxyquinol autoxidizes at a very rapid rate. It would be better to use an assay that can measure oxygen content instead. Overall, based on the *ortho*-cleavage of catechol, we assumed the cell utilizes *ortho*-cleavage for all if its cleavage mechanisms. Biochemical assays can suggest potential pathways, but the link between phenotype and genotype is still missing. The next sections address the missing genotypic information for aromatic catabolism by utilizing a multi-omic dataset.

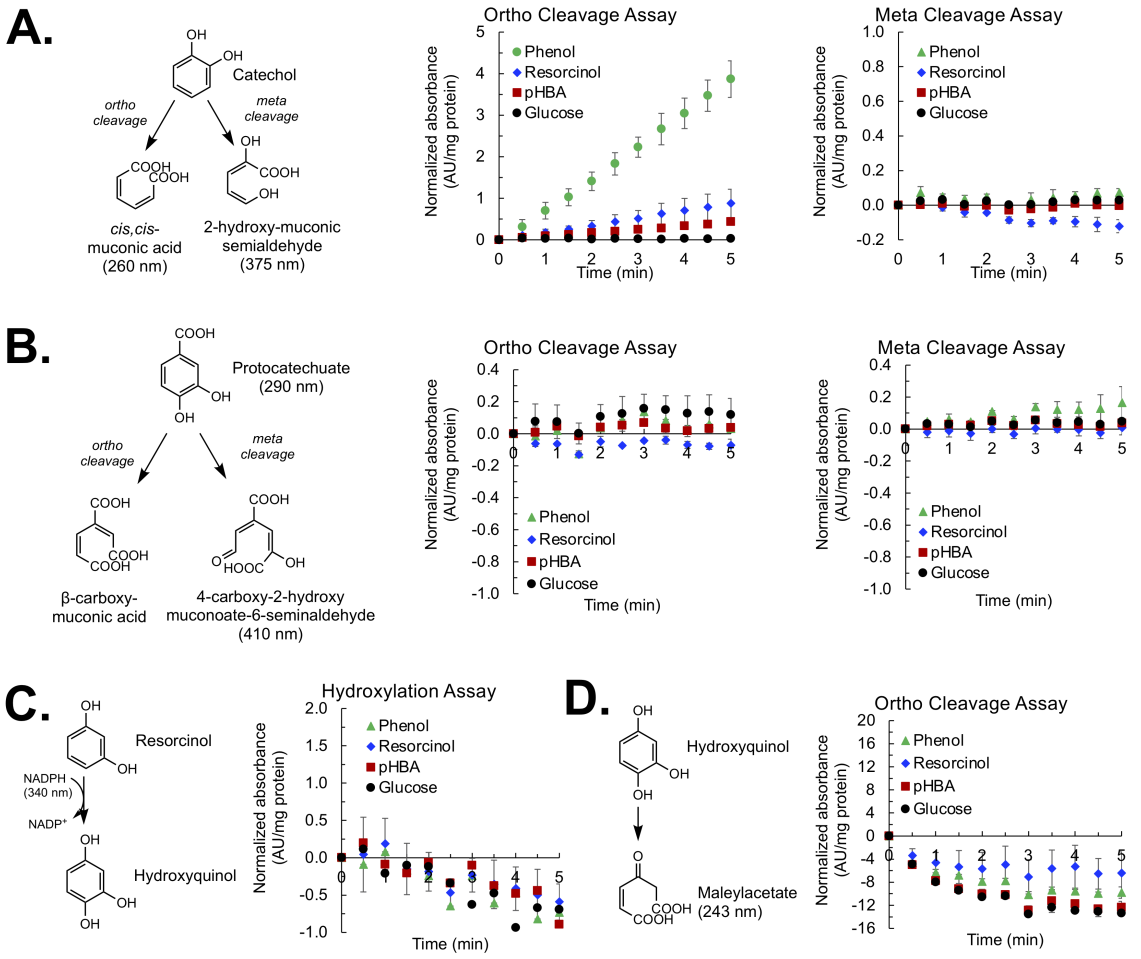


Figure 4.6. Biochemical assay used to determine possible catabolic routes for A. catechol cleavage, B. protocatechuic acid cleavage, C. resorcinol hydroxylation, and D. hydroxyquinol cleavage using clarified lysates from *C. oleaginosus* cells grown with phenol, resorcinol, pHBA, and glucose.

Table 4.10. List enzymes resulting from an initial BLAST search looking for homologs of aromatic funneling pathways.

Enzyme	**Homolog**	**Annotation**
Monooxygenase	CC85DRAFT_240134	putative phenol 2-monooxygenase
	CC85DRAFT_287490	putative phenol 2-monooxygenase
	CC85DRAFT_282431	alcohol dehydrogenase
Dioxygenase	CC85DRAFT_311359	hydroxyquinol 1,2-dioxygenase
	CC85DRAFT_150900	putative catechol 1,2-dioxygenase
Muconate lactonizing enzyme	CC85DRAFT_271894	adenylosuccinate lyase
	CC85DRAFT_274445	muconoate lactonizing enzyme
β-ketoadipate-CoA transferase	CC85DRAFT_260706	succinyl-CoA:α-ketoacid-CoA transferase
	CC85DRAFT_283947	succinyl-CoA:α-ketoacid-CoA transferase

4.4.3 Improving genome annotation by 20%

While bacterial aromatic catabolic pathways are well described genetically, yeast are typically only described biochemically. As noted earlier, incomplete catabolism of ferulic acid, accumulation of vanillic acid, and inability to depolymerize lignin motivates metabolic engineering to enhance the wild-type strain for these properties. However, there is no description for what enzymes or mechanisms are being utilized by *C. oleaginosus*, as is the case for many yeast species. Data mining for aromatic funneling enzymes in the reference genome for *C. oleaginosus* published by JGI yielded incomplete results due to the lacking yeast annotation (Table 4.10), so we utilized RNAseq data to enhance the existing genome annotation. RNA sequences generated from cells grown in phenol, pHBA, resorcinol, and glucose were used to identify new CDS regions and build a *de novo* transcriptome.

The *de novo*, reference-guided genome assembly showed a 26.8% improved

Table 4.11. BUSCO analysis of the *de novo* genome assembly compared to the reference genome and model yeasts.

Genome	**Complete**	**Single**	**Duplicate**	**Fragmented**	**Missing**
S. cerevisiae	92.7%	90.4%	2.3%	0.0%	7.3%
Y. lipolytica	98.7%	97.4%	1.3%	1.0%	0.3%
C. oleaginosus (ref)	69.6%	69.0%	0.7%	8.6%	21.8%
C. oleaginosus (*de novo*)	96.4%	87.5%	8.9%	1.3%	2.3%

Table 4.12. Annotation statistics for the *de novo* genome assembly.

Assembly statistic	**Value**
Assembly size (bp)	17,290,011
No. of protein coding genes	8,869
Protein length (median no. of amino acids)	365
Gene length (median bp)	1,585
% of proteins with ≥4 functional annotations	45.54
% of genes with KEGG annotation	46.78
% of genes with BlastP hit	41.15
% of genes with GO term	40.85

completeness relative to the reference genome (Table 4.11, Figure 4.7). It is near the same level of completeness as model organisms *Saccharomyces cerevisiae* S288c and *Yarrowia lipolytica* CLIB122. There were 8,869 genes identified, of which 283 were unique to the *de novo* assembly and 8,586 were previously identified in the reference genome (Table 4.12). All of the genes in the reference genome were identified in the *de novo* assembly.

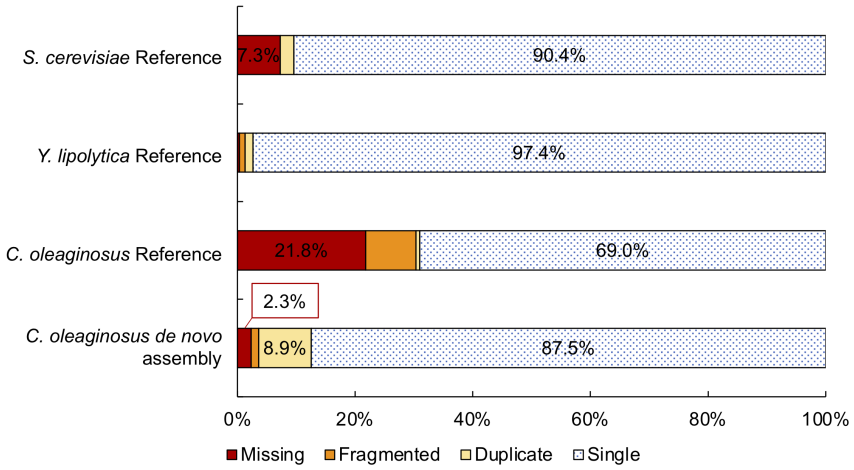


Figure 4.7. Visual representation of Table 4.11 showcasing BUSCO analysis of the *de novo* genome assembly compared to the reference genome and model yeasts.

4.4.4 Multi-omic elucidation of aromatic catabolism

4.4.4.1 Identification of candidate genes

A principal component analysis (PCA) and hierarchical clustering (HC) of the transcriptomic and proteomic datasets both demonstrate clear grouping between different substrates and replicates (Figure 4.8 and 4.9). Visualization of sample trends in both the PCA and HC plots motivated removal of *p*-coumarate 1, ferulate 1, and glucose 3 from the proteomic datasets; however, all transcriptomic datasets were retained. The HC represents clustering of all genes and proteins showcasing the top 20% variance of the total. Each of the substrates show unique clustering and each replicate of each substrate cluster together. The PCA showcases that the group of the samples is sufficient and similar to how other substrates cluster. Within the protein HC, the ferulate and lignin samples tend to cluster differently from the rest of the samples. These samples overall had a lower library abundance than the other substrates and may influence clustering. The pHBA and *p*-coumarate samples cluster together and both are clustered closer to resorcinol than phenol or glucose. pHBA

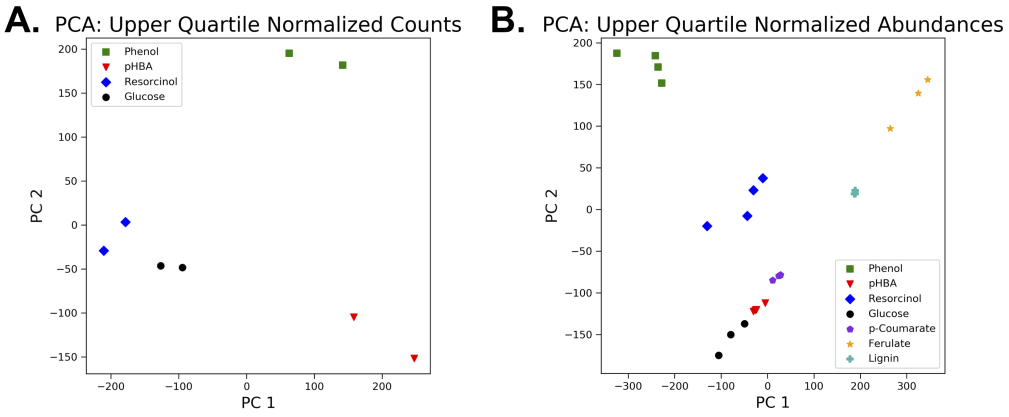


Figure 4.8. Principle component analysis of A. RNAseq samples and B. proteomic samples.

is proposed to be an intermediate of the *p*-coumarate pathway and may explain why these two cluster closely as compared to other substrates [19, 21, 42].

A comparison of the transcriptomic and proteomic log fold changes are displayed as x-y scatter plots in Figure 4.10. Statistically insignificant datapoints are marked with gray, and significant data are colored green in the phenol plot, red in the pHBA plot, and blue in the resorcinol plot. One could argue there is no trend between the RNA and protein differential expression; however, mRNA quantities have been shown to not be linearly correlated to protein quantities, so this is not entirely surprising. It may be interesting to extract the genes that expressed differently than the proteins did and evaluate their functions. This was not done in this work, but is a potential option for future exploration. As with most RNAseq datasets, the log fold change values were spot-checked with qPCR (Figure 4.11). β-actin served as the housekeeping gene to normalize each dataset, and up- and down-regulation is calculated relative to glucose. The pHBA dataset was nearly on top of the x=y line, indicating the RNAseq and qPCR values are nearly identical. The R^2 value is nearly 1. The phenol and resorcinol datasets are less correlated than the pHBA dataset, but still reasonably fitted about the x=y line.

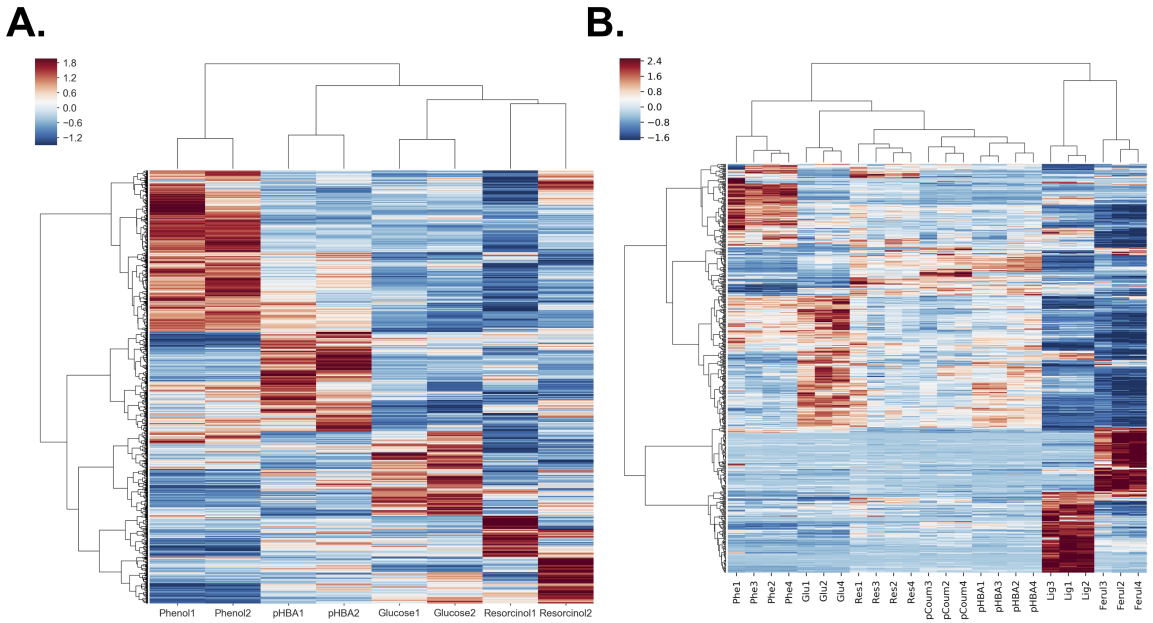


Figure 4.9. Hierarchical clustering of normalized A. transcriptomic counts and B. proteomic abundances after selecting for the genes and proteins showcasing the top 20% variance of the total.

The volcano plots in Figure 4.12 and 4.13 visualize the total proteins and transcripts, respectively, identified versus those that were determined to be statistically significant (log2FC $\geq |1|$, $p_{adj} \leq 0.05$). The venn diagram shown in Figure 4.14A shows the overlap of proteins identified in all conditions. There were only nine proteins common amongst all of the substrates: CC85DRAFT_281704, 258010, 311359, 80588, 317454, 283966, 282431, and 259523 (Figure 4.15). Of these proteins, CC85DRAFT_281704, a zinc finger, was highly upregulated in all conditions except the lignin condition. CC85DRAFT_258010 is a aldolase and is upregulated in all conditions. CC85DRAFT_311359, a putative intradiol dioxygenase, was highly upregulated in all substrate conditions. CC85DRAFT_80588 is GroES-like protein and is upregulated in all conditions to different extents. GroES is classified as a chaperone protein, although the N-terminal of alcohol dehydrogenase-like proteins have a GroES-like fold. In either function, it could have relevance to either aromatic tolerance or aro-

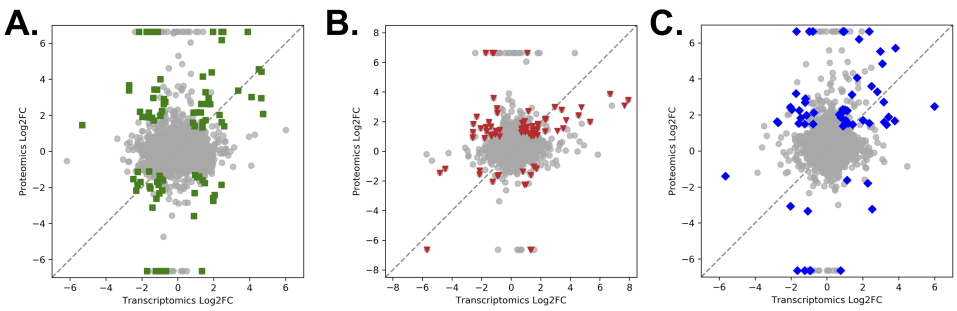


Figure 4.10. Comparison of log-fold change values resulting from transcriptomic and proteomic analysis for the A) phenol, B) pHBA, and C) resorcinol datasets. Gray circles in each plot represent statistically insignificant values, whereas colored markers represent statistically significant values.

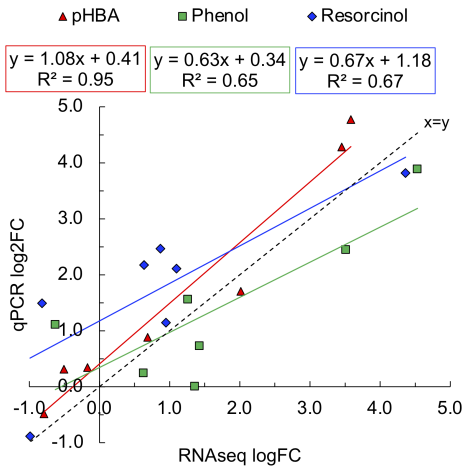


Figure 4.11. Comparison of qPCR and RNAseq log fold change values for pHBA (▲), phenol (■), and resorcinol (◆).

matic catabolism. CC85DRAFT_317454, a general substrate transporter, was upregulated significantly in all conditions except for lignin, which suggests it is a transporter relevant to aromatic monomer transport. CC85DRAFT_283966, a dehydrogenase/reductase, was upregulated in all conditions except in phenol. CC85DRAFT_282431 was originally annotated as an iron-containing alcohol dehydrogenase, but our annotation states it has maleylacetate reductase activity. This enzyme is active highest in phenol (logFC 4.3), ferulate (logFC 3.39), resorcinol (logFC 2.7) and lignin (logFC 3.18). It is minorly upregulated in *p*-coumarate (logFC 1.46) and pHBA (logFC 1.07). Lastly, CC85DRAFT_259523 is an aldehyde dehydrogenase and is upregulated across all conditions.

To elucidate enzymes that are more specific to certain substrates, venn diagrams of three substrate sets were constructed. As these subsets of three generally had a larger overlap than all 6 substrates, the proteins common between all three substrates were extracted and their logFC values plotted pairwise in scatter plots. The first set describes APL-relevant compounds. The APL contains *p*-coumaric acid and ferulic acid as the main monomer compounds released, so a venn diagram comparing these three substrates specifically is shown in Figure 4.14B. There are 21 proteins common between ferulate, *p*-coumarate, and APL. A scatter plot comparing the three conditions shows there are three proteins that do not follow the same up and down regulation trends across all three substrates. Of the proteins that are unique from the nine mentioned above, CC85DRAFT_279786, a nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase, was significantly upregulated in the lignin condition (logFC = 6.64) and in the *p*-coumarate condition (logFC = 3.64), but significantly downregulated in the ferulate condition (logFC = -6.64). This is a protein that transfers acyl groups to lipoproteins. We hypothesize there is little function in aromatic catabolism. Proteins of interest that are regulated similarly across all three datasets,

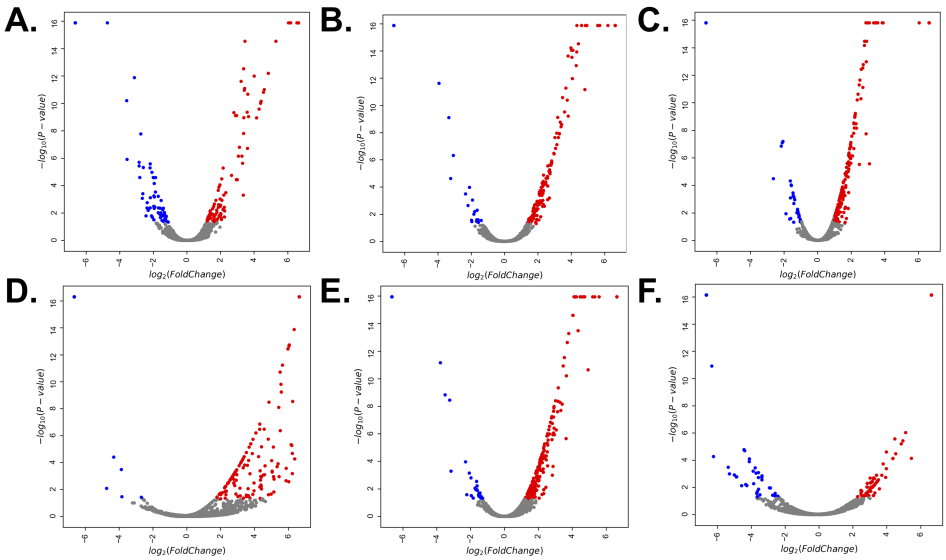


Figure 4.12. Volcano plots of the adjusted p-value versus logFC for the A. phenol, B. resorcinol, C. pHBA, D. ferulate, E. *p*-coumarate, and F. lignin proteomic datasets.

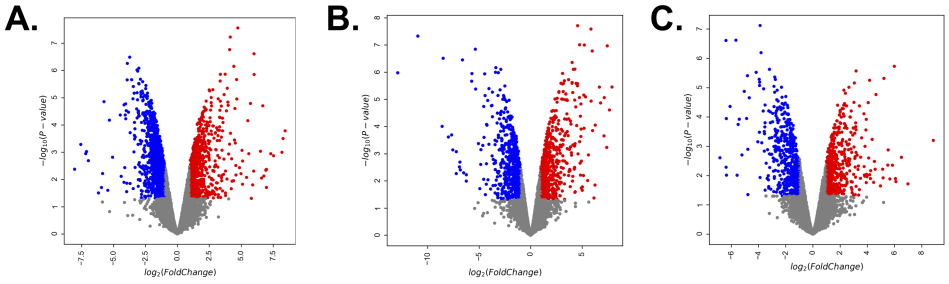


Figure 4.13. Volcano plots of the adjusted p-value versus logFC for the A. phenol, B. resorcinol, and C. pHBA transcriptomic datasets.

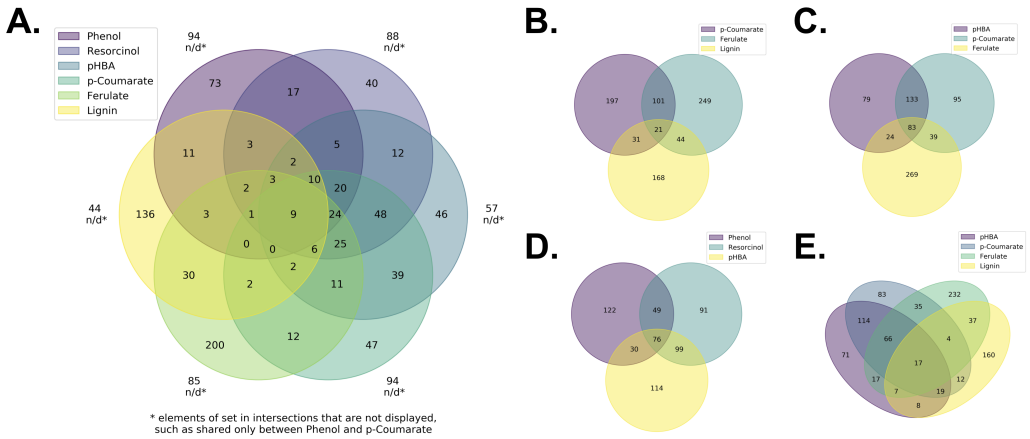


Figure 4.14. Venn diagram of A) the six logFC proteomic datasets, B) APL-relevant datasets (*p*-coumarate, ferulate, and lignin), and C) carboxylic acid-containing monomers (pHBA, p-coumarate, and ferulate), D) model monomer datasets complimentary to current RNAseq datasets (phenol, resorcinol, pHBA), and E) APL-relevant datasets including pHBA to determine protein uniqueness and similarity across the substrates.

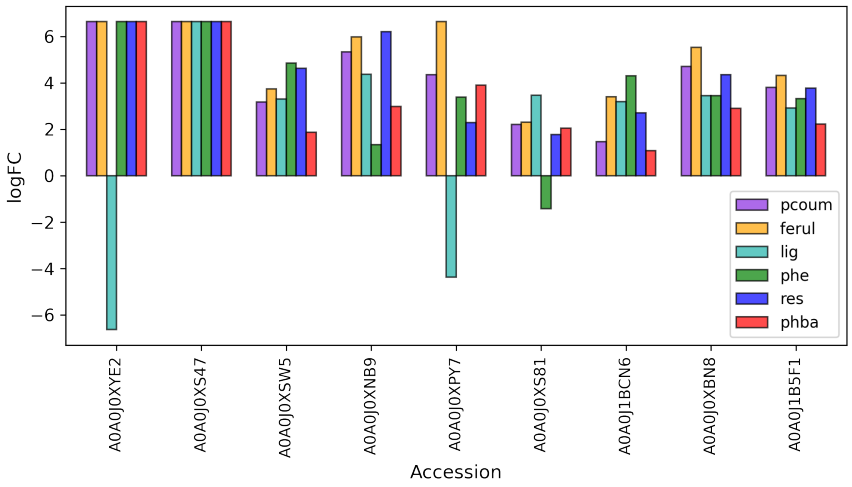


Figure 4.15. Log fold change comparison of the nine proteins identified to be statistically significantly regulated in all substrate conditions relative to glucose.

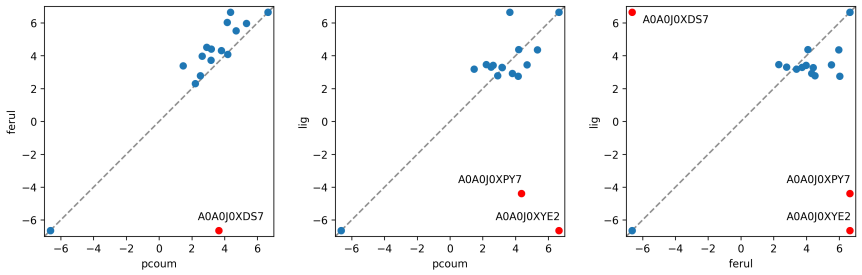


Figure 4.16. Log fold change comparison of the proteins identified to be statistically significantly regulated in across *p*-coumarate, ferulate, and lignin relative to glucose. Proteins that are oppositely regulated across the substrates are marked by red and labeled with their Accession ID.

and unique from the nine listed previously in Figure 4.14A, are CC85DRAFT_272667 (2,3-butanediol dehydrogenase) and CC85DRAFT_285545 (formate dehydrogenase). A previously uncharacterized protein, CC85DRAFT_251704, has been annotated as an oxidoreductase with monooxygenase activity in our Trinotate annotation of the RNAseq work and may have an important role in aromatic catabolism, as it is upregulated by 3.18, 4.4, and 3.28 logFC in *p*-coumarate, ferulate, and lignin, respectively.

The next venn diagram examined explores catabolic routes of *p*-coumarate and ferulate, the two aromatic rings containing conjugated carboxylic acids. Because *p*-coumarate is hypothesized to converge with pHBA, another carboxylic acid containing substrate, pHBA was included as the third comparison in this venn diagram. This set shares 83 of the same proteins (Figure 4.14C). A common step to funnel carboxylic acid-functionalized monoaromatics is to first decarboxylate them. CC85DRAFT_309111 is a putative 2,3-dihydroxybenzoic acid decarboxylase that is highly upregulated in *p*-coumarate, ferulate, and pHBA (logFC of 4.04, 4.3, and 2.79, respectively). There are a significant number of genes across these datasets relevant to hydroxylation of the aromatic rings to allow recognition for ring cleavage, many of which are novel annotations. CC85DRAFT_285638 is upregulated in three conditions. This was originally annotated as a cytochrome P450, which are

often implicated in redox and monooxygenase reactions, but our annotation defines it as a benzoate 4-monooxygeanse. CC85DRAFT_287509 is originally annotated as a pirin family protein, but our annotation suggests it has quercetin 2,3-dioxygenase activity. This protein is upregulated by 1.34, 1.9, and 2.85 logFC in pHBA, *p*-coumarate, and ferulate, respectively. CC85DRAFT_277598 was originally annotated as a putative FAD monooxygenase, PheA/TfdB family, but our annotation further specified it as a 3-hydroxybenzoate 4-monooxygenase. It is expressed the highest in ferulate conditions (logFC 6.64), followed by *p*-coumarate (logFC 5.92), and finally pHBA (logFC 2.65). CC85DRAFT_287490 was predicted to be involved in aromatic catabolism based off our initial BLASTp search. It has high homology to a phenol 2-monooxygenase from *Cutaneotrichosporon cutaneum*. This enzyme is not significantly upregulated in pHBA (logFC 0.98), but is in *p*-coumarate (logFC 1.97) and ferulate (logFC 3.21). A previously unannotated protein, CC85DRAFT_251704, is now annotated as a heme binding monooxygenase/oxidoreductase, and upregulated in pHBA (logFC 1.67), *p*-coumarate (logFC 3.18), and ferulate (logFC 4.4). Many enzymes active in the funneling processes require NADP as a cofactor. There are some proteins in this dataset that were originally annotated as NADP-binding proteins, but CC85DRAFT_78849 and CC85DRAFT_283572 were re-annotated as also having oxidoreductase activity. The former enzyme is upregulated at logFC of 3.58, 3.66, and 3.87 for pHBA, *p*-coumarate, and ferulate, respectively, whereas the latter enzyme is upreglated at 2.84, 2.37, and -6.64, respectively. CC85DRAFT_301122 is an alcohol dehydrogenase with upregulation at logFC of 3.5, 5.21, and 6.05 for pHBA, *p*-coumarate, and ferulate, respectively. CC85DRAFT_301024 was originally annotated as a GroES-like protein, which has functions regarding oxidoreduction, but our annotation specifies it may have alcohol dehydrogenase activity. CC85DRAFT_272155 was specified as a *p*-coumarate-CoA ligase and has exceptionally high upregulation

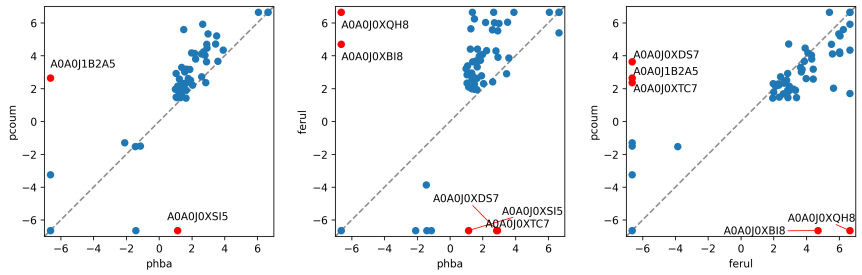


Figure 4.17. Log fold change comparison of the proteins identified to be statistically significantly regulated in across *p*-coumarate, ferulate, and lignin relative to glucose. Proteins that are oppositely regulated across the substrates are marked by red and labeled with their Accession ID.

across all three conditions (logFC 6.64).

Collectively, it seems that CC85DRAFT_287509, 277598, 251704, 78849, 283572, 301024, and 272155 all have additional or novel annotations that suggest they play a significant role in catabolism in pHBA, *p*-coumarate, and ferulate. Monooxygenase and oxidoreductases strongly implicated to be active on ferulate include CC85-DRAFT_277598, 287490, 78849, and 301122. This differs from pHBA which seems to require CC85DRAFT_251704, 78849, and 283572. Figure 4.17 shows that *p*-coumarate shows similar protein expression profiles as pHBA. This supports the hypothesis that *p*-coumarate funnels through pHBA during catabolism. *p*-Coumarate shares similar expression profiles to ferulate, though less strongly correlated with pHBA, as might be expected since both have conjugated carboxylic acids decorating their aromatic rings. Lastly, pHBA and ferulate have the least degree of correlation.

The existing transcriptomic dataset encompasses differential gene expression derived from model monoaromatic compounds phenol, resorcinol, and pHBA as compared to glucose. The third and last venn diagram model created describes the proteomic comparison of these substrates. This dataset shows 76 proteins common across all three compounds (Figure 4.14C), and overall, these monomers show less correlation to one another than the previous comparisons (Figure 4.18). CC85DRAFT_287490,

a phenol hydroxylase, is strongly upregulated in phenol (logFC 6.64) and resorcinol (logFC 5.71). It is much higher expression in phenol and resorcinol than pHBA, *p*-coumarate, or ferulate, and was not expressed significantly in APL at all. This suggests substrate specificity for phenol and resorcinol. An oxidoreductase, CC85DRAFT_267523, is upregulated by 2.21 logFC in pHBA and 2.13 logFC in resorcinol is downregulated by -1.39 in phenol. CC85DRAFT_309111 is a putative 2,3-dihydroxybenzoic acid decarboxylase, but our annotation notes this as having carboxy-lyase activity. It is more upregulated in phenol (logFC 4.56) and resorcinol (logFC 4.85) than in pHBA (logFC 2.79). In the analysis above, this enzyme is strongly upregulated in *p*-coumarate (logFC 4.04) and ferulate (logFC 4.3), as well. The carboxy lyase reaction converts a muconolactone to β-ketoadipate enol-lactone and is present in the predicted ortho cleavage pathways for all of the substrates. While it could be that it is a benzoic acid decarboxylase, the upregulation in phenol and resorcinol indicate it may be a carboxy lyase responsible for muconolactone conversion. However, since *p*-coumaric acid, ferulic acid, and pHBA all converge at protocatechuic acid, if it were a muconolactone decarboxylase, it would be upregulated more significantly in pHBA. Though the enzyme clearly has a role in catabolism of all five monoaromatic compounds, its precise role remains unclear. CC85DRAFT_287593, originally annotated as a CMD-domain-containing protein, has been reannotated as having 4-carboxymuconolactone decarboxylase activity. This enzyme is also a carboxy lyase that has the same function to convert muconolactones to β-ketoadipate enol-lactone. It has logFC values of 2.09, 3.34, and 3.68 in phenol, pHBA, resorcinol, respectively, and when searching the full proteomic dataset manually, is also upregulated in *p*-coumaric acid at a logFC value of 3.41. Once β-ketoadipate is formed, it first converted to β-ketoadipyl-CoA then to succinyl-CoA and acetyl-CoA. This second conversion requires a succinyl-CoA:3-ketoadcid-CoA transferase. In the

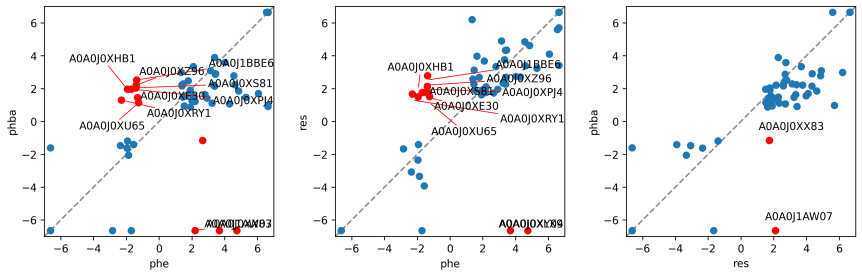


Figure 4.18. Log fold change comparison of the proteins identified to be statistically significantly regulated in across phenol, pHBA, and resorcinol relative to glucose. Proteins that are oppositely regulated across the substrates are marked by red and labeled with their Accession ID.

this dataset, CC85DRAFT_283947 was uniquely identified to be upregulated across phenol (logFC 4.61), pHBA (logFC 2.25), and resorcinol (logFC 3.39).

The inherent breadth of information a multi-omics study captures means that tolerance mechanisms may also be explored. Separately from catabolism, a pH-response regulator protein, CC85DRAFT_287214, was highly upregulated across phenol, pHBA, and resorcinol (all logFC of 6.64). A Family A GPR-like protein with ion channel activity was upregulated in phenol by 1.46 logFC, pHBA by 2.29 logFC, and resorcinol by 3.13 logFC. CC85DRAFT_327926 is a superoxide dismutase that may have a role in controlling concentration of oxidizing equivalents, which are necessary to degrade highly reduced aromatic rings [43, 44]. It is lowly upregulated in pHBA and *p*-coumarate (logFC of 1.34 and 1.9, respectively) and to a lesser extent than in ferulate conditions (logFC 2.85). There are a number of general transporters identified and upregulated; however, to explore them is outside the scope of this work and will not be discussed in more detail. Future work exploring these mechanisms will certainly identify interesting conclusions.

The identification of genes and proteins relevant to aromatic funneling was a combination of data mining and manual curation. Manual curation significantly lowering identity cutoff values and cross-checking conserved functional domains created

a list of putative genes. These genes were then identified for differential gene and protein expression in the corresponding dataset. It was combined with the datasets described above to give a holistic view of aromatic catabolism across all the substrates tested. Each set of genes is broken into funneling, ring-cleavage, and β-ketoadipate steps and key genes are highlighted in Figure 4.19. As with all 'omics work, the findings described are predictive in nature. The genetic tools to pursue knockout studies and functional genomics within *C. oleaginosus* are unavailable. Thus, other methods were pursued to enhance our knowledge from predictions to concrete functional information.

4.4.5 Functional genomics

The multi-omic work provides significant knowledge with regard to potential metabolic routes *C. oleaginosus* may use to convert aromatic compounds to central metabolites. However, predictions are sometimes incorrect, and the holy grail for catabolism is to connect the biochemical assay results discussed in section 4.4.2 to the genetic predictions identified in section 4.4.4. This section describes two methods used for functional genomics. One method uses *in vitro* translation of proteins of interest and a second method produces proteins of interest using *E. coli* as an expression host.

The NEB PURExpress kit supplies all reagents necessary to translate a protein of interest by providing a DNA template in the form of a PCR product or a plasmid. Complementary DNA (cDNA) was synthesized from RNA extracted from cells grown in phenol, resorcinol, pHBA, and glucose. The cDNA was amplified such that it contained the T7 promoter, RBS, and terminator sequence required for *in vitro* translation. Of the 10 putative phenol hydroxylase and catechol dioxygenase genes chosen for amplification, 9 amplified via PCR, and three were chosen

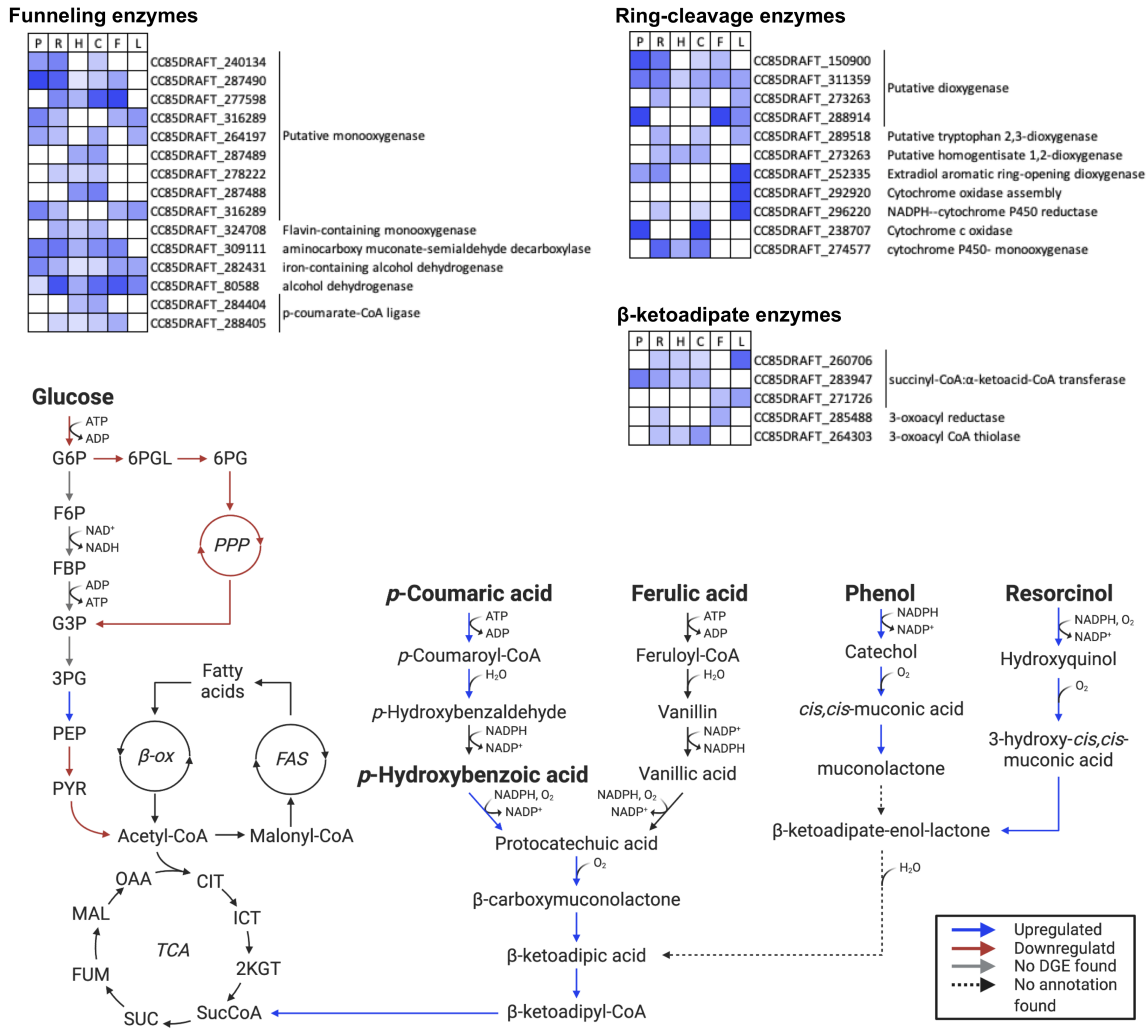


Figure 4.19. Putative genes involved in aromatic catabolism and proposed catabolic pathways for aromatic conversion.

for *in vitro* translation. Two were phenol hydroxylases (CC85DRAFT_287490 and CC85DRAFT_316289) and one was a catechol dioxygenase (CC85DRAFT_311359). CC85DRAFT_287490 was chosen because it was upregulated only in the phenol condition, whereas CC85DRAFT_316289 was upregulated in phenol, pHBA, and resorcinol. CC85DRAFT_311359 was upregulated in only pHBA and was hypothesized to be a protocatechuate 3,4-dioxygenase, rather than a catechol 1,2-dioxygenase. Of the three proteins that were translated using the PURExpress kit, two translated successfully (Figure 4.20). cDNA template concentration influences translation rate, so the template concentration was varied between 250 ng, 500 ng, and 750 ng (Figure 4.20B). It did appear that a faint band for CC85DRAFT_287490 appeared when 750 ng of cDNA template was provided for the translation reaction. CC85DRAFT_316289 protein translation was not correlated with cDNA template concentration. CC85DRAFT_311359 protein translation was positively correlated with cDNA template concentration.

Proteins resulting from the PURExpress kit in solution with the translational machinery and cofactors. Translation was successful in both presence and absence of cofactors (NADH and NADPH for the phenol hydroxylases and $FeSO_4$ and $MnSO_4$ for the catechol dioxygenase). The proteins of interest were also successfully purified by removing the translational machinery with a Ni-NTA column and buffer exchanging them into the buffer used for enzymatic assays. In all cases, enzyme assays of all proteins showed no activity (Figure 4.21). In these assays, phenol hydroxylase activity towards phenol, pHBA, and resorcinol was evaluated by depletion of NADH or NADPH, measured by absorbance at 340 nm (Figure 4.21A). Catechol dioxygenase activity towards catechol and protocatechuic acid as substrates using $FeSO_4$ or $MnSO_4$ was evaluated by measuring accumulation of the *ortho* product at 260 nm. Different protein concentrations were also tested in the enzyme assays, with no change in

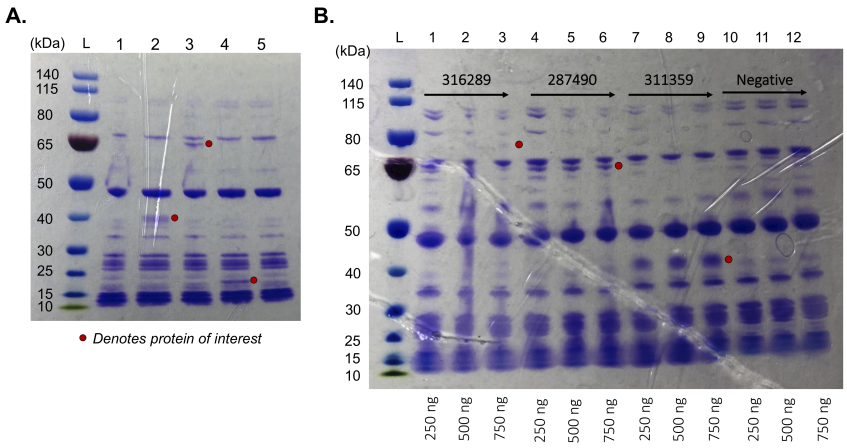


Figure 4.20. SDS-PAGE of *in vitro* translated proteins: A) Lanes are as follows: L, ladder; 1, 287490 (75.3 kDa); 2, 311359 (37.8 kDa); 3, 316289 (64.9 kDa); 4, DHFR (positive control) (24.0 kDa); 5, water (negative control). B) Testing template concentration. Lanes are as follows: L, ladder; 1-3, 287490 (75.3 kDa); 4-6, 316289 (64.9 kDa); 7-9, 311359 (37.8 kDa); 10-12, negative control; with each range using 250, 500, and 750 ng DNA as template, respectively. The red dot on each gel marks the protein of interest.

results. In these cases, CC85DRAFT_287490 and CC85DRAFT_316289 were loaded at 5, 10, 15, and 20 μL of protein and evaluated for conversion of phenol to catechol by measuring NADH and NADPH depletion at 340 nm. CC85DRAFT_311359 was similarly evaluated, but for catechol intradiol cleavage at 260 nm (Figure 4.22. It does not appear that protein concentration is the limiting factor for enzyme assay.

In vitro translation does mot add post-translational modifications (PTMs) nor does it include chaperones that may be required for successful folding. While the primary structure looks sound on an SDS-PAGE, the tertiary structure or PTMs may not be sufficient to exhibit proper enzyme activity. Protein production using *E. coli* BL21(DE3) and *E. coli* T7SHuffle Express as expression hosts could solve these problems. *E. coli* BL21(DE3) is an *E. coli* strain designed for protein production through the knockdown of protease activity and inclusion of an inducible T7 RNA polymerase expression system. T7SHuffle is derived from this cell line and includes a

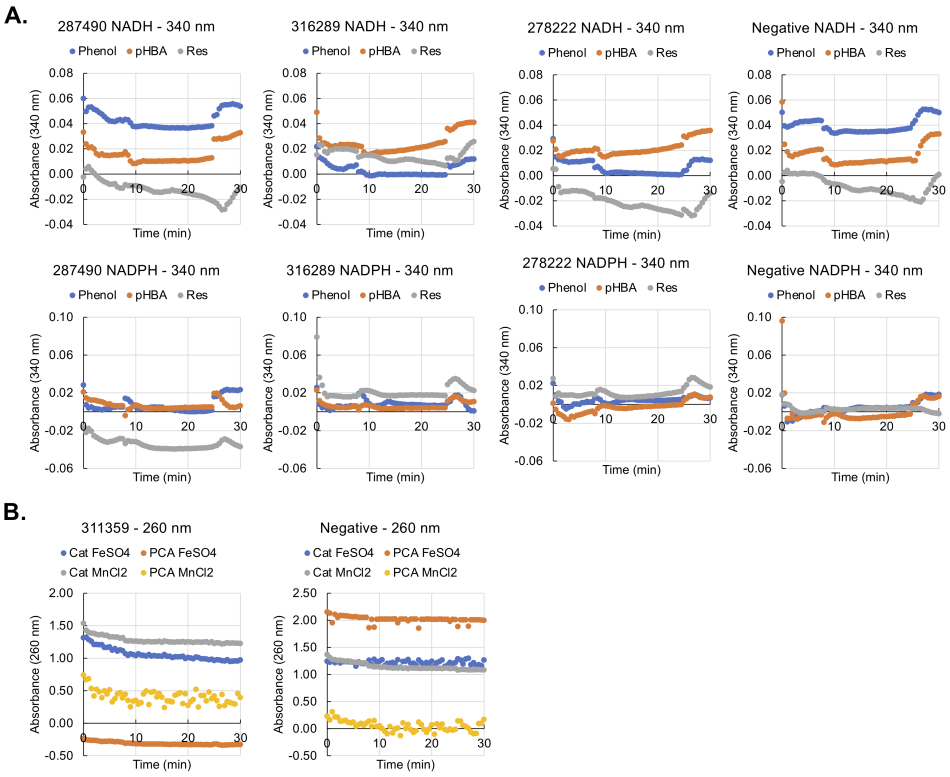


Figure 4.21. Enzymatic assays for the proteins translated by the PURExpress kit alongside cofactors. A) Comparison of 3 phenol hydroxylase activity using phenol, pHBA, or resorcinol as substrates with either NADH and NADPH as cofactors, as compared to a negative control containing cofactors and substrates but no protein. B) Comparison of catechol dioxygenase activity using either catechol (CAT) or protocatechuic acid (PCA) as substrates with either $FeSO_4$ or $MnSO_4$ as cofactors, as compared to a negative control containing cofactors and substrates but no protein.

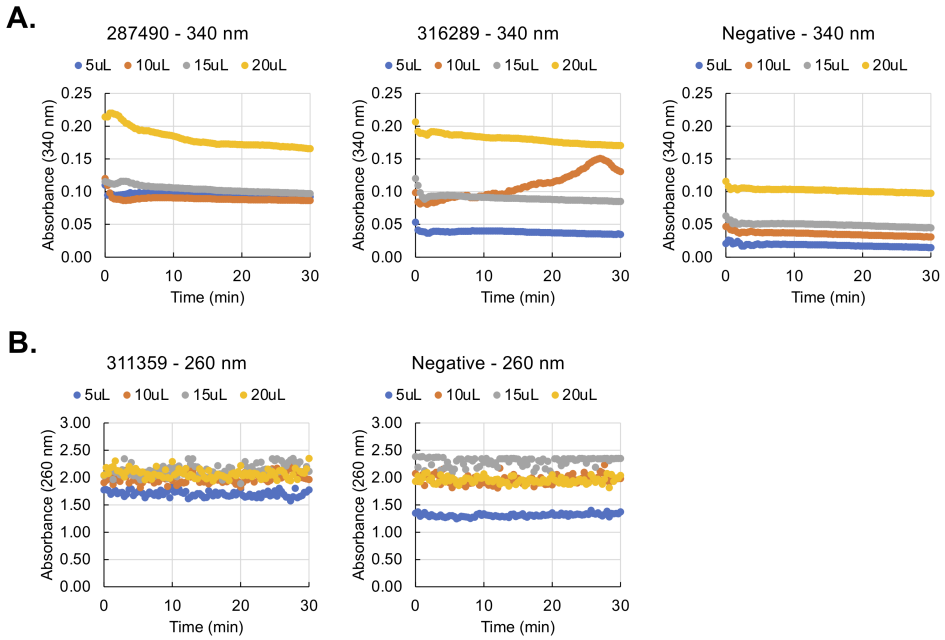


Figure 4.22. Enzymatic assays evaluating effect of protein loading for A) conversion of phenol to catechol by two different phenol hydroxylase enzymes, and B) cleavage of catechol by one catechol dioxygenase.

constitutively expressed disulfide bond isomerase, DsbC, which has been shown to act as a protein chaperone. Both strains were transformed with CC85DRAFT_311359 in a pET21b plasmid backbone. It was shown that the T7SHuffle cells did not express the protein of interest (37.8 kDa), but the BL21(DE3) cells did express them when induced in the presence of IPTG or lactose (Figure 4.23A). Thus, BL21(DE3) cells were used to express the catechol dioxygenase in the presence and absence of cofactors $FeSO_4$, $FeCl_3$, and $MnSO_4$ (Figure 4.23B). Expression studies used mCherry as a positive control for expression conditions and negative control for enzyme assays. As see in Figure 4.24, cells cultured in autoinduction media with different cofactors turned a different color than cells cultured in autoinduction media not containing lactose (uninduced).

The proteins expressed from BL21(DE3) cells in different cofactor conditions

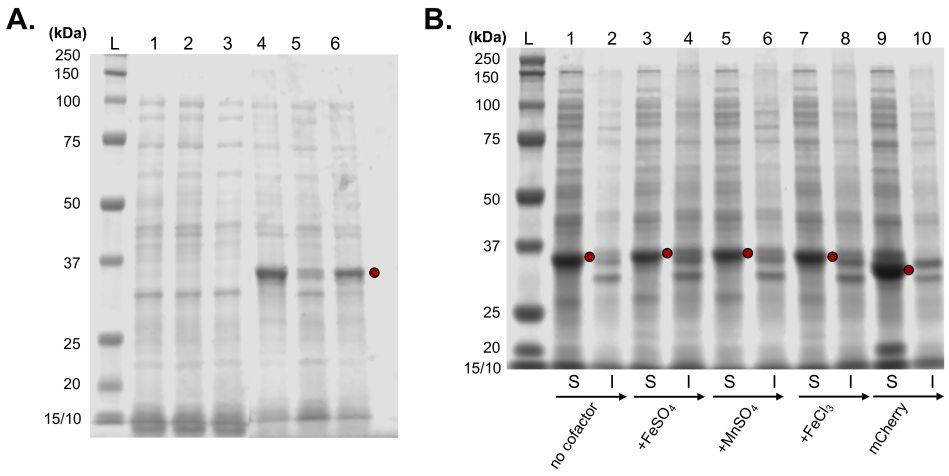


Figure 4.23. SDS-PAGE of a catechol dioxygenase (37.8 kDa) expressed from A) *E. coli* T7SHuffle Express cells or BL21DE3 cells in autoinduction (AI) media such that lanes 1 and 4 are induced with lactose, 2 and 5 are uninduced (-lactose), and 3 and 6 are induced with IPTG. B) Compares BL21(DE3) expression in the presence of no cofactors supplemented to the AI media (lanes 1 and 2), in the presence of $FeSO_4$ (lanes 3 and 4), in the presence of $MnSO_4$ (lanes 5 and 6), and in the presence of $FeCl_3$ (lanes 7 and 8) as compared to a control strain expressing mCherry (lanes 9 and 10). Even lanes are insoluble protein fractions and odd lanes are soluble protein fractions.

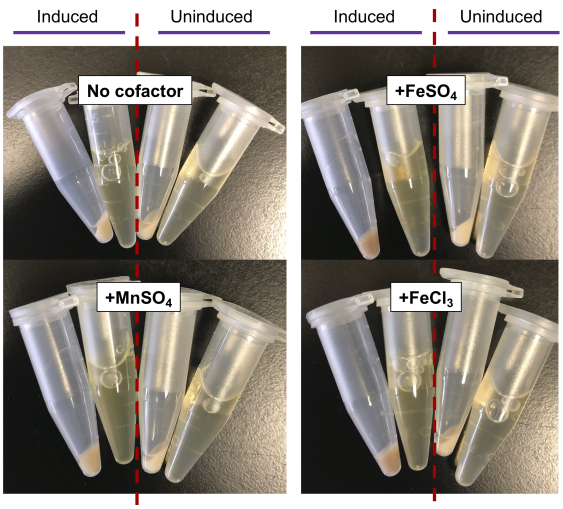


Figure 4.24. Cell pellets generated by expression of CC85DRAFT_311359 in autoinduction media containing no cofactors (upper left) or with $MnSO_4$ (bottom left), $FeSO_4$ (upper right), or $FeCl_3$ (lower left).

were used in enzymatic assays to probe catechol and protocatechuic acid cleavage (Figure 4.25). It appears that the dioxygenase expressed with $FeSO_4$ showed the best activity for catechol intradiol cleavage in the presence of $MnSO_4$ or no cofactor (Figure 4.25A). The same enzyme and expression condition showed enzyme activity for catechol extradiol cleavage in the presence of $MnSO_4$ and $FeSO_4$. This may indicate this enzyme has both intra and extradiol activity. By contrast, the dioxygenase did not seem to have activity towards protocatechuic acid, regardless of expression or assay conditions (Figure 4.25C and D). Together, the data indicate the CC85DRAFT_311359 enzyme is a catechol dioxygenase with both intradiol and extradiol cleavage activity.

4.5 Discussion and Conclusions

Lignin hydrolysates are complex mixtures of aromatic compounds that often change in composition depending on the lignin pretreatment method and source of the lignin stream. Biological valorization requires microorganisms to tolerate and metabolize a wide variety of aromatics and byproducts found in hydrolysates. Understanding the mechanisms utilized is key to facilitate robust lignin conversion to value-added products, such as oleochemicals. Oleaginous yeasts are particularly well suited for this; however, few are able to metabolize aromatic monomers and remain oleaginous [18, 37]. *Cutaneotrichosporon oleaginosus* has shown promising ability to accumulate high quantities of lipids when grown with monoaromatics [17, 18]. In this work, we evaluate the growth and lipid production of cells grown in a lignin hydrolysate. In doing so, we were able to determine this yeast tolerates a lignin hydrolysate, remains oleaginous in nitrogen-limited conditions, and consumes released aromatic monomers. Yeast genome annotation for aromatic metabolism lags signifi-

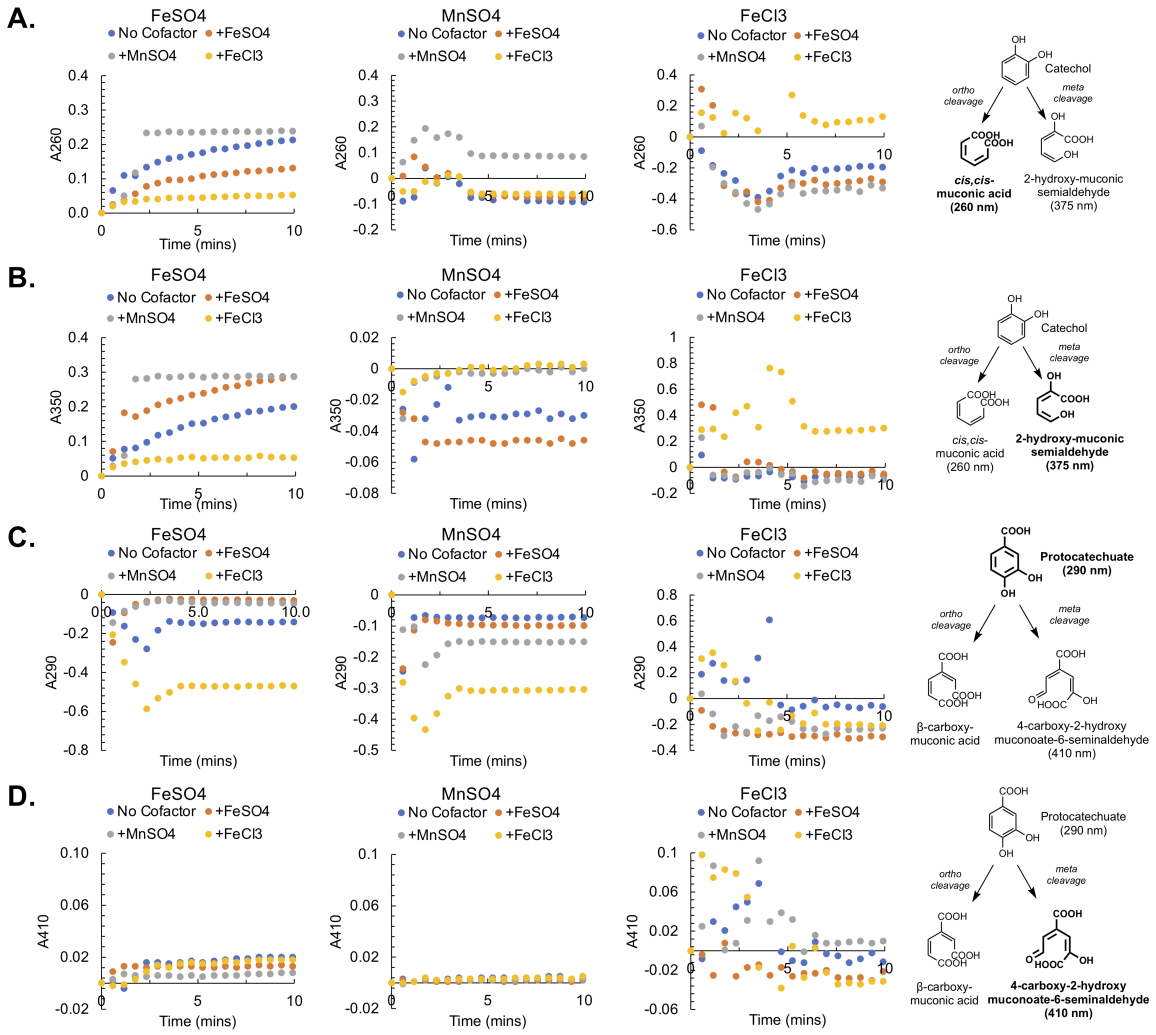


Figure 4.25. Enzyme assays for CC85DRAFT_311359 expressed from *E. coli* BL21(DE3). Cultures were spiked with cofactors to evaluate the effect of their presence of $FeSO_4$ (left column), $MnSO_4$ (middle column), and $FeSO_4$ (right column) on enzyme folding. Enzyme activity for A) catechol *ortho*-cleavage, B) catechol *meta*-cleavage, C) PCA *ortho*-cleavage, and D) PCA *meta*-cleavage was evaluated over 10 mins in the presence of no cofactors (blue), $FeSO_4$ (orange), $MnSO_4$ (gray), and $FeCl_3$ (yellow).

cantly behind bacteria. To address this, five aromatic monomers – phenol, resorcinol, *p*-hydroxybenzoic acid (pHBA), ferulic acid, and *p*-coumaric acid – and an alkaline pretreated corn stover lignin (APL) hydrolysate were used as substrates for comparative transcriptomics and proteomics in order to identify differentially expressed genes. We have identified a number of novel genes and provide insights into yeast aromatic catabolism that can serve as targets for engineering *C. oleaginosus* for enhanced lignin valorization.

Cutaneotrichosporon oleaginosus has shown a promising ability to grow in a lignin hydrolysate without any supplementation. In a nitrogen limited media, it retains oleaginous behavior by accumulating lipids at greater than 30% of its dry cell weight. It showed uptake of aromatic monomers released during hydrolysis – *p*-coumaric acid and ferulic acid – but was unable to fully catabolize the latter. Other organisms have similarly been grown in lignin hydrolysates, including *Pseudomonas putida* KT2442, *Sphingobium* sp. SYK-6, and *Rhodococcus opacus* PD630 [24, 45–50]. While *C. oleaginosus* was unable to depolymerize the higher molecular weight fractions, laccases and peroxidases from lignin degrading organisms can heterologously expressed, as has been done for many other organisms, such as *Rhodosporidium toruloides* [51]. Fungal secretomes, known for their superb ligninolytic capabilities, have been used in combination with *P. putida* acting as a microbial sink to facilitate full lignin metabolism [24]. A similar approach may be used with *C. oleaginosus*, rather than moving towards consolidated bioprocessing. Regardless of the approach used to depolymerize the high molecular weight lignin, genetic engineering to enhance the range of substrates utilized and rate of consumption will facilitate efficient lignin valorization.

While the catabolic pathways are well described biochemically in yeast, there is a lacking connection between reactions observed and the genes and proteins cat-

alyzing for them [16, 19, 21]. Without this knowledge, rationally engineering aromatic catabolism becomes infeasible. In yeasts related to *C. oleaginosus*, phenol metabolism genes have been discovered and biochemically characterized. *Cutaneotrichosporon cutaneum* has both demonstrated aromatic catabolism. In *C. cutaneum*, phenol hydroxylase has been well described and annotated [52–57]. Phenol has been described to be cleaved at an *ortho*-position and converted to muconic acid [58]. We see many enzymes that have putative functions for *ortho*-cleavage of phenol and resorcinol that regulated differently than for pHBA, *p*-coumarate, and ferulate. In *C. cutaneum*, resorcinol is converted through hydroxyquinol [58]. A single dioxygenase is expressed commonly across every aromatic substrate, including alkaline pre-treated lignin. There are other dioxygenases reported as well that seem to have substrate specific regulation. There is a pHBA cleavage pathway described biochemically for *C. cutaneum* that proceeds through protocatechuic acid, but there is an alternative route described that converts pHBA to hydroxyquinol, rather than being ring-cleaved or converted to catechol [59, 60]. It seems *C. oleaginosus* may converts pHBA through protocatechuic acid, as there are dioxygenases that are highly upregulated by phenol or resorcinol, and not in pHBA, *p*-coumarate, or ferulate. Overall, a litany of monooxygenases, oxidoreductases, alcohol and aldehyde dehydrogenases, dioxygenases, and other funneling enzymes are reported in this chapter. The exact function of some of these genes is not well understood, and the functions proposed are hypothesized. Still, they are important for understanding yeast aromatic catabolism and provide targets for future engineering work.

Functional genomics is the ultimate goal for any predictive study. Several attempts to express a suite of putative monooxygenases and dioxygenases through *in vitro* and *in vivo* methods were unable to show confirmation of functional predictions with biochemical assays. A lack of genetic tools means genes cannot be knocked out

to show loss of function and with phenotype rescue by complementation. The phenol hydroxylase from *Cutaneotrichosporon cutaneum* was the first eukaryotic phenol hydroxylase to be expressed and characterized. It was successfully amplified from the cDNA and transformed into *E. coli* for functional and kinetic assays on a variety of substrates, and has since been mutagenized and characterized further [52–57]. This approach was taken with a putative catechol 1,2-dioxygenase from *C. oleaginosus*, CC85DRAFT_311359. The data suggest the putative dioxygenase accepts catechol as a substrate and not protocatechuic acid, and that cleavage occurs both intradiol and extradiol. It is possible a cell lysis protocol with a buffer compatible for enzyme activity, such as BugBuster, would be a better lysis method. This would eliminate the possibility that the protein is damaged during the freeze-thaw lysis. Other proteins, such as the putative phenol hydroxylases, should be cloned for functional confirmation. Metabolomics is another route to prove the proposed reactions take place by searching for metabolites along the funneling pathways. If ^{13}C labeling is used, flux through these pathways can also be identified and a metabolic model made. It is quite expensive, and sometimes impossible, to get labeled versions of certain aromatic compounds, so it may be that the aromatic compounds remain unlabeled and the yeast be co-fed with fully labeled glucose. It will be difficult to identify regulatory effects caused by feeding a sugar with an aromatic, but may help identify metabolites and reactions, nonetheless. Functional genomics within *C. oleaginosus* will only be achieved with the development of better genetic tools. Progress towards this is discussed in Chapter 5.

4.6 Acknowledgments

This work was performed in preparation for publication in *Nucleic Acids Research*.

This research was made possible, in part, with support from the Clemson University Genomics and Bioinformatics Facility, which receives support from an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM109094. We would like to acknowledge Dr. Gregg Beckham and his team for access to alkaline pretreated lignin hydrolysate and access to analytical equipment and personnel necessary for lignin characterization.

This material is based upon work supported by the Department of Energy SCGSR fellowship, NREL, and ORNL. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of Energy, NREL, ORNL, or Clemson University.

Bibliography

[1] Novaes, E.; Kirst, M.; Chiang, V., *et al.* Lignin and biomass: a negative correlation for wood formation and lignin content in trees. *Plant Physiol*, **2010**, 154, 555–561.

[2] Bertella, S.; Luterbacher, J. Lignin Functionalization for the Production of Novel Materials. *Trends in Chem*, **2020**, 2, 440–453.

[3] Ponnusamy, V.; Nguyen, D.; Dharmaraja, J., *et al.* A review on lignin structure, pretreatments, fermentation reactions and biorefinery potential. *Biores Tech*, **2019**, 271, 462–472.

[4] Thakur, V.; Thakur, M.; Raghavan, P., *et al.* Valorization of Biomass: Deriving More Value from Waste. *Science*, **2012**, 337, 695–699.

[5] Kumar, A.; Anushree, J.; Bhaskar, T. Utilization of lignin: A sustainable and eco-friendly approach. *Journal of the Energy Institute*, **2020**, 93, 253–271.

[6] Davis, R.; Tao, L.; Tan, E., *et al.* In *Process Design and Economics for the Conversion of Lignocellulosic Biomass to Hydrocarbons: Dilute-Acid and Enzymatic Deconstruction of Biomass to Sugars and Biological Conversion of Sugars to Hydrocarbons*. NREL, **2013**.

[7] Iskigor, F.; Becer, R. Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers.. *Poly Chem*, **2015**, 5, 447.

[8] Vu, H.; Nguyen, L.; Vu, M., *et al.* A comprehensive review on the framework to valorise lignocellulosic biomass as biorefinery feedstocks. *Sci Total Evinron*, **2020**, 743, 140630.

[9] Klein-Marcuschamer, D.; Simmons, B.; Blanch, H. Techno-economic analysis of a lignocellulosic ethanol biorefinery with ionic liquid pre-treatment. *Biofuels, Bioproducts, and Biorefining*, **2011**, 5, 562–569.

[10] Bbosa, D.; Mba-Wright, M.; Brown, R. More than ethanol: a techno-economic analysis of a corn stover-ethanol biorefinery integrated with a hydrothermal liquefaction process to convert lignin into biochemicals. *Biofuels, Bioproducts, and Biorefining*, **2018**, 12, 497–509.

[11] Rijn, R.; Nieves, I.; Shanmugam, K., *et al.* Techno-Economic Evaluation of Cellulosic Ethanol Production Based on Pilot Biorefinery Data: a Case Study of Sweet Sorghum Bagasse Processed via L+SScF. *BioEnergy Res*, **2019**, 11, 414–425.

[12] Bozell, J.; Holladay, J.; Johnson, D., *et al.* In *Top Value Added Chemicals from Biomass*. PNNL, **2007**.

[13] Valdivia, M.; Galan, J.; Laffarga, J., *et al.* Biofuels 2020: Biorefineries based on lignocellulosic materials. *Microb Biotechnol*, **2016**, 9, 585–594.

[14] Yaguchi, A.; Spagnuolo, M.; Blenner, M. Oleaginous yeast for biofuel and oleochemical production. *Curr Opin Biotechnol*, **2019**, 57, 73–81.

[15] Yaguchi, A.; Rives, D.; Blenner, M. The New Kids on the Block: Emerging Oleaginous Yeast of Biotechnological Importance. *AIMS Microbiology*, **2017**, 3(2), 227–247.

[16] Yaguchi, A.; Spagnuolo, M.; Blenner, M. Engineering yeast for utilization of alternative feedstocks. *Curr Opin Biotechnol*, **2018**, 53, 122–129.

[17] Yaguchi, A.; Robinson, A.; Mihealsick, E., *et al.* Metabolism of aromatics by Trichosporon oleaginosus while remaining oleaginous. *Microb Cell Fact*, **2017**, 16(1), 206.

[18] Yaguchi, A.; Franaszek, N.; O'Neill, K., *et al.* Identification of oleaginous yeasts that metabolize aromatic compounds. *JIMB*, **2020**, 47, 801–813.

[19] Lubbers, R.; Dilokpimol, A.; Visser, J., *et al.* A comparison between the homocyclic aromatic metabolic pathways from plant-derived compounds by bacteria and fungi. *Biotech Adv*, **2019**, 37, 107396.

[20] Becker, J.; Whittmann, C. A field of dreams: Lignin valorization into chemicals, materials, fuels, and health-care products. *Biotech Adv*, **2019**, 37, 107360.

[21] Brink, D.; Ravi, K.; Liden, G., *et al.* Mapping the diversity of microbial lignin catabolism: experiences from the eLignin database. *Appl Microbiol Biotechnol*, **2019**, 103, 3979–4002.

[22] Kourist, R.; Bracharz, F.; Lorenzen, J., *et al.* Genomics and Transcriptomics Analyses of the Oil-Accumulating Basidiomycete Yeast Trichosporon oleaginosus: Insights into Substrate Utilization and Alternative Evolutionary Trajectories of Fungal Mating Systems. *MBio*, **2015**, 6(4), e00918.

[23] Karp, E.M.; Donohoe, B.S.; O'Brien, M.H., *et al.* Alkaline Pretreatment of Corn Stover: Bench-Scale Fractionation and Stream Characterization. *ACS Sus Chem Eng*, **2014**, 2(6), 1481–1491.

[24] Salvachúa, D.; Katahira, R.; Cleveland, N.S., *et al.* Lignin depolymerization by fungal secretomes and a microbial sink. *Green Chemistry*, **2016**, 18(22), 6046–6062.

[25] Mansfield, S.; Kim, H.; Lu, F., *et al.* Whole plant cell wall characterization using 2D-NMR. *Nature Protocol*, **2012**, 7, 1579–15 moleculhe s89.

[26] Wen, J.-L.; Sun, S.-L.; Xue, B.-L., *et al.* Recent Advances in Characterization of Lignin Polymer by Solution-State Nuclear Magnetic Resonance (NMR) Methodology. *Materials*, **2013**, 6(1), 359–391.

[27] Sluiter, J.B.; Ruiz, R.O.; Scarlata, C.J., *et al.* Compositional Analysis of Lignocellulosic Feedstocks. 1. Review and Description of Methods. *Journal of Agricultural and Food Chemistry*, **2010**, 58(16), 9043–9053.

[28] Bolger, A.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data.. *Bioinformatics*, **2014**, 30, 2114–2120.

[29] Davidson, N.M.; Oshlack, A. Necklace: combining reference and assembled transcriptomes for more comprehensive RNA-Seq analysis. *GigaScience*, **2018**, 7.

[30] Dujon, B.; Sherman, D.; Fischer, G., *et al.* Genome evolution in yeasts. *Nature*, **2004**, 430, 35–44.

[31] Kim, D.; Langmead, B.; Salzberg, S. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*, **2015**, 12, 357–360.

[32] Liao, Y.; Smyth, G.; Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **2014**, 30, 923–930.

[33] McCarthy, D.; Chen, Y.; Smyth, G. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nuc Acids Res*, **2012**, 40, 4288–4297.

[34] Grabherr, M.; Haas, B.; Yassour, M., *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnol*, **2011**, 29, 644–652.

[35] Moriya, Y.; S., I.M.; Yoshizawa, A., *et al.* KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*, **2007**, 35, W182–W185.

[36] Alexa, A.; Rahnenfuhrer, J.; Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, **2006**, 22, 1600–1607.

[37] Simao, F.; Waterhouse, R.; Ioannidis, P., *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **2015**, 31, 3210–3212.

[38] Mahiudddin, M.; Fakhruddin, A.; Abdullah-Al-Mahin, Degradation of Phenol via Meta Cleavage Pathway by Pseudomonas fluorescens PU1. *Int Scholarly Res Notices*, **2012**, 2012.

[39] Iwagami, S.; Yang, K.; Davies, J. Characterization of the Protocatechuic Acid Catabolic Gene Cluster from Streptomyces sp. Strain 2065. *Appl Environ Microbiol*, **2000**, 66(4), 1499–1508.

[40] Ohta, Y.; Ribbons, D. Bacterial Metabolism of Resorcinylic Compounds: Purification and Properties of Orcinol Hydroxylase and Resorcinol Hydroxylase from Pseudomonas putida ORC. *Eur. J. Biochem*, **1976**, 61, 259–269.

[41] Hatta, T.; Nakano, O.; Imai, N., *et al.* Cloning and Sequence Analysis of Hydroxyquinol 1,2-Dioxygenase Gene in 2,4,6-Trichlorophenol-Degrading Ralstonia pickettii DTP0602 and Characterization of Its Product. *J Biosci and Bioeng*, **1999**, 87(3), 267–272.

[42] Kamimura, N.; Takahashi, K.; Mori, K., *et al.* Bacterial catabolism of lignin-derived aromatics: New findings in a recent decade: Update on bacterial lignin catabolism. *Environ Microbiol Rep*, **2017**, 9, 679–705.

[43] Henson, W.R.; Campbell, T.; DeLorenzo, D.M., *et al.* Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus. *Metab Eng*, **2018**, 49, 69–83.

[44] Yoneda, A.; Henson, W.R.; Goldner, N.K., *et al.* Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630. *Nucleic Acids Res*, **2016**, 44(5), 2240–2254.

[45] Zhao, C.; Xie, S.; Pu, Y., *et al.* Synergistic enzymatic and microbial lignin conversion. *Green Chem*, **2016**, 18(5), 1306–1312.

[46] Le, R.; Wells, T.; Das, P., *et al.* Conversion of corn stover alkaline pre-treatment waste streams into biodiesel via Rhodococci. *RSC*, **2017**, 7, 4108–4115.

[47] He, Y.; Li, X.; Ben, H., *et al.* Lipid production from dilute alkali corn stover lignin by Rhodococcus strains. *ACS Sus Chem Eng*, **2017**, 5, 2302–2311.

[48] Vardon, D.R.; Franden, M.A.; Johnson, C.W., *et al.* Adipic acid production from lignin. *Energy Environ Sci*, **2015**, 8, 617–628.

[49] Salvachua, D.; Rydzak, T.; Auwae, R., *et al.* Metabolic engineering of Pseudomonas putida for increased polyhydroxyalkanoate production from lignin. *Microb Biotechnol*, **2020**, 13(1), 290–298.

[50] Masai, E.; Katayama, Y.; Fukuda, M. Genetic and Biochemical Investigations on Bacterial Catabolic Pathways for Lignin-Derived Aromatic Compounds. *Biosci, Biotech, and Biochem*, **2007**, 71, 1–15.

[51] Yaegashi, J.; Kirby, J.; Ito, M., *et al.* Rhodosporidium toruloides: a new platform organism for conversion of lignocellulose into terpene biofuels and bioproducts. *Biotech for Biofuels*, **2017**, 10(241).

[52] Kalin, M.; Neujahr, H.; Weissmahr, R., *et al.* Phenol hydroxylase from Trichosporon cutaneum: gene cloning, sequence analysis, and functional expression in Escherichia coli. *J of Bacteriol*, **1992**, 174, 7112–7120.

[53] Neujahr, H.; Gaal, A. Phenol hydroxylase from yeast. Purification and properties of the enzyme from Trichosporon cutaneum.. *European J of Biochem*, **1973**, 35(2), 386–400.

[54] Peelen, S.; Rietjens, I.; Boersma, M., *et al.* Conversion of phenol derivatives to hydroxylated products by phenol hydroxylase from Trichosporon cutaneum. A comparison of regioselectivity and rate of conversion with calculated molecular orbital substrate characteristics.. *European J of Biochem*, **1994**, 227, 284–291.

[55] Xu, D.; Ballou, D.; Massey, V. Studies of the mechanism of phenol hydroxylase: mutants Tyr289Phe, Asp54Asn, and Arg281Met.. *Biochem*, **2001**, 40, 12369–12378.

[56] Gerginova, M.; Manasiev, J.; Shivarova, N., *et al.* Influence of various phenolic compounds on phenol hydroxylase activity of a Trichosporon cutaneum strain. *J of Biosci*, **2006**, 62(1-2), 83–86.

[57] Gerginova, M.; Zlateva, P.; Peneva, N., *et al.* Influence of phenolic substrates utilised by yeast Trichosporon cutaneum on the degradation kinetics. *Biotechnol Biotechnol Equip*, **2014**, 28(1), 33–37.

[58] Gaal, A.; Neujahr, H.Y. Metabolism of Phenol and Resorcinol in Trichosporon-Cutaneum. *J Bacteriol*, **1979**, 137(1), 13–21.

[59] Spence, E.; Scott, H.; Dumond, L., *et al.* Characterisation of the hydroxyquinol degradation pathway in Rhodococcus jostii RHA1 and Agrobacterium sp.: an alternative pathway for degradation of protocatechuic acid and lignin degradation fragments. *Appl Environ Microbiol*, **1990**, 136, 1491–1495.

[60] Suzuki, K.; Itoh, M. Metabolism of p-Hydroxybenzoate via Hydroxyquinol by Trichosporon cutaneum WY2-2: Characterization of the Pathway using Superoxide Dismutase as a Stabilizer of Hydroxyquinol. *Plant Cell Physiology*, **1986**, 27(8), 1451–1460.

Chapter 5

Metabolic engineering and synthetic biology towards novel lipid production

5.1 Abstract

Non-model oleaginous yeasts are particularly well suited to produce value-added oleochemicals from low-cost substrates at industrially relevant titers, but often remain genetically intractable. Tools development is becoming more common, as researchers seek to capitalize on the inherent benefits of non-model organsisms. This chapter details the expansion of the genetic toolkit for *Cutaneotrichosporon oleaginosus*, noted for its ability to robust utilize all components of lignocellulosic biomass.

5.2 Introduction

Oleochemicals are chemicals derived from animal fats and plant oils that are used for a number of applications including surfactants, lubricants, cosmetics, biofuels, biodiesels and more. The increasing demand for biodiesel is contributing to agricultural demand, and soon we will surpass the oilseed production capability. The rise in healthy fat supplements, such as omega-3 fatty acids, are contributing to overfishing. Use of plant-derived oils such as palm oil and cocoa butter in cosmetics and skincare items are contributing to drastic deforestation. There has also been a long-standing interest in replacing petroleum-based products as a matter of economic and national security. Alternative sources are needed to meet the increasing demand of oleochemicals. Oleaginous yeasts are those that are able to accumulate more than 20% of its biomass as lipids. A significant amount of work has been done with *Yarrowia lipolytica*, one of the most well-studied oleaginous yeast. However, non-model yeasts such as *Rhodosporidium toruloides*, *Lipomyces starkeyi*, and *Cutaneotrichosporon oleaginosus* have unique features, such as natural utilization of low-cost substrates, that motivate developing them like *Y. lipolytica*. Particularly, *C. oleaginosus* is noted for its ability to utilize all components of lignocellulosic biomass, including by-products generated during processing [1–3]. There are few tools available for this yeast, and it remains a challenging yeast to engineer [4, 5]. This chapter highlights progress made towards expanding the genetic toolkit for *C. oleaginosus*.

5.3 Materials and Methods

5.3.1 Cloning

5.3.1.1 Cloning pRF and pUC19 plasmids

pRF-HU2 plasmids were purchased from Fungal Genetics Stock Center (FGSC)] and transformed into DH10β cells (NEB). All plasmids were miniprepped using the Zyppy Plasmid Kit from Zymo Research. Genomic DNA was extracted using the YeaStar Genomic DNA Kit from Zymo Research. All genes were synthesized by Genscript (Piscataway, NJ, USA) and codon optimized based off the codon usage table for Cyrptococcus curvatus (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=57679). All oligos were synthesized by IDT and are listed in Table 5.2. Sequencing was performed either by Eurofins Genomics or IDT.

The pRF plasmids were assembled in two different SLIC reactions. In the first reaction, a 1 kb promoter (P_1, P_2, P_4, and P_6) and GPD terminator, T_{GPD}, were PCR amplified from the *C. oleaginosus* genome. The GPD terminator was amplified using Term_For and Term_Rev. The promoters were amplified using P_x_For and P_x_Rev, where x corresponds to the number 1, 2, 4, or 6. The P_x, T_{GPD}, and pRF plasmid backbone were assembled using a 3-piece SLIC and verified by restriction digest and sequencing. In the second reaction, the trpC promoter, P_{trpC}, driving *hygR* was replaced with a *C. oleaginosus* promoter. The previously assembled plasmid was amplified using For_pTrpC and Rev_pTrpC. Each promoter (1, 2, 4, and 6) were amplified using For_ and Rev_P_x_Hyg primers, where x corresponds to 1, 2, 4, or 6. The two PCR fragments were assembled with a 2-piece SLIC. Successful assemblies were verified by restriction digest and sequencing.

pUC plasmids were designed by Stephen Lee. Individual pieces for the pUC19

vectors were PCR amplified. Fragment 1 consisted of promoter 2 and GOI CS (primers 3/4) and Fragment 2 consisted of the hygromycin cassette being driven by promoter 1 and the GPD terminator with homology to the GOI CS (primers 1/2). Both fragment were amplified from the pRF plasmids described above. Fragment 3 was the pUC19 backbone (primers 22/23). The three fragments were assembled using a 3-piece SLIC and verified by sequencing. The resulting plasmid is called pUC19-p1.2.

Genes of interested (GOIs) were PCR amplified from pUC57 vectors containing codon-optimized genes synthesized by Genscript. Codon optimization was based on the Kasuza codon usage table for *Cryptococcus curvatus* ATCC 20509, currently the only codon usage table available. A table of codon optimized genes are listed in Table 5.1. To insert a gene of interest (GOI), the pRF plasmids can be digested with BsqQI, while the pUC19-p1.2 vector can be PCR amplified. The GOI was assembled with either the pRF or pUC19 plasmid using a 2-piece SLIC. At present, cpFAH12 has been inserted into pRF-p2 and pRF-p4 and igASE2 has been inserted into pRF-p1. All three constructs have been sequence verified.

Table 5.2. Table of primers used to clone the pRF and pUC19 plasmids.

Primer	**Sequence (5' → 3')**
Term_For	GAAGAGCTTCTGGGAATGCCATGCTCTTC TCTTTCTAGGTTGTAGCATGGTTTCCGCG TA
Term_Rev	cgtcgagatctagaggatcccccgactagATCCGCTGACA TTGGACCTTTTGGTGCA
For_pTrpC	ggcctgatcatcgatgggc
Rev_pTrpC	cgatatgaaaaagcctgaactcacc
P_1_For	gagcggagaattaagggagtcacgaagctCCGTATCAGG GTTTGGGTAC

Continued on next page

Table 5.2 – *Continued from previous page*

Primer	**Sequence (5' → 3')**
P_1_Rev	AGAAGAGCATGGCATTCCCAGAAGCTCTT CATGAACCACGTCAGCGGTTGA
P_2_For	gagcggagaattaagggagtcacgaagctCCGGCGTCTA AAGCCAGAGGAT
P_2_Rev	AGAAGAGCATGGCATTCCCAGAAGCTCTT CAAGAAACGGGGGGAGTGATGG
P_4_For	gagcggagaattaagggagtcacgaagctGTCGCCATCG CACCTTCGGTT
P_4_Rev	AGAAGAGCATGGCATTCCCAGAAGCTCTT CAGATGAATGAGGATGGAGTGAC
P_6_For	gagcggagaattaagggagtcacgaagctCGTGGAACTA GATACCGCAT
P_6_Rev	AGAAGAGCATGGCATTCCCAGAAGCTCTT CAtggagCGTCGCCACTTTAAGTA
For_P_1_Hyg	gcccatcgatgatcaggccCCGTATCAGGGTTTGGG TAC
Rev_P_1_Hyg	ggtgagttcaggctttttcatatcgTGAACCACGTCAGC GGTTGA
For_P_2_Hyg	gcccatcgatgatcaggccCCGGCGTCTAAAGCCAG AGGAT
Rev_P_2_Hyg	ggtgagttcaggctttttcatatcgAGAAACGGAGTGAT GG
For_P_4_Hyg	gcccatcgatgatcaggccGTCGCCATCGCACCTTC GGTT
Rev_P_4_Hyg	ggtgagttcaggctttttcatatcgGATGAATGAGGATG GAGTGACTGTGA
For_P_6_Hyg	gcccatcgatgatcaggccCGTGGAACTAGATACCG CAT
Rev_P_6_Hyg	ggtgagttcaggctttttcatatcgtggagCGTCGCCACTT TAAGTA
Primer 1	GCCTGCAGGTCGACTCTAGAGGATCGGtgg caggatatattgtggtgtaaac
Primer 2	TGAAGAGCTTCTGGGAATGC
Primer 3	GCATTCCCAGAAGCTCTTCAAGAAACGGG GGGAGTGATGG
Primer 4	TCGAGCTCGGTACCCGGGGATCCCGGCGT CTAAAGCCAGAGGAT
Primer 22	GATCCTCTAGAGTCGACCTG
Primer 23	GATCCCCGGGTACCGAG

Table 5.1. A list of genes that were codon optimized for *C. oleaginosus*.

Source Organsism	**Gene**	**Function**
Claviceps purpurea	cpFAH12	Δ12-desaturase
Hiptage benghalensis	hbFAH12	
Ricinus communis	rcFAH12	
Isochrysis galbana	igASE2	C18-Δ9 elongase
Escherichia coli	ecCPS	cyclopropane synthase
Gossypium hirsutum	ghCPS1	
Sterculia foetida	sfLPAT	lysophosphatidic acid acyltransferase
Klebsiella pneumoniae	neoR	aminoglycoside 3'-phosphotransferase
Streptoalloteychus hindustanus	bleoR	phleomycin resistance

5.3.2 Agrobacterium-mediated transformation

Cells were transformed using *Agrobacterium tumefaciens*-mediated transformation (ATMT) per Gorner, et al [4]. Membranes used for recovery of transformed cells include Whatman non-sterile cellulose nitrate membrane, 0.45 μm (cat. no. 7184-004), Whatman nylon membrane, 0.45 μm (cat. no. 7404-004), and Whatman cellulose acetate membrane, 0.45 μm (cat. no. 7184-004) from GE Healthcare (Little Chalfont, Buckinghamshire, UK). Cells were recovered on membranes at 24 °C for 48 hours before membranes were transferred to YPD agar plates containing 200 μg/mL hygromycin B and 300 μg/mL cefotaxime and incubated at 28 °C until colonies were observed. Separately, cells were washed from the membrane after recovery using sterile water. The entire cell resuspension was plated onto selective YPD agar plates and incubated at 28 °C until colonies were observed. Colonies were screened for positive transformation using colony PCR using primers Hyg-F_cPCR_AY (5' ATAGCTGC GCCGATGGTTTC 3') and Hyg-R_cPCR_AY (5' GGACGATTGCGTCGCATCGA 3').

5.3.3 Electroporation

5.3.3.1 Method 1: Sorbitol buffer, no pretreatment

Yeast cells were grown in YPD to mid-exponential phase and harvested by centrifugation. The yeast cell pellet was washed two times with 250 mL of ice-cold deionized water. Then the cells were washed once with 50 mL of ice-cold 1 M sorbitol solution and the cell pellet was resuspended in 0.5–1 mL of ice-cold 1 M sorbitol solution [6]. The final cell suspension was at a density of about 1×10^{10} cells/mL.

5.3.3.2 Method 2: Tris buffer, DTT pretreatment

Yeast cells were grown in YPD to mid-exponential phase and harvested by centrifugation. After centrifugation, the cell pellet was resuspended in 80 mL of deionized water. Then 10 mL of 10X TE buffer and 100 mL of 1 M LiAc were mixed with the cell suspension and the cells were shaken gently at 25 °C for 45 min, then 2.5 mL of 1 M DTT buffer was added and were shaken at the same conditions for 15 min. Then the cells were collected by centrifugation to remove the supernatant. The cells pellet was washed twice with the ice-cold deionized water and washed once with ice-cold 1 M sorbitol. Finally, the cell pellet was resuspended with 0.5–1 mL of ice-cold 1 M sorbitol [7]. The final cell suspension was at a density of about 1×10^{10} cells/mL.

5.3.3.3 Method 3: Tris buffer, LiAc and DTT pretreatment

Yeast cells were grown in YPD to mid-exponential phase and harvested by centrifugation. The cell pellet was resuspended in 10 mM Tris-Cl buffer (pH 7.5) containing 1 mM $MgCl_2$, 100 mM LiAc, 270 mM sucrose, and 5 mM DTT, and the cells were shaken gently at 25 °C for 60 min. Then the cells were subjected to a

centrifugation at 4,000 x g for 5 min at 4 °C, and the cells were washed twice with 10 mM Tris-Cl buffer (pH 7.5) containing 1 mM $MgCl_2$ and 270 mM sucrose and resuspended in the same buffer [8, 9]. The final cell suspension was at a density of about 1×10^{10} cells/mL.

5.3.3.4 Method 4: Tris buffer, DTT pretreatment

Optimization of conditions Plasmids were miniprepped using the Zyppy Miniprep kit and 6 μg was digested with either PacI/NotI or PacI/PmeI. The digested DNA reaction was not purified and was used directly for electroporation. Yeast cells were grown to mid-exponential in YPD media. Cells were harvested by centrifugation at 3,200 x g for 4 min. Supernatant was decanted and cells were resuspended in 5 mL ice-cold EB buffer (10 mM Tris-HCl (pH 7.5), 270 mM sucrose, 1 mM $MgCl_2$). Cells were centrifuged as before and then washed and resuspended in 5 mL of IB buffer (25 mM DTT, 20 mM Hepes (pH 8.0) in YPD). Cells were incubated at 30 °C for 30 minutes, shaking at 250 rpm. Cells were harvested by centrifugation at 3,200 x g for 5 mins. Cells were resuspended with 5 mL ice-cold EB buffer and centrifuged as before. Finally, cells were resuspended in 500 μL of ice-cold EB buffer [5]. Experiments to optimize electroporation conditions changed voltage between 800 and 1800 V and mass of DNA transformed between 2, 4, and 6 μg. After electroporation, cells were recovered in 1 mL YPD and incubated in 50 mL Falcon tubes at 30 °C, shaking at 250 rpm, for 4 hours. Cells were plated on YPD + 200 μg hygromycin.

Electroporation of protoplasts Electroporation of protoplasts followed the same protocol above, but instead of cells being harvested from a YPD culture, the prepared Glucanex-treated protoplasts in Buffer 3 were used. Cells were transformed using 1800 V with 6 μg of digested plasmid DNA.

5.3.4 Protoplasting

5.3.4.1 Glucanex

Glucanex was purchased from Sigma-Aldrich (L1412-5G). A protocol originally designed for fusion of *Rhodotorula rubra* with *Saccharomyces cerevisiae* was adapted for *C. oleaginosus* [10]. *Cutaneotrichosporon oleaginosus* cells were grown in YPD overnight or until stationary phase at 28 °C and 250 rpm. Cells were harvested by centrifugation and washed twice with TE buffer (50 mM Tris, pH 6.5 + 5 mM EDTA, pH 8). Cells were resuspended to a final concentration of 10^9 cells/mL in the TE buffer. Glucanex powder was added to Buffer 3 (TE buffer + 1.5 M mannitol) to a concentration of 1.5% (w/v). Cells were incubated undisturbed on the benchtop during treatment for 2 hours total, with samples being taken at 1 hour and 2 hour mark.

5.3.4.2 Lallzyme-MMX

Lallzyme-MMX was purchased from Scott Laboratories (016207). This protocol was originally designed to generate protoplasts for fission yeast, specifically *Schizosaccharomyces pombe*, but was adapted for *C. oleaginosus* here [11]. *C. oleaginosus* cells were grown in a 2 mL YPD pre-culture overnight at 28 °C and 250 rpm. The pre-culture was used to inoculate 15 mL of fresh YPD in a 50 mL baffled flask. Cells were incubated as before until reaching a cell count of 6-8 x 10^6 cells/mL. Cells were centrifuged at 1,500 x g for 5 mins and washed twice with 15 mL of SCS buffer. SCS buffer was made by mixing equal volumes of SCSa (20 mM sodium citrate + 1 mM D-sorbitol) and SCSb (20 mM citric acid + 1 mM D-sorbitol). The SCS buffer was adjusted to a pH of 5.8 by adding more SCSb buffer. Cells were resuspended in 0.5 mL SCS buffer containing 0.1 g/L Lallzyme-MMX and incubated on the benchtop

undisturbed for a total of 3.5 hours. Cells were evaluated for protoplast formation at 5, 10, 20, 90, 150, and 210 minutes.

5.3.4.3 Protoplasting efficiency measurement

Protoplast efficiency was evaluated by how sensitive cells were to SDS treatment for treated cells (digested with protoplasting enzymes) and control cells (cells incubated in identical conditions without protoplasting enzymes). To test protoplast efficiency, cells were incubated in a 1% SDS solution for 5 minutes and then measured in a cuvette at 600 nm on a Nanodrop 2000. Simultaneously, cells were incubated in water for 5 minutes and then measured in a cuvette at 600 nm on a Nanodrop 2000. The protoplast efficiency is the ratio between the cells incubated in 1% SDS versus water [10].

5.3.5 Live cell transformation

Cells were transformed using a live cell transformation method orignially developed for improved homologous recombination in *Yarrowia lipolytica* [12]. Briefly, *C. oleaginosus* cells were grown in a 2 mL YPD pre-culture overnight at 28 °C and 250 rpm. Cells were inoculated into a 125 mL baffled flask containing 25 mL of YPD to an OD of 0.5 and cultured as previously. After 3 hours of growth, solid hydroxyurea was added directly to the media to a final concentration of 100 mM. Cells were incubated on the benchtop undisturbed for 2 hours. After, cells were centrifuged at 4,000 x g for 2 min at 4 °C. Cells were washed with 25 mL of sterile ddH_2O and spun as before. Cells were resuspended in a volume equivalent to the pellet (~ 1 mL). 50 μL of the resuspension was aliquoted into a fresh 1.7 mL tube. Each aliquot received 9 μL of unpurified PCR product and 92 μL of transformation mix (80 μL of 60%

PEG 4,000, 5 μL 2 M DTT (aq), 5 μL 2 M lithium acetate (aq) (pH 6.0), and 2 μL 10 mg/mL single-stranded salmon sperm DNA). The cells, DNA, and transformation mix were vortexed well and heat shocked at 39 °C for 1 hour. Cells were centrifuged as before and the supernatant was aspirated off. The cells were resuspended in 1 mL YPD and transferred to a culture tube. Cells were incubated at 28 °C overnight and then plated on YPD plates containing 300 μg/mL hygromycin B.

5.4 Results

5.4.1 Successful implementation of four novel, strong, constitutive promoters

The top ten most highly expressing genes across all tested conditions were identified from the previously published dataset by Kourist, et al. and are detailed in Table 5.3 [13]. Roughly 1 kb upstream of each gene was designated as the promoter regions. Of the top ten promoters identified, four were successfully amplified (p1, p2, p4, and p6) (Figure 5.1). These four were carried forward for cloning. It should be noted that my efforts to clone the GPD promoter as detailed by Gorner, et al. were unsuccessful [4]. I presume it is the high GC content of the promoter, and subsequent work engineering this yeast used a TEF promoter rather than the GPD promoter [5].

These promoters were added upstream of a cloning site (CS) containing two sites for BspQI, a type II restriction enzyme (Fig. 5.2). The gene cassette was placed between the left-bound (LB) and right-bound (RB) regions so that the plasmids could be used for *Agrobacterium tumefaciens*-mediated transformation (ATMT).

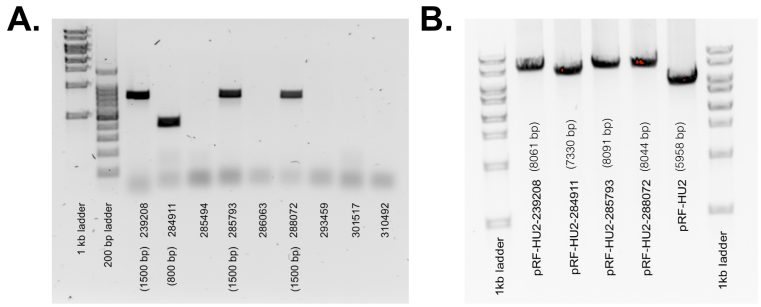


Figure 5.1. A. Set of 10 strong, constitutive promoters that were amplified for use in genetic engineering of *C. oleaginosus* and B. assembly of the promoters into pRF plasmids for use in *Agrobacterium*-mediated transformation.

Table 5.3. The top ten highly expressing genes selected for a suite of strong constitutive promoters.

Name	**Gene identifier**	**Gene Function**
p1	CC85DRAFT_286063	Glutamate/Phenylalanine dehydrogenase
p2	CC85DRAFT_288072	Nucleotide binding
p3	CC85DRAFT_293459	Glutamine synthetase
p4	CC85DRAFT_301517	Glycine hydroxymethyltransferase
p5	CC85DRAFT_310492	Pyruvate carboxyltransferase
p6	CC85DRAFT_239208	Glycogen synthase kinase-3
p7	CC85DRAFT_276138	Glyceraldehyde 3-phosphate dehydrogenase
p8	CC85DRAFT_284911	Chitinase
p9	CC85DRAFT_285494	Ketose-biphosphate aldolase, class II
p10	CC85DRAFT_285793	Transaldolase AB

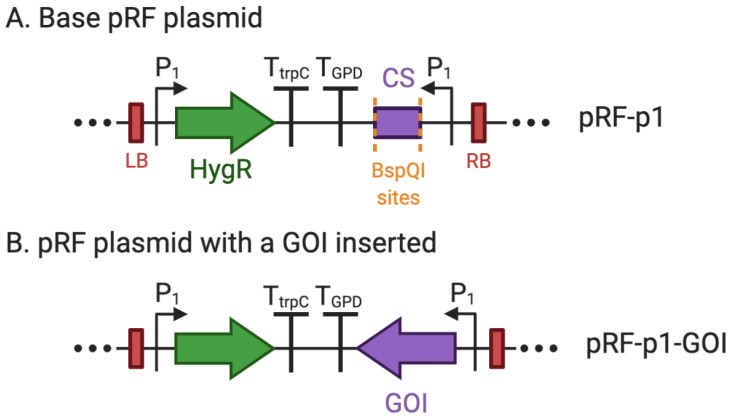


Figure 5.2. Schematic for the pRF plasmids created during cloning. LB = left bound DNA, P_1 = promoter, HygR = hygromycin resistance gene, T_{trpC} = *E. coli* tryptophan terminator, T_{GPD} = *C. oleaginosus* GPD terminator, CS = cloning site, RB = right bound DNA, and GOI = gene of interest.

pRF plasmids are a medium copy plasmid in *E. coli* and we desired a high copy plasmid backbone. The pRF plasmid DNA from the LB to the RB regions were transferred into a pUC19 backbone. Simultaneously, the promoter driving the hygromycin resistance gene was deliberately changed to be different from the promoter upstream of the gene of interest. This allows PCR amplification of the DNA between the LB and RB regions.

5.4.2 Optimization of Agrobacterium-mediated transformation

Agrobacterium tumefaciens-mediated transformation (ATMT) resulted in very few colonies and can take anywhere from 5 days to several weeks. In my hands, using ATMT to introduce a hygromycin resistance gene resulted in colonies on the order of 1 to 13 after approximately one week. Nylon and cellulose acetate membranes resulted in 13 CFU and 10 CFU, respectively, while the cellulose nitrate membranes

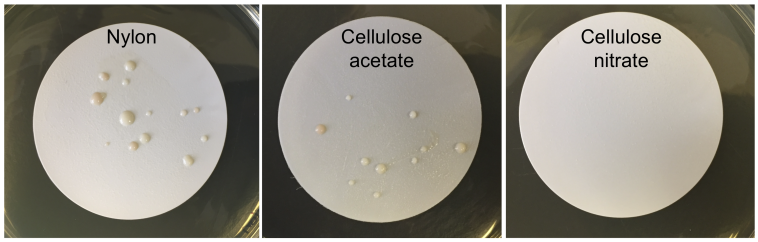


Figure 5.3. Colonies resulting from Agrobacterium-mediated transformation.

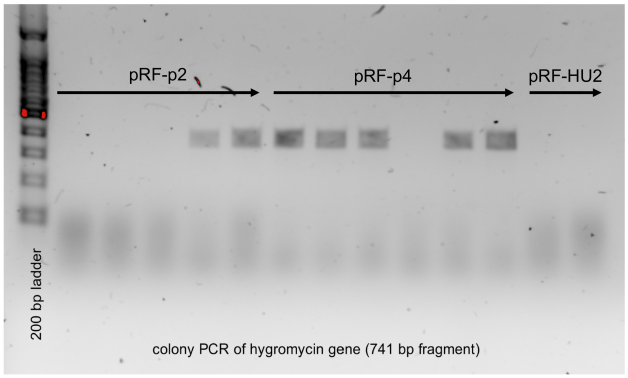


Figure 5.4. Successful integration of the hygromycin resistance gene using *Agrobacterium*-mediated transformation method.

did not result in any colonies (Fig. 5.3). Washing cells from the membrane after recovery did not result in any colonies. Colonies were analyzed from each membrane through colony PCR (Fig. 5.4). Not all colonies from each plate were analyzed. The pRF-p2 colonies were screened from cellulose acetate membranes and the pRF-p4 colonies were screened from nylon membranes. The nylon membranes resulted in five out of six successfully intergrated colonies while the cellulose acetate showed two out of five successfully integrated colonies.

5.4.3 Protoplasting

Some yeast cell membranes are covered by a thick cell wall to protect the cell. These protective cell walls are typically composed of chitin, 1,6-β-glucan, 1,3-β-glucan, and mannoproteins (Fig. 5.5). This shell often inhibits both DNA extraction

and DNA introduction. Often, the cell wall must first be digested, a process called protoplasting. Because the cell wall provides structure and protection, removal of the cell wall results in very delicate cells, and they need to be suspended in an osmostabilizer after being protoplasted. As every cell wall is composed of different ratios of chitin, β-glucan, and mannoproteins, optimization with different enzymes, ratios of enzymes, and time of digestion will prevent over-digestion of the cell wall, which can decrease cell survival.

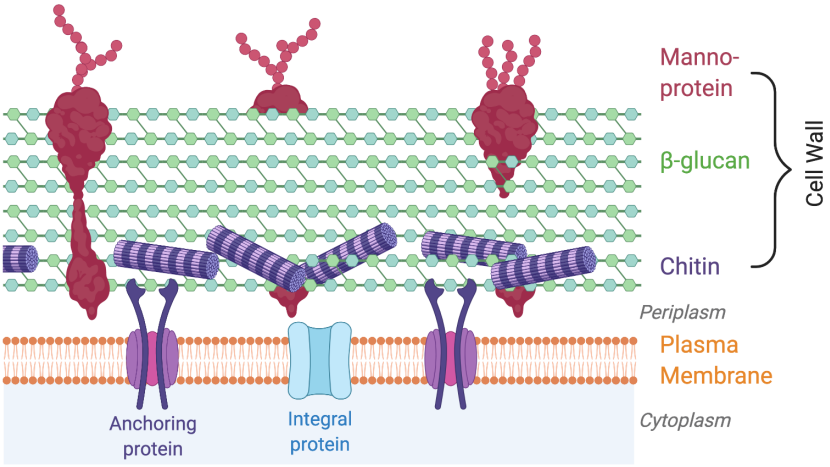


Figure 5.5. Schematic of a yeast cell wall structure.

Two enzyme products were used to protoplast *C. oleaginosus* cells. Glucanex is a cocktail derived from *Thrichoderma harzinanum* consisting of β-glucanase, cellulase, protease, and chitinase. Glucanex is a replacement for Novozyme 234, which was discontinued. Lallzyme-MMX is a combination of glucosidase and pectinase and is used in wine making to break down yeasts after fermentation and improve filterability. The mechanism for glucosidase and chitinase are shown in Figure 5.6. The red arrows indicate where cleavage occurs.

Protoplasting was shown to be effective (Figure 5.7). Across all conditions, control cells, or those not treated by protoplasting enzymes, were shown to have lit-

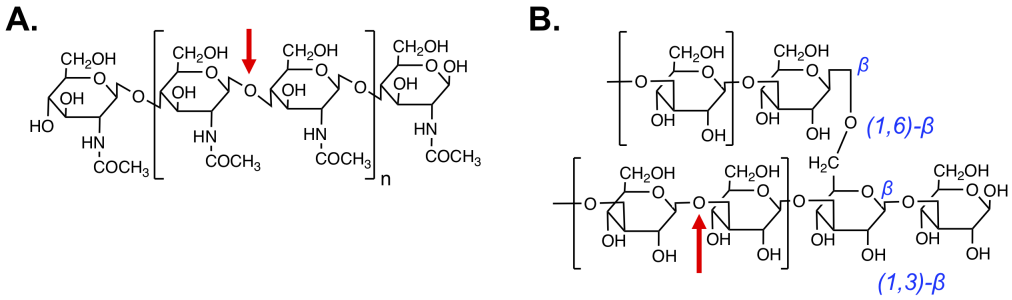


Figure 5.6. Mechanism for cleavage (red arrow) for A. chitinase and B. glucosidase. Data are for single replicates.

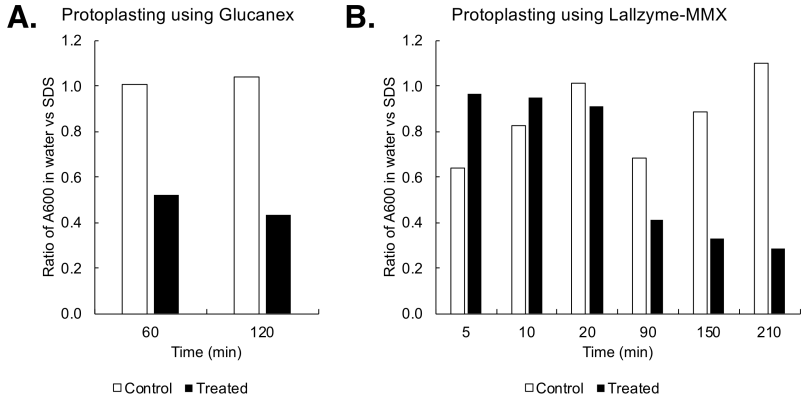


Figure 5.7. Protoplast efficiency for A. Glucanex and B. Lallyzme-MMX treatments. Data are for single replicates.

tle decrease in A600 in the presence of 1% SDS. This indicates they are resistant to SDS lysing. Treated cells, or those digested with protoplasting enzymes, were shown to have a decreased A600 in the presence of 1% SDS. This indicates treated cells are sensitive to SDS, presumably due to a digested cell wall decreasing osmostability. Glucanex does not seem to require longer than 60 minutes to digest, as there is little difference in the effect of the SDS incubation at 120 minutes (Figure 5.7A). Incubation of cells with Glucanex can be shortened to see if less time is required for the same protoplasting efficiency. Lallzyme-MMX required longer than 20 minutes to show protoplast formation. There is a point of diminishing returns after 90 minutes, after which the protoplast formation seems to significantly slow (Figure 5.7B). To improve protoplasting further, cells can be incubated with reducing agents such as β-mercaptoethanol (BME) or dithiolthrietol (DTT). The presence of BME with Glucanex slightly improved protoplasting efficiency (Figure 5.8). While there is a tradeoff between cell protoplast efficiency and cell survival, this was not probed in this experiment. The protoplasts were used for transformation of either water (-DNA) or linearized DNA (+DNA). Unfortunately, these transformations did not result in any colonies.

5.4.4 Electroporation and live cell transformation

Stephen Lee optimized the electroporation protocol based on that of Koivuranta, et al. [5]. It was found that optimal transformation conditions are at 1600 V with 4 μg of linearized DNA (Figure 5.9). Other pre-treatments (lithium acetate and DTT) and osmostabilizers (Tris, sucrose, and sorbitol) did not result in efficiencies higher than that of Koivuranta's method original method with electroporation conditions of 1600 V and 4 μg DNA (Figure 5.10). The electroporation protocol was not

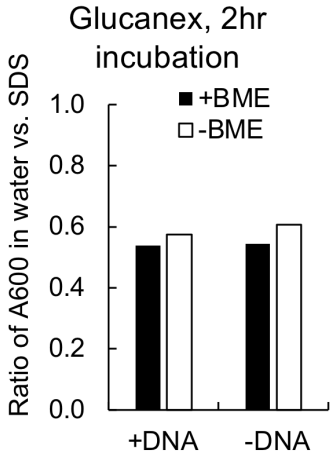


Figure 5.8. Protoplast efficiency with and without BME in preparation for transformation. Data are for single replicates.

re-optimized for use in combination with protoplasting.

A live cell transformation protocol from *Yarrowia lipolytica* was applied to *C. oleaginosus* with success. The protocol is both easier to work with and faster than both ATMT and electroporation. All seven of the colonies tested for integration of the hygromycin resistance gene showed success (Figure 5.11).

5.5 Discussion

A significant body of work characterizing and developing tools for the model oleaginous yeast *Yarrowia lipolytica* provides a guideline for the tools necessary to develop for *Cutaneotrichosporon oleaginosus* and other non-model yeast. Genetic engineering in *C. oleaginosus* is currently laborious and slow. There are only two published promoters and one selection marker. Future work engineering multiple genes into the genome or creating knock-outs will require much more robust tools.

The data presented here demonstrate use of four novel strongly expressing, constitutive promoters. Each promoter was used to drive expression of a hygromycin

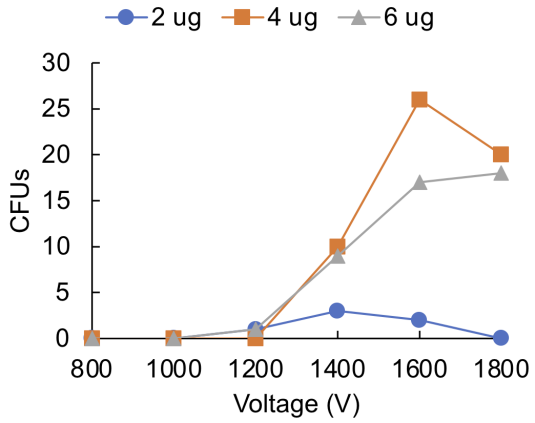


Figure 5.9. Optimization of electroporation protocol. Data are for single replicates.

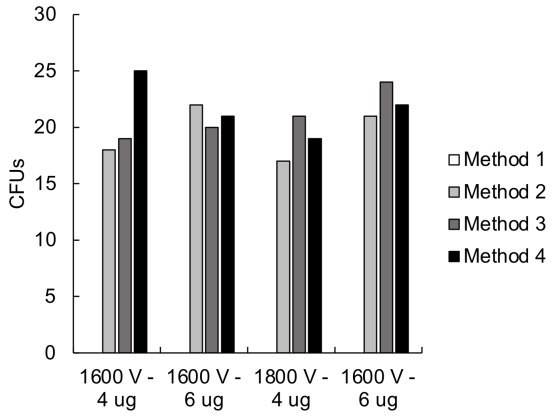


Figure 5.10. Optimization of electroporation protocol using 4 different methods: method 1: sorbitol buffer, no pretreatment [6]; method 2: Tris buffer, DTT pretreatment [7]; method 3: Tris buffer, LiAc and DTT pretreatment [8, 9]; method 4: Tris and sucrose buffer, DTT pretreatment [5]. Data are for single replicates.

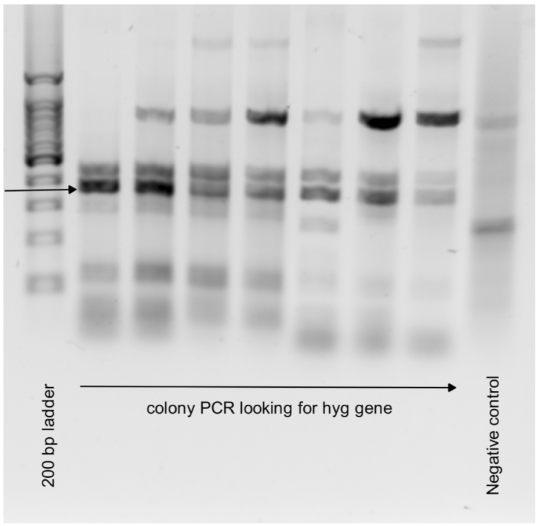


Figure 5.11. Successful integration of the hygromycin resistance gene using the live cell transformation method.

resistance gene and genomically integrated using three different transformation methods. Previous studies have used the native GPD promoter [4] and the native TEF promoter [5]. I was unable to clone the GPD promoter and did not try the TEF promoter.

The transformation methods tested included *Agrobacterium*-mediated transformation (ATMT), electroporation, and a live cell integration method adapted from *Y. lipolytica* [4, 5, 12]. ATMT was a reliable method, frequently used with fungi and basidiomycete yeast, that always resulted in colonies and integration. ATMT is a laborious, time-consuming method, and other methods such as electroporation and live cell transformations have been developed to have higher throughput and more approachable protocols. The fastest, easiest, and most efficient transformation method was the live-cell method. *Cutaneotrichosporon oleaginosus* has a more recalcitrant shell of carbohydrates surrounding its cell membrane, and the addition of DTT likely increased the cell wall permeability. The electroporation protocol was optimized by Stephen Lee and showed slight improvements over the existing protocol. He tried

varying methods to include different osmostabilizers, such as Tris, sucrose, and sorbitol, and pretreatments, such as DTT and LiAc. Of the methods tried, following the protocol by Koivuranta, et al., but with 1600 V and 4 μg of DNA resulted in the most colonies after selection. To enhance transformation efficiency, protoplasting was successfully used to degrade the carbohydrate shell surrounding the cell and increase cell wall permeability. The same electroporation method was used to transform protoplasted cells, which was likely too harsh of a method for the delicate cells. Future studies should decrease the voltage used. Since the cell wall is more permeable, it's possible less DNA will be required for successful transformation.

Other attempts to improve the genetic toolkit include developing an auxotrophic URA3 deficient strain. Establishing an auxotrophic strain would be beneficial for the implementation of self-replicating plasmids. 1 kb homology arms to target the URA3 gene flanked the integration cassette such that the entire region between the LB and RB would be integrated into the URA3 gene. The URA3 targeting cassette was transformed with ATMT and selection was performed with 5-FOA. No successful integrations were observed. 35 bp homologies were used instead of 1 kb arms, transformed using the live cell method, and was also unsuccessful [12]. Regions of high GC are difficult to work with, so instead of removing the entire URA3 gene, a smaller knockout region with a more optimal GC content could be targeted. Additionally, random mutagenesis coupled to 5-FOA screening may achieve the same goal in a much faster, non-targeted way. This will be less ideal since mutations would be generated across the entire genome.

5.6 Conclusions

Genetic engineering is necessary to facilitate tailored oleochemical production in *Cutaneotrichosporon oleaginosus*. The combination for robust aromatic catabolism and superb lipid production makes this yeast a strategic starting organism to produce value-added oleochemicals. However, the low throughput and efficiency of the existing transformation methods prevent this yeast from being genetically tractable. In this chapter, I discuss enhancements to the electroporation protocol for this yeast as well as propose a novel live cell method that is easier and faster than the electroporation method. A suite of strong constitutive promoters were identified from RNAseq data and tested for functionality. Even with these improvements and additions to the genetic toolbox for *C. oleaginosus*, tools such as self-replicating plasmids, aromatic-inducible promoters, auxotrophic strains, and targeted homologous recombination tools will be invaluable for future engineering efforts.

5.7 Acknowledgments

This material is based upon work supported by the US Department of Agriculture Sun Grant, Creative Inquiry funds. I would like to acknowledge Stephen Lee for his work optimizing the electroporation protocols and designing the pUC19 vectors.

Bibliography

[1] Yaguchi, A.; Rives, D.; Blenner, M. The New Kids on the Block: Emerging Oleaginous Yeast of Biotechnological Importance. *AIMS Microbiology*, **2017**, 3(2), 227–247.

[2] Yaguchi, A.; Spagnuolo, M.; Blenner, M. Engineering yeast for utilization of alternative feedstocks. *Curr Opin Biotechnol*, **2018**, 53, 122–129.

[3] Yaguchi, A.; Spagnuolo, M.; Blenner, M. Oleaginous yeast for biofuel and oleochemical production. *Curr Opin Biotechnol*, **2019**, 57, 73–81.

[4] Görner, C.; Redai, V.; Bracharz, F., *et al.* Genetic engineering and production of modified fatty acids by the non-conventional oleaginous yeast Trichosporon oleaginosus ATCC 20509. *Green Chem.*, **2016**, 18(7), 2037–2046.

[5] Koivuranta, K.; Castillo, S.; Jouhten, P., *et al.* Enhanced triacylglycerol production with genetically modified Trichosporon oleaginosus. *Front Microbiol*, **2**, 9, 1337.

[6] Becker, D.; Guarente, L. High-efficiency transformation of yeast by electroporation. *Guide to Yeast Genetics and Molec Biol Methods in Enzymology*, **1991**, 182–187.

[7] Becker, D.; Lundblad, V. Metabolic engineering of Corynebacterium glutamicum for the production of cis, cis-muconic acid from lignin-vol. 20, Wiley Online Library, 2001.

[8] Takahashi, S.; Okada, H.; Abe, K., *et al.* Metabolism of Gallic Acid and Syringic Acid by Pseudomonas putida, journal = Appl Biochem Microbiol, volume = 50, pages = 624-628, year = 2014, type = Journal Article. *???*.

[9] Lin, X.; Wang, Y.; Zhang, S., *et al.* Functional integration of multiple genes into the genome of the oleaginous yeast Rhodosporidium toruloides. *FEMS yeast res*, **2014**, 14(4), 547–555.

[10] Evans, C.T.; Conrad, D. An improved method for protoplast formation and its application in the fusion of Rhodotorula rubra with Saccharomyces cerevisiae. *Arch Microbiol*, **1987**, 148, 77–82.

[11] Flor-Parra, I.; Zhurinsky, J.; Bernal, M., *et al.* A Lallzyme MMX-based rapid method for fission yeast protoplast preparation. *Yeast*, **2013**, 31(2), 61–66.

[12] Tsakraklides, V.; Brevnova, E.; Stephanopoulos, G., *et al.* Improved Gene Targeting through Cell Cycle Synchronization. *PLoS ONE*, **2015**, 10(7), e0133434.

[13] Kourist, R.; Bracharz, F.; Lorenzen, J., *et al.* Genomics and Transcriptomics Analyses of the Oil-Accumulating Basidiomycete Yeast Trichosporon oleaginosus: Insights into Substrate Utilization and Alternative Evolutionary Trajectories of Fungal Mating Systems. *MBio*, **2015**, 6(4), e00918.

Chapter 6

Conclusions and Future Directions

With the rise of biorefineries, lignin will be produced at unprecedented rates and its valorization is key for economic success. Yet, lignin's heterogeneous and recalcitrant nature make its utilization a complex and difficult problem to tackle. The work in this dissertation was motivated by a need to convert lignin-based feedstocks into value-added biochemicals. The unique ability of *Cutaneotrichosporon oleaginosus* to convert lignin-derived monoaromatic compounds and remain highly oleaginous hinted at its ability to metabolize lignin hydrolysates. When grown with a corn stover lignin hydrolysate, cells were able to metabolize released monoaromatics and in low nitrogen conditions, retain oleaginous behavior. A more fundamental study to understand the genetics that facilitate growth on and tolerance to phenolic compounds identified a number of novel genes found in funneling, ring-cleavage, and β-ketoadipate pathways. Collectively, *C. oleaginosus* has proven to be a strong potential model for lignin to oleochemical conversion, and the fundamental understanding of aromatic catabolism identified for this yeast will contribute valuable knowledge for aromatic catabolism across all yeast species.

6.1 Conclusions

In Chapter 2, 36 oleaginous yeast were screened for catabolism of six aromatic compounds. It was observed that basidiomycetes were stronger metabolizers than ascomycetes, and the best growth across the most number of aromatic compounds were seen from *Rhodosproridium*, *Trichosporon*, and *Cutaneotrichsporon* genera. Batch culturing of *Cutaneotrichosporon oleaginosus* resulted in lipid accumulation up to 50% on a dry cell basis.

Chapter 3 furthered catabolic analysis in high and low nitrogen sources to probe lipid accumulation, lipid profile, and substrate utilization of three aromatic compounds. Cells cultured using fed-batch fermentation resulted in 69.9% lipid accumulation on a per cell basis — the second highest value reported for this yeast and one of the highest reported lipid accumulation of any wild-type oleaginous yeast utilizing model lignin monoaromatics as the sole carbon source.

In Chapter 4, growth studies were expanded from model aromatic compounds to lignin hydrolysate and established that monomers released during hydrolysis were catabolized towards lipid production, though high molecular weight lignin remained polymeric. A mutli-omic investigation to elucidate genes related to aromatic catabolism discovered previously unidentified genes involved in funneling, ring-cleavage, and β-ketoadipate pathways. Differential gene expression of redox balancing proteins, transporters, and proteins altering membrane structure provide insights for tolerance mechanisms utilized by this yeast. These are novel discoveries and provide important information for understanding aromatic catabolism in yeast species.

The novel genes discovered in Chapter 4 provides a roadmap for future engineering work to increase flux through metabolism, enhance aromatic tolerance, eliminate metabolic bottlenecks, and shunt flux through competing pathways to allow

accumulation of desired products. However, the genetic inaccessibility of *C. oleaginosus*, as is the case with many non-model organisms, is a greatly limiting factor preventing these engineering efforts from being realized. The genetic tools developed in Chapter 5 will not only facilitate functional studies to support the predictions made in Chapter 4, but allow future strain engineering to create a even more robust lignin to oleochemical production strain.

6.2 Recommendations

The recommendations discussed are the opinions of this author. They are broken into two categories: 1) fundamental work towards understanding yeast aromatic catabolism and tolerance and 2) production of natural products.

6.2.1 Fundamental work

6.2.1.1 Discover ligninolytic activity from oleaginous yeasts

It was noted in Chapter 4 that *C. oleaginosus* cannot depolymerize high molecular weight lignin. There is debate whether consolidated bioprocessing or a step-wise process is better for lignin valorization. Consolidated bioprocessing will require fungal or bacterial ligninolytic genes to be expressed in *C. oleaginosus*. This is not outside the realm of possibilities, given that heterologous expression has successfully implemented in this yeast. Basidiomycete yeasts, which are more related to fungal systems than ascomycetes, may be better for expressing fungal enzymes required for lignin depolymerization. However, it may be better to forego consolidated bioprocessing and instead search for a yeast that can be co-cultured alongside *C. oleaginosus* that can do the depolymerization. In this manner, *C. oleaginosus* can be a microbial sink for

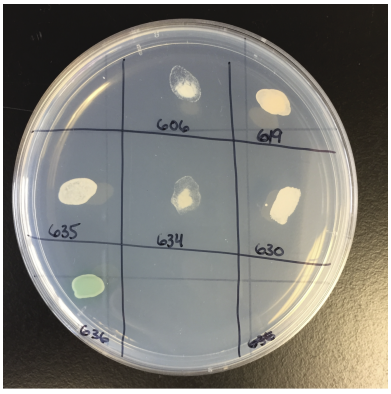


Figure 6.1. ABTS plate-based screening to detect laccase activity using strains: 606 (*Solicoccozyma phenolicus*, 619 (*Rhodotorula araucariae*), 635 (*Pseudohyphozyma borgoiensis*), 634 (*Naganishia albida*), 630 (*Filobasium magnus*), and 636 (*Vanrija musci).*

the released aromatic monomers, creating a driving force for lignin depolymerization, removing monomers to prevent re-polymerization, and eliminate toxicity through the accumulation of monomers. The same Phaff collection yeasts screened in Chapter 2 can be screened for ligninolytic capability by growing them on minimal media containing ABTS. Yeasts spotted on such plates will exhibit a green halo in the presence of ABTS if they express and secrete active peroxidases or laccases. A preliminary study was carried out and has identified *Vanrija musci* as such a yeast (Figure 6.1) based on a protocol from the literature [1]. It remains untested if *V. musci* can grow in a lignin hydrolysate, break polymeric lignin into monomers, or be grown in a co-culture with *C. oleaginosus*. Similarly, perhaps a two-stage fermentation can be developed such that in the first phase, an organism secreting ligninolytic enzymes can be first cultured with lignin. In the second phase, the fermentation broth from phase one can be used to culture *C. oleaginosus*, which can then utilize the released monomers. The secreted ligninolytic enzymes will still be present in the fermentation broth, and can perhaps continue to depolymerize the lignin as *C. oleaginosus* consumes the aromatics released.

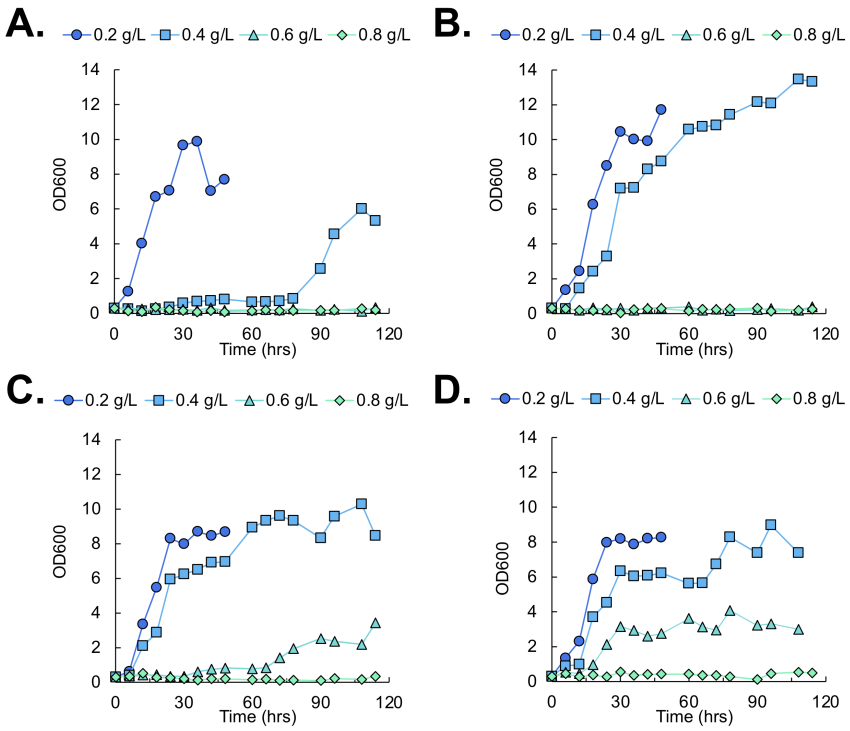


Figure 6.2. Growth of *C. oleaginosus* in TOHN media utilizing guaiacol, ferulic acid, syringic acid, and *p*-coumaric acid as sole carbon sources at concentrations of 0.2 g/L (circles), 0.4 g/L (squares), 0.6 g/L (triangles), and 0.8 g/L (diamonds). Data are for single replicates.

6.2.1.2 Catabolism of G- and S-type lignin

Softwood lignin is primarily comprised of G lignin and hardwood lignin is primarily S lignin. Both G and S monomers have at least one methoxy group on the aromatic monomer ring and are more toxic than H monomers. Wild type *C. oleaginosus* was found to metabolize up to 0.6 g/L syringic acid as the sole carbon source in TOHN media (Figure 6.2). It also grew robustly in 0.4 g/L guaiacol, and eventually grew on 0.6 g/L after a 66 hour lag phase. Additionally, a buffer has been shown to be useful in increasing the amount of carbon that can be delivered to the cells in a given batch culture (Figure 6.3). This motivated the buffering system seen in Chapters 2 and 4.

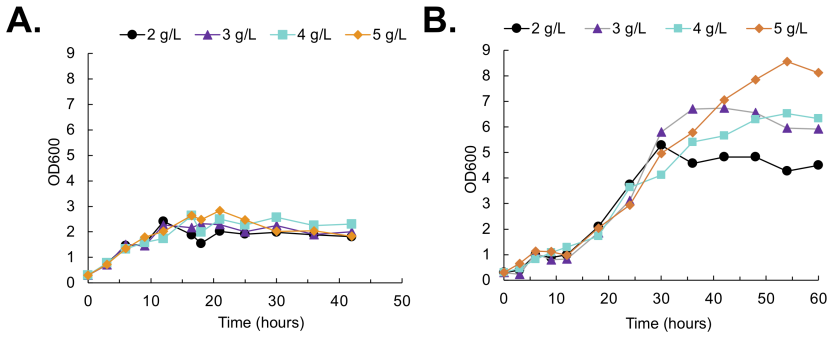


Figure 6.3. Growth of *C. oleaginosus* in TOHN media containing A. 0 mM phosphate buffer or B. 50 mM phosphate buffer while utilizing 2 g/L (black circle), 3 g/L (purple triangle), 4 g/L (cyan square), and 5 g/L (orange diamond) protocatechic acid as the sole carbon source. Data are for single replicates.

The evolved strain, mentioned in section 6.2.1.4 below, can likely tolerate even higher concentrations. Understanding the demethylation and funneling steps, in particular, behind G- and S-type monomers specifically will greatly enhance efforts to valorize of lignin derived from woody biomass.

6.2.1.3 Functional genomics

The discoveries in 4 help improve our understanding of the genetic basis for aromatic metabolism in yeast; yet, predictions can only be as robust as the data they are based from. In the case of aromatic catabolism in yeast, the foundational knowledge on which we base our genetic predictions has significant room for improvement. While there are biochemical assays confirming which reactions are taking place in yeast, the genes responsible are often understudied. Functional genomics bridges these two pieces of information. With the advent of CRISPR, most would think to use this high-throughput, highly tunable, targeted genome engineering technique for targeted knockouts coupled to loss of function screening. However, CRISPR requires design of a guide RNA for every target and suffers from off-target effects. Selection

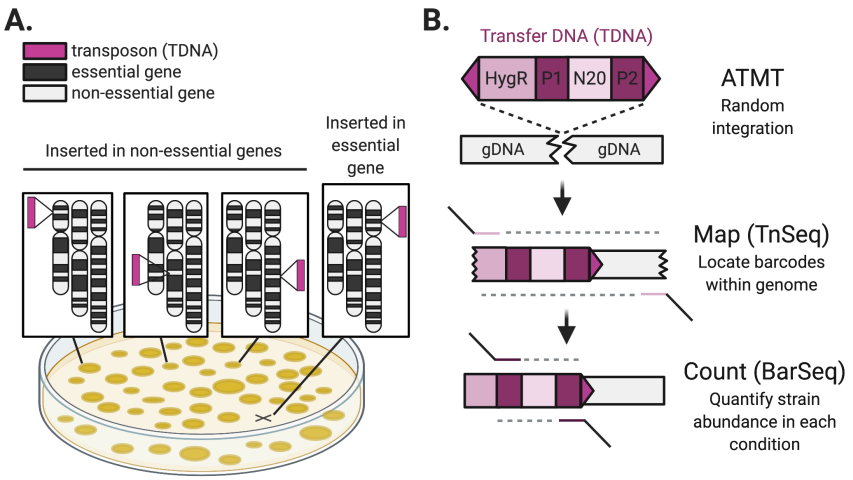


Figure 6.4. Schematic of the A. generation of mutant libraries and B. the randomly barcoded transposon DNA required for RB-Tnseq. Based off [2] and [4].

can also be a difficult phenotype to probe in a high-throughput manner.

Transposon-insertion sequencing methods combine large-scale transposon mutagenesis with next-generation sequencing to gauge essentiality and fitness contribution of genetic features in a genome. Transfer DNA (TDNA) are randomly integrated in the genome, but can be located to identify site of insertion (Tnseq). The scalability and precision of Tnseq is dramatically improved when the TDNA contains a short barcode sequence, a method called random barcode Tnseq (RB-Tnseq) (Figure 6.4). Not only can you see where a barcode is integrated, you can quantify its occurence and look at hundreds of conditions at a time. The randomness of integration also opens the solution space to include unexpected genes or genes of unknown function. This technique was used in *Rhodosporidum toruloides* to identify over 1,000 essential genes and 150 genes having significant roles in lipid accumulation [2]. RB-Tnseq was also used to identify catabolic pathways of lignin-derived aromatic compounds in *Novosphingobium aromaticivorans* [3]. The same method could be developed for *C. oleaginosus*.

6.2.1.4 Evolved strain

An evolved strain of *C. oleaginosus* was created through serial passaging on 1 g/L phenol in a semi-defined, minimal media. 28 passages was enough to generate an evolved strain that grows 52% faster on 1 g/L phenol and metabolizes 60% higher concentration of phenol (Appendix A). There are assuredly genomic and transcriptomic differences between the wild type and evolved strains. Common trends for similar adaptive evolution studies include gene duplication and up-regulation of transport and metabolic genes.

To study this, I recommend comparing wild-type grown with both 1 g/L phenol and 1 g/L glucose to the evolved strain with 1 g/L phenol, 1.4 g/L phenol, and 1 g/L glucose. The comparisons would provide more validation for phenol utilization pathways and potentially showcase tolerance mechanisms, such as expression of transporters, membrane altering proteins, DNA repair proteins, and redox proteins. The genome of the final evolved strain should be resequenced using PacBio coupled to Illumina sequencing to determine any SNPs or genomic rearrangements that may have occurred during evolution.

Preliminary studies on the evolved strain indicate adaptation on solely phenol has not resulted in loss of metabolism of other aromatic monomers (Figure 6.5). Experiments examining the full extent of catabolism of carbon sources such as phenol, resorcinol, pHBA, ferulic acid, *p*-coumaric acid, syringic acid, and vanillin would help identify if evolution has improved tolerance across multiple substrates. Future work could investigate enzyme promiscuity to see if aromatic catabolic enzymes have altered kinetic parameters while using different substrates.

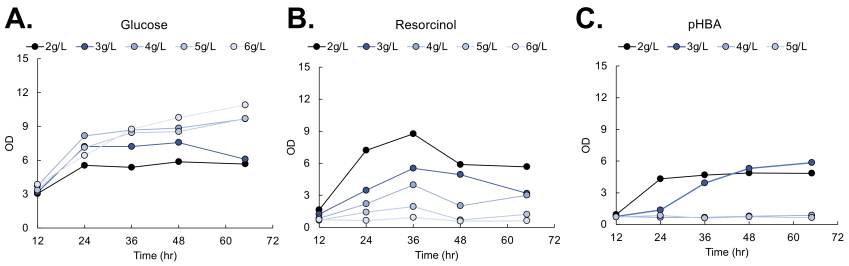


Figure 6.5. Growth of evolved *C. oleaginosus* cells in TOHN media containing A. glucose, B. resorcinol, or C. pHBA as the sole carbon source at 2, 3, 4, 6, and 6 g/L concentrations. Data are for single replicates.

6.2.2 Bioproduction of value-added products

6.2.2.1 Genetic tools

The limited genetic tools available for *C. oleaginosus* inhibits its application for bioproduction of value-added chemicals. Development of basic tools such as an efficient transformation method, stable episomal systems, and a range of weak to strong promoters, and additional terminators are key to facilitate incorporation of heterologous pathways. More advanced tools such as an efficient and targeted recombination tool, such as serine integrases and CRISPR, will allow host engineering for precise gene mutations and knock-outs. Not only would these allow rational host engineering, it would facilitate fundamental knowledge of aromatic catabolism through systematic knockout studies of catabolic genes to prove functionality proposed in 4. While stated simply, this will be an arduous undertaking.

6.2.2.2 Novel products to target

Heterologous pathways are essential to transform naturally-produced neutral lipids into industrially-relevant oleochemicals. There are a number of targets that are currently in the process of being genomically integrated, as mentioned in Chapter 5. Among the desired products are eicosapentaenoic acid (EPA), ricinoleic acid, and

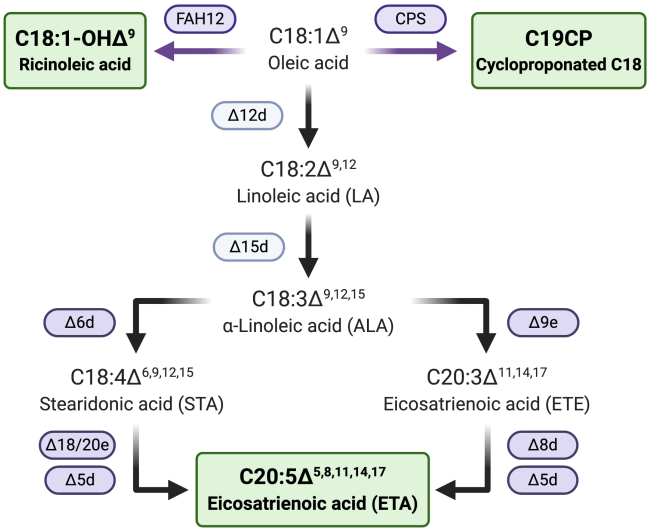


Figure 6.6. A proposed set of novel engineering products (green boxes) and the required metabolic engineering reactions (purple arrows) and enzymes (purple ovals) to convert the naturally occurring products, oleic acid and linoleic acid.

cyclopropane fatty acids (Figure 6.6). EPA precursor, eicosatrienoic acid (ETE), has been produced in *C. oleaginosus* [5].

In terms of economic considerations, a low-value product requiring a low-cost feedstock to economically make production feasible would be best if utilizing lignin. A main target in this case could be bulk chemicals, such as biodiesel or fatty alcohols. Other feasible products include any requiring significant amounts of acetyl-CoA or succinate as precursors (Figure 6.7). Fatty acids are a natural extension of this (biofuels, biodiesels, adipic acid, etc). Terpenes, specifically tanshinones, are another product that a junior graduate student in my group is looking into; however, I foresee significant competition from natural carotenoid producers, such as *Rhodosporidium toruloides*. Bioplastics such a polyhydroxyalkanoates (PHAs) and polyhydroxybutyrates (PHBs) are derived from the β-oxidation cycle, which breaks fatty acids down to acetyl-CoA. In these biopolymers, it may be possible to utilize lignin to introduce unique aromatic side chains to mimic polystyrene or PET. This would require PHA

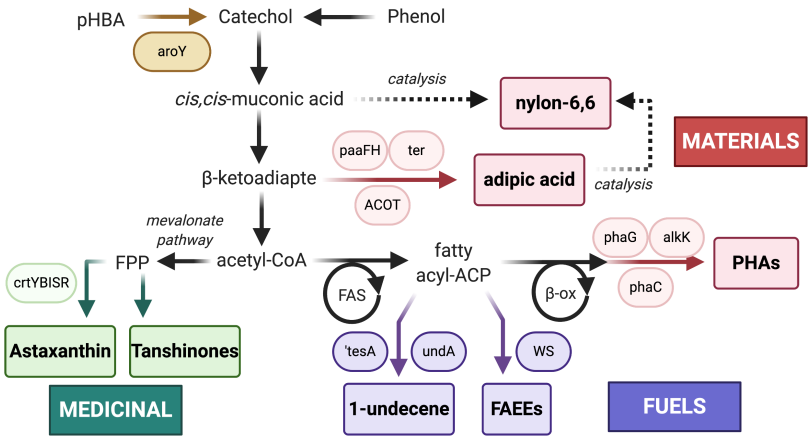


Figure 6.7. A proposed set of novel engineering products based on funneling intermediates to produce materials (red boxes), fuels (purple boxes), and medicinal (green boxes) products.

and PHB synthases that can accept aromatic substrates and for lignin metabolism to be shunted through knockouts of the ring-cleavage genes. Other avenues could capitalize on flux through the β-ketoadipate metabolism pathway to focus on bio-products like muconic acid and β-ketoadipate-CoA, both precursors for nylon. A more ambitious project could be engineering lipid droplets to be a pseudo-organelle for synthesis of hydrophobic products. Given the extremely high lipid accumulation of this yeast and propensity to store them as droplets, proteins could be scaffolded to a lipid bilayer and then metabolites transported into the lipid droplets.

Bibliography

[1] Gu, C.; Zheng, F.; Long, L., *et al.* Engineering the Expression and Characterization of Two Novel Laccase Isoenzymes from Coprinus comatus in Pichia pastoris by Fusing an Additional Ten Amino Acids Tag at N-Terminus. *Plos ONE*, **2014**, 9, e93912.

[2] Coradetti, S.; Pinel, D.; Geiselman, G., *et al.* Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides. *eLife*, **2018**, 7, e32110.

[3] Cecil, J.; Garcia, D.; Giannone, R., *et al.* Rapid, parallel identification of catabolism pathways of lignin-derived aromatic compounds in Novosphingobium aromaticivorans. *Appl Environ Microbiol*, **2018**, 84, e01185.

[4] Michel, A.; Hatakeyama, R.; Kimmig, P., *et al.* Functional mapping of yeast genomes by saturated transposition. *eLife*, **2017**, 6, e23570.

[5] Görner, C.; Redai, V.; Bracharz, F., *et al.* Genetic engineering and production of modified fatty acids by the non-conventional oleaginous yeast Trichosporon oleaginosus ATCC 20509. *Green Chem.*, **2016**, 18(7), 2037–2046.

Appendices

Appendix A Adaptive evolution for enhanced phenol metabolism

This appendix describes the work to create a strain of *Cutaneotrichosporon oleaginosus* that more robustly tolerates and metabolizes phenol.

A.1 Motivation

The wild-type strain metabolizes a maximum of 1.2 g/L phenol. Adaptive laboratory evolution (ALE) can enhance metabolism and/or tolerance to toxic compounds by selectively passaging cells with the desired phenotype. In this case, the desired phenotype is a decreased lag phase before cells are able to grow in phenol-containing media. While not specifically selected for, evolved strains were expected to tolerate higher concentrations of phenol due to their faster phenol metabolism.

A.2 Methods

A.2.1 Adaptive laboratory evolution

Cutaneotrichosporon oleaginosus ATCC 20509 cells were cultured in a 2 mL YPD pre-culture overnight. Cells were washed three times with sterile water and used to inoculate three 50 mL baffled flasks containing 15 mL TOHN with 1 g/L phenol as the sole carbon source to an OD of 0.3. Samples were monitored until they reached mid-exponential, which was previously established to be an OD of approximately 2.0 (refer to Chapter 3). The two samples that reached this OD first was harvested by centrifugation and resuspended in fresh TOHN media. Six fresh 50 mL baffled flasks containing 15 mL TOHN + 1 g/L phenol were prepared. Each harvested sample was used to inoculate three flasks to an OD of 0.3. Samples were again monitored

until mid-exponential phase. The two fastest growers were harvested as previously and used to inoculate three fresh flasks each. This was repeated and is schematically described in Figure A.1).

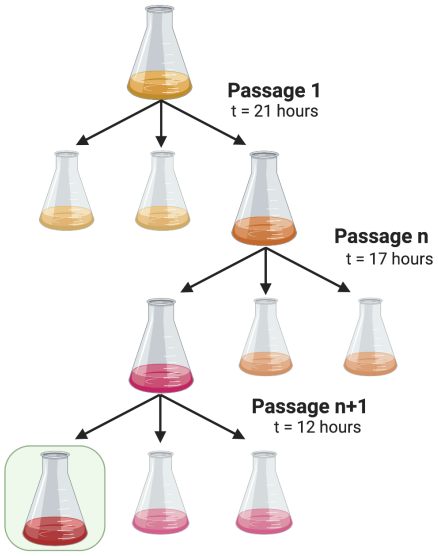


Figure A.1. Schematic describing the process for adaptively evolving *C. oleaginosus* for enhanced phenol metabolism.

Cells with a marked improved phenotype were glycerol stocked. An aliquot of cells were also saved for downstream RNA extraction. Stocked strains were grown in TOHN + 1 g/L phenol to ensure strain stability and growth consistency (data not shown).

A.3 Results

As each sample was passaged, the time it took to reach mid-exponential, a proxy for lag phase, was recorded. A plot of each generations time to mid-exponential versus generation number shows the evolution trend of decreasing time (Fig. A.2). In the plot, each dot represents a generation. The dots that are colored green indicate

which samples have RNA samples saved. More generations are saved as glycerol stocks. There is a distinct point of diminishing returns seen around generation 20.

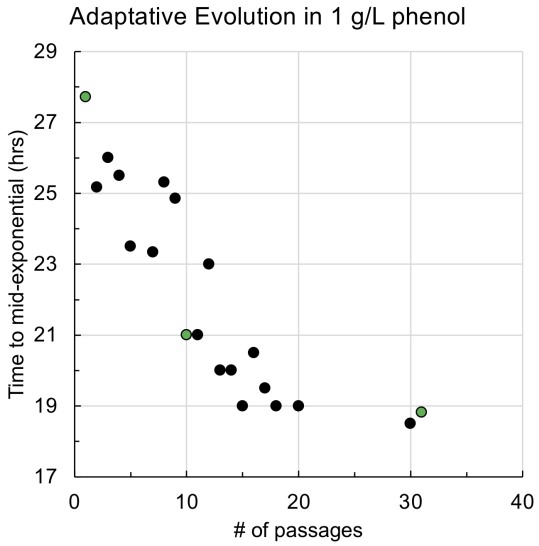


Figure A.2. Demonstration of how time to reach mid-exponential phase decreased with every generation produced from ALE.

At the end of the ALE, the evolved strain showed a decrease in lag phase when cells were cultured in 1 g/L phenol in TOHN media (Figure A.3A). The growth rate for wild-type cells grown in phenol is 0.044 hr^{-1}, in pHBA is 0.094 hr^{-1}, and 0.156 hr^{-1} for cells grown in glucose. The evolved strain showed a marked improved growth rate of 0.067 hr^{-1}. This is a growth rate improvement of 52.3%. The evolved strain is also able to tolerate more phenol as compared to the wild-type cells.

The evolved strain is able to tolerate up to 1.6 g/L phenol (Fig. A.3B), whereas the wild-type strain is only able to tolerate up to 1.2 g/L phenol (Fig. A.3C). This results in an improvement in phenol tolerance by 33.3%. Wild-type cells suffer a decreased growth rate when grown in 1.2 g/L phenol as compared to the 1 g/L phenol, but the evolved cells do not grow any differently.

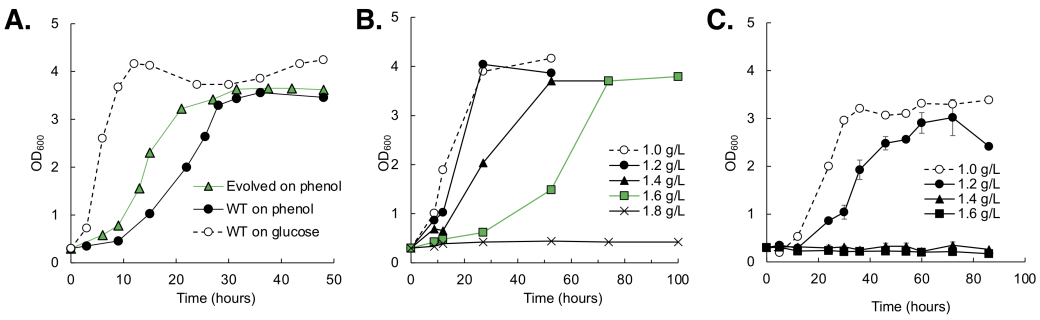


Figure A.3. Comparison of A. the wild-type and evolved strain, B. evolved strain growth in increasing phenol concentrations, and C. wild-type growth in increasing phenol concentrations.

The RNA samples harvested for all the samples in this project were from cells grown in TOCM media. Since the ALE was performed in TOHN, I checked how different the cell growth was when grown in TOCM media (Fig. A.4). Cells had an enhanced growth rate, grew to a slightly higher OD, and had a slightly longer lag phase, all probably because of the higher concentration of yeast extract. Lipid production was not tested. This data may be helpful for future studies involving RNAseq.

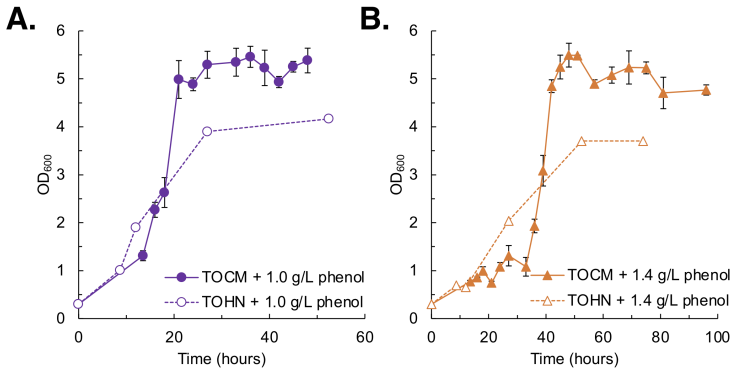


Figure A.4. Comparison of the evolved strain grown in TOHN and TOCM in A. 1 g/L phenol and B. 1.4 g/L phenol.

A.4 Future work

In Fig A.2, the point of diminishing returns is due to a lack of driving force for evolution. The evolutionary pressure needed to be made more stringent and I think the easiest one to pursue would be to increase the concentration of phenol. Because the evolved cells grow no differently between 1 g/L and 1.2 g/L phenol, if one wanted to continue improving the growth rates and tolerance thresholds, I suggest starting a new ALE experiment using 1.4 g/L phenol.

The cell pellets for RNA extraction were harvested during ALE, but were not stored with RNA protect. I would suggest regrowing the isolated generation to extract fresh RNA. I think an RNAseq comparison between the evolved strain and the wild-type would be the best use for the evolved strain. I would compare wild-type with 1 g/L phenol and 1 g/L glucose to the evolved strain with 1 g/L phenol, 1.4 g/L phenol, and 1 g/L glucose. The comparisons would provide more validation for phenol utilization pathways and potentially showcase tolerance mechanisms. The genome of the final evolved strain should be resequenced using PacBio coupled to Illumina sequencing. This will determine any SNPs or genomic rearrangements that may have occurred during evolution.

Similarly, to probe mutations that occur due to the nature and stochasticity of evolution, wild type cells could be adaptively evolved in 1 g/L glucose for the same number of generations as the C31 strain. The genome could be resequenced to rule out overlapping SNPs or trends in genome rearrangement that may not be due to phenol as the driving force, but could instead be artifacts of the evolution process.

Appendix B Metabolism of alternative sugars at high density

This appendix describes our characterization of sweet sorghum syrup, mock and real cellulosic biomass hydrolysates, and high density single and dual sugar feedstocks. This work was funded by the U.S. Department of Agriculture (USDA) Sun Grant (2014-38502-22598) from 2016-2019.

B.1 High gravity sugar fermentation

Cutaneotrichosporon oleaginosus is well known for its alternative sugars metabolism 1. In this section, I describe the work involving glucose, xylose, and mixed feedstocks for high gravity fermentation. High gravity fermentations are desirable for their beneficial processing parameters, such as lowered requirements for process water and energy. High osmotic stress associated with high sugar density can additionally address contamination issues.

High gravity cultures were based in a high nitrogen YSC media buffered with a 50 mM phosphate buffer. Both glucose and xylose were tested as individual carbon sources in 20, 60, 100, 150, and 200 g/L concentrations (B.1 and B.2). Growth curve (closed marker, solid line) was measured at A600 using a Nanodrop 2000. Substrate utilization (open mark, dashed line) was measured using a Waters 600 system with an Aminex HPX 87C column, a mobile phase of ddH_2O, temperature of 85 °C, flow rate of 0.6 mL/min, and a Waters 2400 RI detector.

It appears both glucose and xylose are fully consumed by the end of the culturing period (120 hours) for 20, 60, and 100 g/L. However, 150 g/L and 200 g/L for both glucose and xylose are too high of a concentration to be fully metabolized by

the end of the culturing period, but I think they would be with enough time. For the glucose cultures, all concentrations higher than 20 g/L resulted in oleaginous behavior of approximately 40% (B.4 A). The xylose showed less oleaginous behavior overall, with the worst oleaginous behavior coming from the culture containing 200 g/L (B.4 B). The cultures containing 100 g/L glucose seemed to have the peak biomass titer, lipid titer, and lipid accumulation. In contrast, the xylose cultures continued to having increasing amounts of biomass and a plateaued lipid production with increasing xylose concentration.

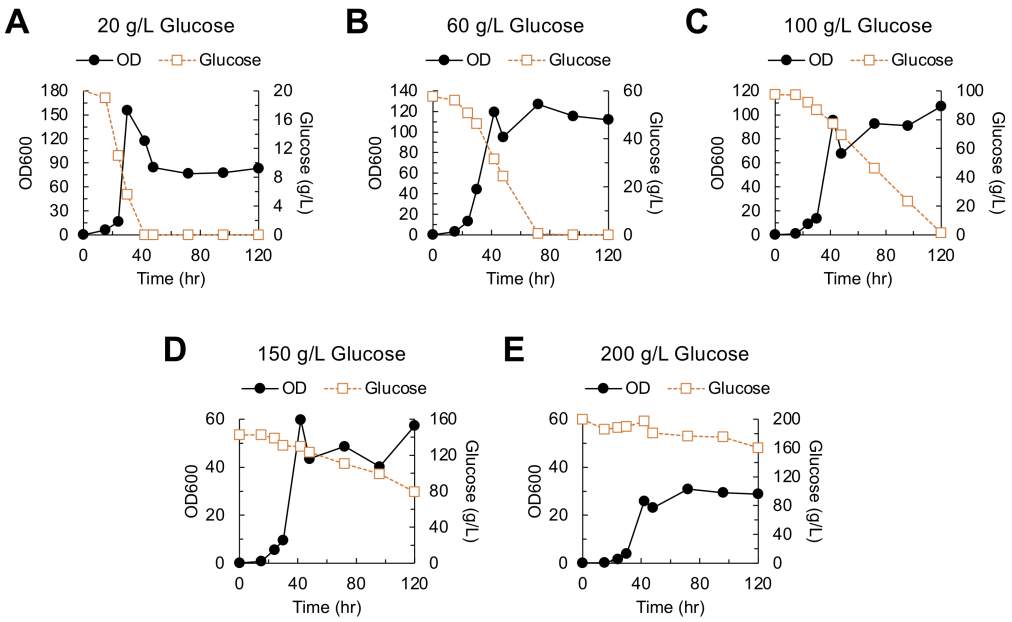


Figure B.1. High gravity glucose fermentation growth and substrate utilization profiles.

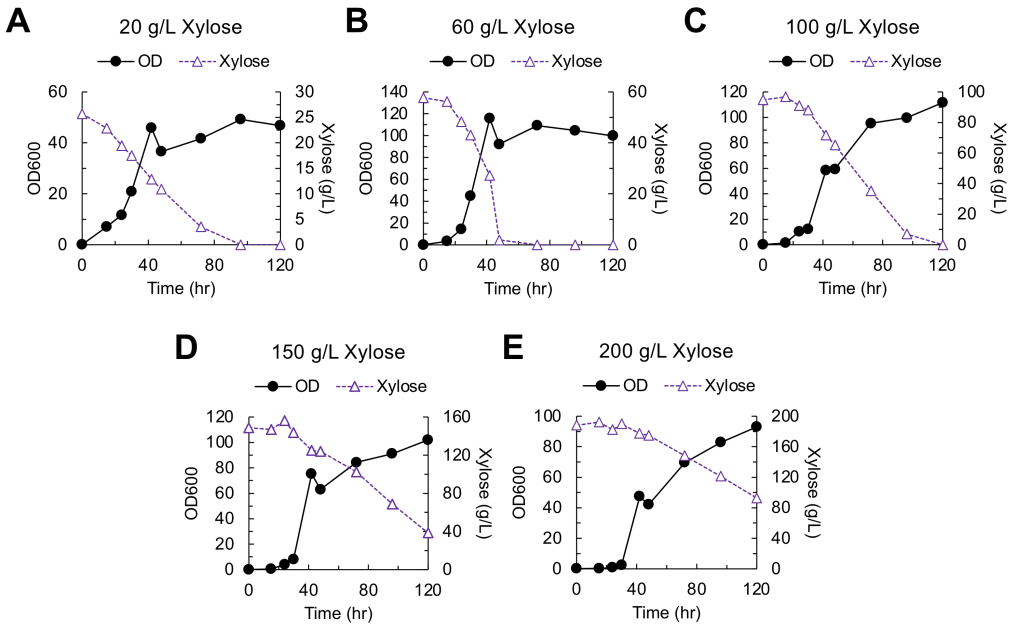


Figure B.2. High gravity xylose fermentation growth and substrate utilization profiles.

Individual sugars seem to be limited to 120 hours for full uptake. It was not known if blending the sugars as a dual carbon feedstock would require the total sugar concentration to be under 150 g/L or if each sugar had to be 150 g/L individually. It was also unknown how the ratio of glucose to xylose would affect uptake. Thus, I tested a total sugar concentration of 100 g/L (B.3 A-C) and 200 g/L (B.3 D-F) at ratios of 1:1, 2:1, and 1:2.

It seemed that a total sugar concentration of 200 g/L of sugars prevented full metabolism of the sugars by 120 hours, and it does seem that uptake rate is negatively affected at such a high concentration. For the cultures containing 100 g/L total sugar concentrations, glucose was depleted before xylose for all glucose to xylose ratios. In all cases, xylose was utilized simultaneously as glucose, though it does not become utilized at a significant amount until the glucose depleted by approximately 50% of the starting glucose concentration. In both the 100 g/L and 200 g/L total sugar concentration cultures, the 1:1 glucose to xylose ratio produced an optimal lipid

accumulation percent (B.4 C and D).

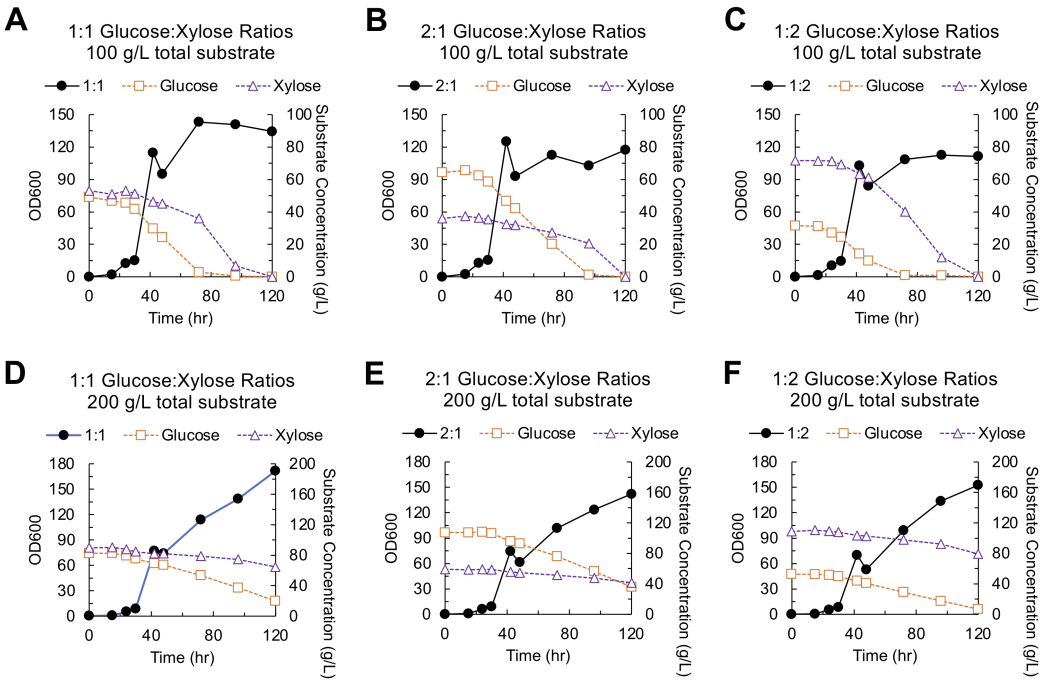


Figure B.3. High gravity glucose:xylose blend fermentation growth and substrate utilization profiles.

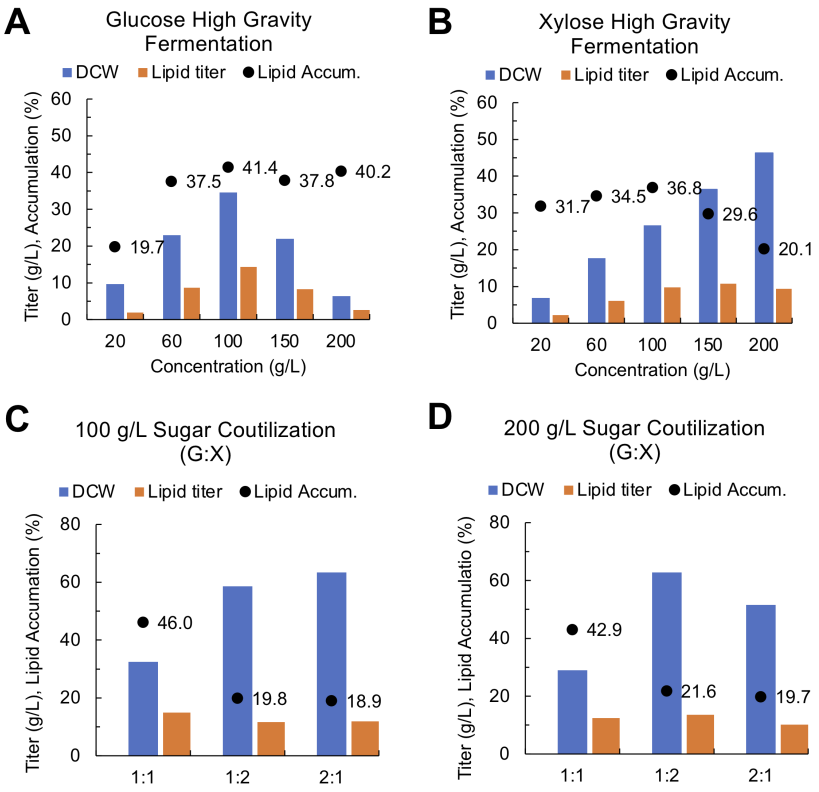


Figure B.4. High gravity fermentation biomass titer, lipid titer, and lipid accumulation.

B.2 Sorghum syrup

Pure Missouri Ozark Country sweet sorghum syrup was purchased from Amazon and used as the sorghum syrup source in the experiments in this section. In all cases, the syrup was mixed into the media and then the final media was sterilized by syringe filtering.

We have characterized *C. oleaginosus* growth and lipogenesis using pure sweet sorghum syrup as sole carbon source with various nitrogen concentrations. Positive growth was observed using sorghum syrup as a sole carbon source (B.5D), synthetic sorghum syrup as sole carbon source (B.5C), and individual single sugars (B.5A & B).

Cells were compared for their growth and lipid production in high nitrogen (C:N = 4.9) and low nitrogen (C:N = 241) conditions. As expected, cells exhibit oleaginous behavior in low nitrogen media when grown on sorghum syrup (B.1 and B.6).

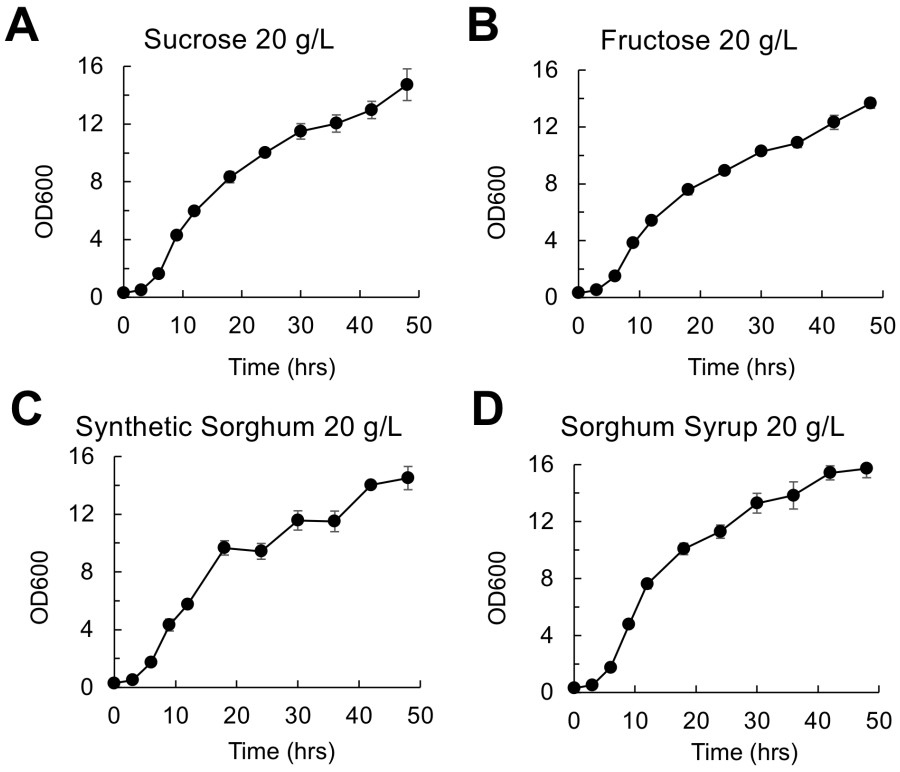


Figure B.5. Growth profiles and substrate utilization derive)d from sorghum syrup feedstock.

Table B.1. Measurement of dry cell weight, lipid titer, and percentage lipid accumulation from sorghum syrup cultures.

Substrate	**Lipid Titer(g/L)**	**Biomass Titer (g/L)**	**Lipid Accumulation (%)**
Sorghum (HN)	0.94±0.06	5.40±0.11	17.43±1.25
Sorghum (LN)	4.60±0.51	7.04±0.16	65.30±6.78

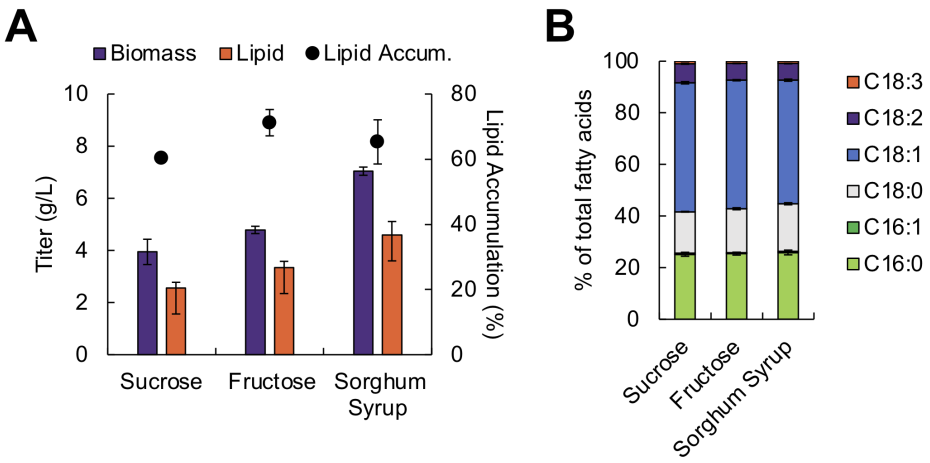


Figure B.6. Growth profiles and substrate utilization derived from sorghum syrup feedstock.

To investigate the possibility of using whole lignocellulosic biomass hydrolysate as a feedstock, we spiked the sweet sorghum syrup with a model aromatic compound, resorcinol (B.7). Cells lose oleaginicity when grown with both sorghum syrup and resorcinol, a monoaromatic compound (B.2). The lipid titers and biomass titers are significantly less than using the sorghum syrup alone, so the resorcinol may have been too toxic to process alongside the sugars. The cells may also be suffering from catabolite repression.

Table B.2. Measurement of dry cell weight, lipid titer, and percentage lipid accumulation from sorghum syrup cultures.

Substrate	**Lipid Titer(g/L)**	**Biomass Titer (g/L)**	**Lipid Accumulation (%)**
Sorghum (HN)	0.41±0.02	3.89±0.01	9.99±0.77
Sorghum (LN)	0.75±0.04	2.62±0.23	28.78±2.79

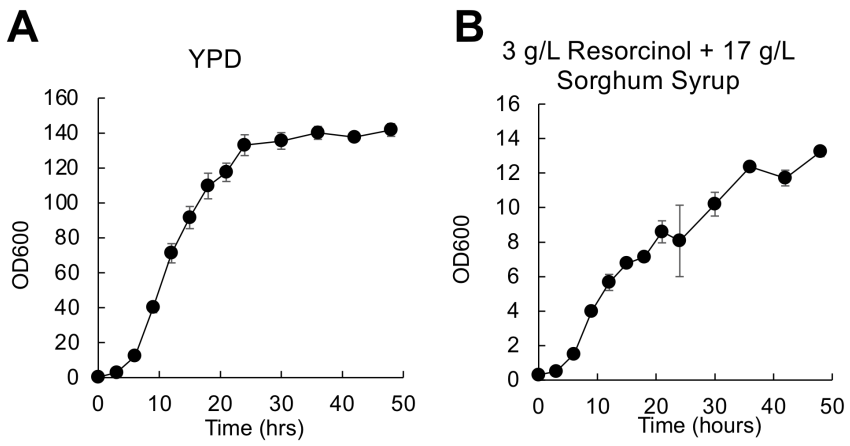


Figure B.7. Growth profile of sorghum syrup mixed with resorcinol.

To probe whether this phenomenon is due to inherent toxicity of the aromatic compound or catabolite repression, we explored two-phase culturing. Surprisingly, cells do not regain oleaginous characteristics when they are grown in two stages (B.3). The first stage media is in YPD to generate high biomass, and the second stage media contains resorcinol (B.8).

Table B.3. Measurement of dry cell weight, lipid titer, and percentage lipid accumulation from sorghum syrup cultures.

Substrate	**Lipid Titer(g/L)**	**Biomass Titer (g/L)**	**Lipid Accumulation (%)**
3 g/L	0.62±0.18	5.02±0.35	16.48±6.45
6 g/L	0.64±0.20	3.51±0.39	18.91±3.31
12 g/L	0.60±0.12	3.26±0.20	18.66±4.32
20 g/L	0.57±0.03	4.27±0.15	13.42±0.47

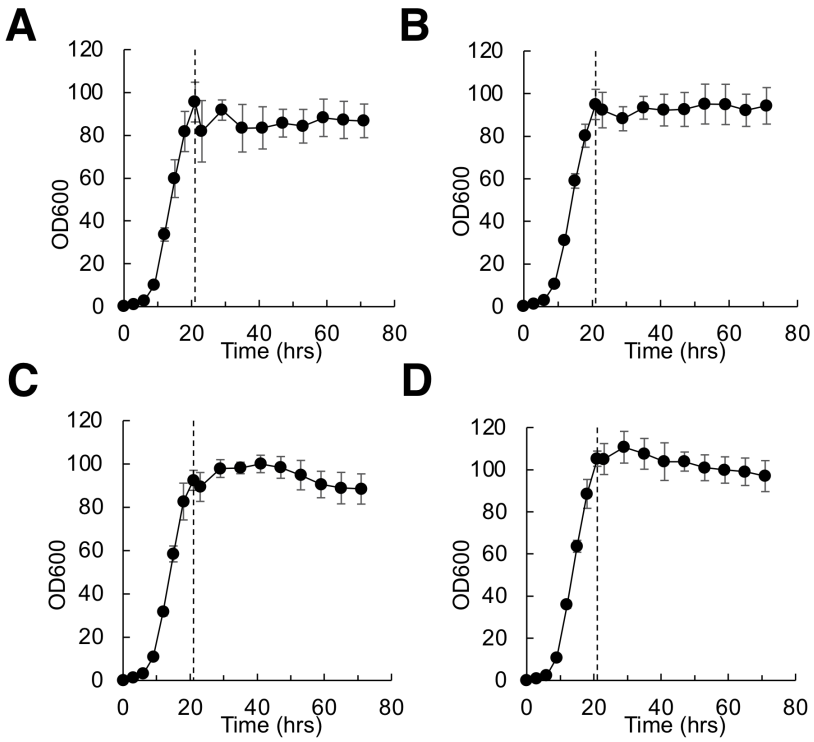


Figure B.8. Growth profile of two-stage YPD:resorcinol culturing.

B.3 Eucalyptus and switchgrass cellulosic hydrolysate

We received a cellulosic hydrolysate derived from a blend of eucalyptus and switchgrass from John Gladden at the Lawrence Livermore National Lab. Its ABPDU

batch record number is No. 110205R1. The composition, listed as of September 14, 2012, was 55.95 g/L glucose, 19.30 g/L xylose, and 3.76 g/L cellobiose. The hydrolysate was generated using C1C2imOAc, an ionic liquid, which was washed off the biomass prior to hydrolysis (personal communication). The hydrolysate was supplemented with a YSC media base. A synthetic version of the media, based on the reported composition, was blended for use as a point of comparison for the hydrolysate. Sugars were detected using HPLC-RID using the method reported in the high gravity section. Cells produced higher lipid accumulation values with the hydrolysate as compared to the synthetic media (B.10). This indicates there is likely something else in the media contributing to lipid accumulation.

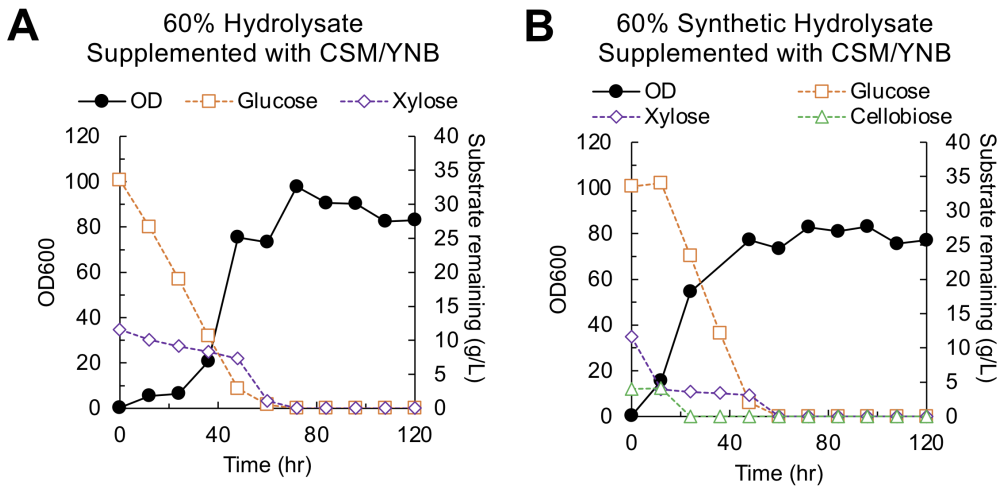


Figure B.9. Growth profiles and substrate utilization derived from eucalyptus/switchgrass cellulosic hydrolysate feedstock.

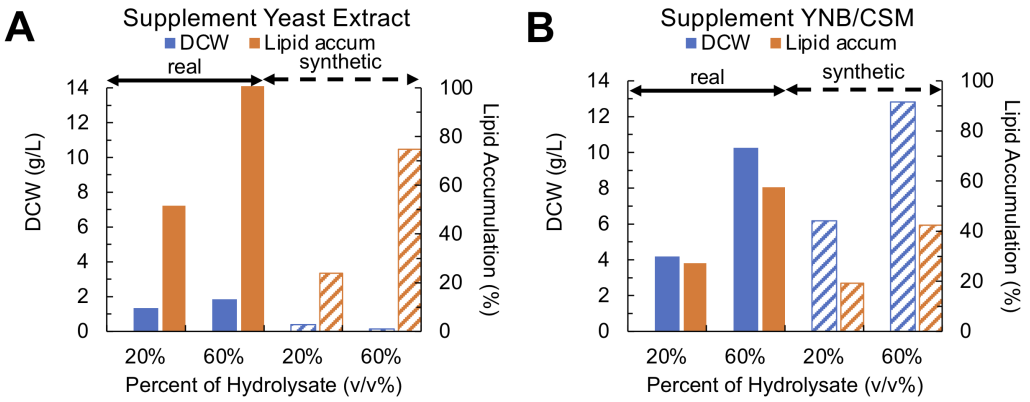


Figure B.10. Biomass titer, lipid titer, lipid accumulation, and lipid profile derived from eucalyptus/switchgrass cellulosic hydrolysate feedstock.

B.4 Future work

Further work is required to understand why there is such variation between the high density fermentation studies. Analysis probing transporter activity and targeted qPCR of cells grown with xylose and glucose blends may elucidate feedback inhibition and other regulatory effects of having a pentose and a hexose sugar. Additional studies with the hydrolysate to examine common hydrolysate by-products may be useful to identify what may be contributing to the enhanced lipid accumulation seen in the hydrolysate as compared to the synthetic hydrolysate media.

Bibliography

[1] Yaguchi, A; Rives, D; Blenner MA. The New Kids ont he Block: Emerging Oleaginous Yeast of Biotechnological Importance. *AIMS Microbiology*. **2017**, **3**:2, 227-247.

Appendix C In-house python scripts for transcriptomic work for *C. oleaginosus*

This appendix describes the Jupyter Notebooks developed for Python. These scripts normalize gene counts and visualize outputs. Soft copy Notebooks are provided to the Blenner Lab and have more thorough annotations.

C.1 Normalization of counts using upper quartile normalization and heirarchical clustering of normalized counts

C.1.1 Normalization

```
# import packages
import pandas as pd
import numpy as np
import scipy
import matplotlib.pyplot as plt
import seaborn as sns

# import files and format
Counts = pd.read_table("counts_new_2.txt",
  delim_whitespace=True, low_memory=False, index_col=0)
Counts = Counts.drop(Counts.columns[[0, 1, 2, 3, 4]], axis=1)
Counts.rename(columns={"9-MAB.sorted.bam":"Phenol1",
                       "10-MAB.sorted.bam":"Phenol2",
                       "11-MAB.sorted.bam":"pHBA1",
                       "12-MAB.sorted.bam":"pHBA2",
                       "13-MAB.sorted.bam":"Resorcinol1",
                       "14-MAB.sorted.bam":"Resorcinol2",
                       "15-MAB.sorted.bam":"Glucose1",
                       "16-MAB.sorted.bam":"Glucose2",},
                       inplace=True)

# calculate upper quartile
rankPhe1 = Counts.sort_values(by='Phenol1', ascending=False)
phe1Q3 = np.percentile(rankPhe1.Phenol1, 75)

rankPhe2 = Counts.sort_values(by='Phenol2', ascending=False)
phe2Q3 = np.percentile(rankPhe2.Phenol2, 75)
```

```
### repeat above for all samples and duplicates ###

# normalize ranked data by their 75th percentile and drop non-
  normalized data
Counts_Norm = Counts
Counts_Norm['Phe1Norm'] = (Counts.Phenol1 / phe1Q3)
Counts_Norm.drop(labels='Phenol1', axis=1, inplace=True)

### repeat for all samples and duplicates ###

Counts_Norm.to_csv("Counts_quartnorm.csv")
```

C.1.2 Heirarchical clustering

This code is a continuation of the above, but is broken into two sections for readability.

```
# eliminate rows comprised of zeroes and calculate variance
Counts_norm_trim = Counts_Norm.loc[~(Counts_Norm==0).all(axis=1)]
Counts_var = Counts_norm_trim.var(axis=1)
Counts_norm_trim['Counts_var'] = Counts_var


# keep top 200 genes with highest variance
Counts_top200 = Counts_norm_trim.nlargest(200, 'Counts_var')
Counts_top200 = Counts_top200.drop(labels='Counts_var', axis=1)


# plot
sns.set_style('ticks', rc={"xtick.major.size":2,"ytick.major.size":2
                                    , "axes.labelsize":'large'})
Counts_cg = sns.clustermap(Counts_top200, method='ward', metric='
                                    euclidean', robust=True, z_score=0
                                    , figsize=(15,50))
```

C.2 Principal component analysis (PCA) of normalized counts

```
# import packages
import pandas as pd
import numpy as np
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
%matplotlib inline

# import CSV output from upper quartile normalization and format
counts = pd.read_csv("Counts_quartnorm.csv", header=0, index_col=0)
counts_T = counts.T
substrate = pd.read_excel('181213_SampleType_Quart.xlsx', header=0,
                                    index_col=0)
sample = substrate.index
counts_T['substrate'] = substrate

# perform PCA
x = counts_T.loc[:,counts_T.columns !='substrate'].values
y = counts_T.loc[:,['substrate']].values
pca = PCA(n_components=2)
principalComponents = pca.fit_transform(x)
principalDf = pd.DataFrame(data = principalComponents, columns = ['
                                    PC 1', 'PC 2'], index=sample)
finalDf = pd.concat([principalDf, counts_T[['substrate']]], axis = 1
                                    )

# print degree to which each PC explains the variance
print(pca.explained_variance_ratio_)

# visualize scores on the PCAs
fig = plt.figure(figsize = (6,6))
ax = fig.add_subplot(1,1,1)
ax.set_xlabel('PC 1', fontsize = 15)
ax.set_ylabel('PC 2', fontsize = 15)
ax.set_title('PCA: Upper Quartile Normalized Counts', fontsize = 20)
targets = ['Phenol', 'pHBA', 'Resorcinol', 'Glucose']
colors = ['g', 'r', 'b', 'k']
markers = ['s', 'v', 'D', 'o']
for target, color, marker in zip(targets,colors,markers):
    indicesToKeep = finalDf['substrate'] == target
    ax.scatter(finalDf.loc[indicesToKeep, 'PC 1']
               , finalDf.loc[indicesToKeep, 'PC 2']
               , c = color
               , marker = marker
               , s = 40)
ax.legend(targets, frameon=True)
ax.grid(False)
ax.tick_params(direction='out', length=5)
```

```
fig.savefig("PCA_Quartile.png")

# evaluate loadings
loadings = pca.components_
loadingsDf = pd.DataFrame(loadings, index=['PC1','PC2'], columns=
                                    counts.index)
loadingsDf = loadingsDf.T
loadingsDf_maxPC1 = loadingsDf.nlargest(50, columns='PC1')
loadingsDf_maxPC1.drop('PC2', axis=1)
loadingsDf_maxPC2 = loadingsDf.nlargest(50, columns='PC2')
loadingsDf_maxPC2.drop('PC1', axis=1)

# visualize the loadings on PCs
fig = plt.figure(figsize = (7,7))
plt.scatter(x=loadingsDf['PC1'], y=loadingsDf['PC2'] ,alpha=0.5);
plt.title("Loading plot", fontsize=20);
plt.xlabel("Loadings on PC1", fontsize=15);
plt.ylabel("Loadings on PC2", fontsize=15);
plt.grid(False);
plt.tick_params(direction='out', length=5)
fig.savefig("Loading_Quartile.png")
```

C.3 Volcano plots

```
# import packages
import pandas as pd
import numpy as np
import scipy
import scipy.stats as stats
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.cluster.hierarchy import dendrogram, linkage
from bioinfokit import analys, visuz
import bioinfokit
%matplotlib inline

# import data
phe = pd.read_csv("all_pheVglu.csv", header=0, index_col=0)
phba = pd.read_csv("all_phbaVglu.csv", header=0, index_col=0)
res = pd.read_csv("all_resVglu.csv", header=0, index_col=0)

# plot data and export as PNG
bioinfokit.visuz.gene_exp.volcano(df=phe, lfc='logFC', pv='PValue',
                                    show=True)
bioinfokit.visuz.gene_exp.volcano(df=phe, lfc='logFC', pv='PValue',
                                    r=300, figname='volcano_phe_rna')
```

```
bioinfokit.visuz.gene_exp.volcano(df=phba, lfc='logFC', pv='PValue',
                                  show=True)
bioinfokit.visuz.gene_exp.volcano(df=phba, lfc='logFC', pv='PValue',
                                  r=300, figname='volcano_phba_rna'
                                 )

bioinfokit.visuz.gene_exp.volcano(df=res, lfc='logFC', pv='PValue',
                                  show=True)
bioinfokit.visuz.gene_exp.volcano(df=res, lfc='logFC', pv='PValue',
                                  r=300, figname='volcano_res_rna')
```

Appendix D In-house python scripts for proteomic work for *C. oleaginosus*

This appendix describes the Jupyter Notebooks developed for Python. These scripts analyze proteomics data provided by MUAL at Clemson. Soft copy Notebooks are provided to the Blenner Lab and have more thorough annotations.

D.1 Normalization of abundances using upper quartile normalization and heirarchical clustering of normalized abundances

D.1.1 Normalization

```
# import packages
import pandas as pd
import numpy as np
import scipy
import matplotlib.pyplot as plt
import seaborn as sns

# import files and format
NormAbun = pd.read_excel("200904_AbundancesNorm_ALL_forPython.xlsx",
                                    index_col=0)
NormAbun.fillna(0, inplace=True)

# plot non-normalized data and drop outliers
NormAbun_trim = NormAbun.loc[~(NormAbun==0).all(axis=1)]
cmap = sns.color_palette("RdBu",1000)
normabun_trim_plt = sns.clustermap(NormAbun_trim, method='ward',
                                   metric='euclidean', robust=True,
                                   z_score=0, cmap=cmap)

normabun_drop = NormAbun.drop(labels=['pCoum1','Ferul1','Glu3'],
                                   axis=1)

# calculate upper quartile
rankPhe1 = normabun_drop.sort_values(by='Phe1', ascending=False)
phe1Q3 = np.percentile(rankPhe1.Phe1, 75)
```

```
rankPhe2 = normabun_drop.sort_values(by='Phe2', ascending=False)
phe2Q3 = np.percentile(rankPhe2.Phe2, 75)

rankPhe3 = normabun_drop.sort_values(by='Phe3', ascending=False)
phe3Q3 = np.percentile(rankPhe3.Phe3, 75)

rankPhe4 = normabun_drop.sort_values(by='Phe4', ascending=False)
phe4Q3 = np.percentile(rankPhe4.Phe4, 75)

### repeat above for all samples - neglect dropped outliers ###

# normalize ranked data by their 75th percentile and drop non-
                                    normalized data

quartnorm = normabun_drop

quartnorm['Phe1Norm'] = (normabun_drop.Phe1 / phe1Q3)
quartnorm.drop(labels='Phe1', axis=1, inplace=True)

quartnorm['Phe2Norm'] = (normabun_drop.Phe2 / phe2Q3)
quartnorm.drop(labels='Phe2', axis=1, inplace=True)

quartnorm['Phe3Norm'] = (normabun_drop.Phe3 / phe3Q3)
quartnorm.drop(labels='Phe3', axis=1, inplace=True)

quartnorm['Phe4Norm'] = (normabun_drop.Phe4 / phe4Q3)
quartnorm.drop(labels='Phe4', axis=1, inplace=True)

### repeat for all samples and replicates ###

# format column names and save as csv

quartnorm1 = quartnorm.rename(columns={
    'Phe1Norm':'Phe1', 'Phe2Norm':'Phe2',
    'Phe3Norm':'Phe3', 'Phe4Norm':'Phe4',
    'Res1Norm':'Res1', 'Res2Norm':'Res2',
    'Res3Norm':'Res3', 'Res4Norm':'Res4',
    'pHBA1Norm':'pHBA1', 'pHBA2Norm':'pHBA2',
    'pHBA3Norm':'pHBA3', 'pHBA4Norm':'pHBA4',
    'pCoum2Norm':'pCoum2', 'pCoum3Norm':'pCoum3',
    'pCoum4Norm':'pCoum4',
    'Ferul2Norm':'Ferul2', 'Ferul3Norm':'Ferul3',
    'Ferul4Norm':'Ferul4',
    'Glu1Norm':'Glu1', 'Glu2Norm':'Glu2', 'Glu4Norm':'Glu4',
    'Lig1Norm':'Lig1', 'Lig2Norm':'Lig2', 'Lig3Norm':'Lig3'})

quartnorm1.to_csv("AbundancesAdjusted_QuartNorm_copy5.csv")
```

D.1.2 Heirarchical clustering

This code is a continuation of the above, but is broken into two sections for readability.

```
# eliminate rows comprised of zeroes
quartnorm1_trim = quartnorm1.loc[~(quartnorm1==0).all(axis=1)]


# normalize by z-score and plot
cmap = sns.color_palette("RdBu",1000)
quartnorm1_plt = sns.clustermap(quartnorm1_trim, method='ward',
                                metric='euclidean', robust=True,
                                z_score=0, cmap=cmap)
```

D.2 Principal component analysis (PCA) of normalized abundances

```
# import packages
import pandas as pd
import numpy as np
from sklearn.decomposition import PCA
import matplotlib.pyplot as plt
%matplotlib inline

# import CSV output from upper quartile normalization and format
abun = pd.read_csv("AbundancesAdjusted_QuartNorm_copy5.csv", header=
                                0, index_col=0)
substrate = pd.read_excel('200902_SampleType.xlsx', header=0,
                                index_col=0)
substrate = substrate.drop(labels=['pCoum1','Ferul1','Glu3'])
sample = substrate.index
abun_T = abun.T
abun_T['substrate'] = substrate

# perform PCA
x = counts_T.loc[:,counts_T.columns !='substrate'].values
y = counts_T.loc[:,['substrate']].values
pca = PCA(n_components=2)
principalComponents = pca.fit_transform(x)
principalDf = pd.DataFrame(data = principalComponents, columns = ['
                                PC 1', 'PC 2'], index=sample)
```

```
finalDf = pd.concat([principalDf, counts_T[['substrate']]], axis = 1
                                    )

# print degree to which each PC explains the variance
print(pca.explained_variance_ratio_)

# visualize scores on the PCAs
fig = plt.figure(figsize = (6,6))
ax = fig.add_subplot(1,1,1)
ax.set_xlabel('PC 1', fontsize = 15)
ax.set_ylabel('PC 2', fontsize = 15)
ax.set_title('PCA: Upper Quartile Normalized Abundances',
      fontsize = 20)
targets = ['Phenol', 'pHBA', 'Resorcinol', 'Glucose',
      'p-Coumarate', 'Ferulate','Lignin']
colors = ['g', 'r', 'b', 'k', 'blueviolet', 'orange',
      'lightseagreen']
markers = ['s', 'v', 'D', 'o', 'p', '*', 'P']
for target, color, marker in zip(targets,colors,markers):
    indicesToKeep = finalDf['substrate'] == target
    ax.scatter(finalDf.loc[indicesToKeep, 'PC 1']
               , finalDf.loc[indicesToKeep, 'PC 2']
               , c = color
               , marker = marker
               , s = 40)
ax.legend(targets, frameon=True, loc='lower right')
ax.grid(False)
ax.tick_params(direction='out', length=5)

fig.savefig('PCA_proteome_copy4.png', dpi=300, bbox_inches= 'tight')
```

D.3 Heirarchical clustering of logFC values

```
# import packages
import pandas as pd
import numpy as np
import scipy
import scipy.stats as stats
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.cluster.hierarchy import dendrogram, linkage
%matplotlib inline

# import data

padj = pd.read_excel("200904_Padj_ALL.xlsx", header=0, index_col=0)
```

```
logfc = pd.read_excel("200904_LogFC_ALL.xlsx", header=0, index_col=0
                                    )

# create substrate specific files of logFC values and isolate
                                    statistically significant values (
                                    a = 0.05)

phe = pd.DataFrame(index=logfc.index)
phe['logFC'] = logfc['phe_logfc'].values
phe['padj'] = padj['phe_padj'].values
phe_sig = phe[phe['padj'] <= 0.05]

### repeat with all substrates then combine into single dataframe
                                    ###

siglogfc = pd.concat([phe_sig['logFC'],
                      res_sig['logFC'],
                      phba_sig['logFC'],
                      pcoum_sig['logFC'],
                      ferul_sig['logFC'],
                      lig_sig['logFC']],
                     axis = 1)

siglogfc1 = np.nan_to_num(siglogfc)
siglogfc2 = pd.DataFrame(siglogfc1, index=siglogfc.index)

# plot

cmap = sns.color_palette("RdBu",1000)
logfc_plt1 = sns.clustermap(siglogfc2, method='ward', metric='
                                    euclidean', z_score=1, robust=True
                                    , cmap=cmap)

logfc_plt1.savefig('HC_proteom_logFC.png', dpi=300, bbox_inches='
                                    tight')
```

D.4 Volcano plots of each dataset

```
# import packages
import pandas as pd
import numpy as np
import scipy
import scipy.stats as stats
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.cluster.hierarchy import dendrogram, linkage
```

```
from bioinfokit import analys, visuz
import bioinfokit
%matplotlib inline

# import data files and format

padj = pd.read_excel("200904_Padj_ALL.xlsx", header=0, index_col=0)
logfc = pd.read_excel("200904_LogFC_ALL.xlsx", header=0, index_col=0
                                    )

combine = pd.concat([logfc, padj], axis = 1)
combine.dropna(axis=0, how='any', inplace=True)

# plot data and save as PNG files

bioinfokit.visuz.gene_exp.volcano(df=combine, lfc='res_logfc', pv='
                                    res_padj',
 color=("red","grey","blue"), axlabelfontsize=10,
 legendlabels=('significant up', 'not significant', 'significant
                                     down'),
 plotlegend=True, legendpos='upper right', legendanchor=(1.45,1),
                                     show=True)

bioinfokit.visuz.gene_exp.volcano(df=combine, lfc='res_logfc', pv='
                                    res_padj',
 color=("red","grey","blue"), axlabelfontsize=10,
 legendlabels=('significant up', 'not significant', 'significant
                                     down'),
 plotlegend=True, legendpos='upper right', legendanchor=(1.45,1),
 r=300, figname='volcano_res')

### repeat with all samples
```

D.5 Plot venn diagrams of proteomic logFC values

```
# import packages
import pandas as pd
import numpy as np
import scipy
import scipy.stats as stats
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.cluster.hierarchy import dendrogram, linkage
from venn import venn, pseudovenn
import inspect
%matplotlib inline
```

```
# import data

padj = pd.read_excel("200904_Padj_ALL.xlsx", header=0, index_col=0)
logfc = pd.read_excel("200904_LogFC_ALL.xlsx", header=0, index_col=0
                                    )

# create substrate specific files of logfc and padj values

phe = pd.DataFrame(index=logfc.index)
phe['logFC'] = logfc['phe_logfc'].values
phe['padj'] = padj['phe_padj'].values

### repeat with all substrates ###

# isolate significant data (a=0.05)

phe_sig = phe[phe['padj'] <= 0.05]
res_sig = res[res['padj'] <= 0.05]
phba_sig = phba[phba['padj'] <= 0.05]
pcoum_sig = pcoum[pcoum['padj'] <= 0.05]
ferul_sig = ferul[ferul['padj'] <= 0.05]
lig_sig = lig[lig['padj'] <= 0.05]

siglogfc = pd.concat([phe_sig['logFC'],
                      res_sig['logFC'],
                      phba_sig['logFC'],
                      pcoum_sig['logFC'],
                      ferul_sig['logFC'],
                      lig_sig['logFC']],
                     axis = 1)

siglogfc.to_csv("significant_logfc.csv")

# create datasets for venn

phe_list = pd.DataFrame() # isolate accession number
phe_list['Accession_phe'] = phe_sig.index # create index using
                                    accession number
venn_phe = phe_list.iloc[:,0] # create series
venn_phe = set(venn_phe) # create set from series

### repeat with all substrates ###

dataset = {
    "Phenol": venn_phe,
    "Resorcinol": venn_res,
    "pHBA": venn_phba,
    "p-Coumarate": venn_pcoum,
    "Ferulate": venn_ferul,
    "Lignin": venn_lig
```

```
}

venn = pseudovenn(dataset, legend_loc = "upper left", figsize= (10,
                                    10))

venn.figure.savefig('venndiag.png', dpi=300, bbox_inches='tight')

# Isolate proteins common to all dataset and export as csv

commonall = venn_phe.intersection(venn_res)\
            .intersection(venn_phba).intersection(venn_ferul)\
            .intersection(venn_pcoum).intersection(venn_lig)
commonall = pd.DataFrame(list(commonall))
commonall.set_index(list(commonall)[0],inplace=True)
commonall = commonall.rename_axis("Accession")

commonall1 = commonall.join(pcoum_sig["logFC"], how='left').rename(
                                    columns={"logFC":"pcoum"})\
            .join(ferul_sig["logFC"], how='left').rename(columns={"
                                            logFC":"ferul"})\
            .join(lig_sig["logFC"], how='left').rename(columns={"
                                            logFC":"lig"})\
            .join(phe_sig["logFC"], how='left').rename(columns={"
                                            logFC":"phe"})\
            .join(res_sig["logFC"], how='left').rename(columns={"
                                            logFC":"res"})\
            .join(phba_sig["logFC"], how='left').rename(columns={"
                                            logFC":"phba"})
commonall1.to_csv("venn_commonall.csv")

# make bar plot for comparison against all datasets

color = ['blueviolet', 'orange', 'lightseagreen', 'g', 'b', 'r'] #
                                    match color to PCA plot
ax = commonall1.plot.bar(figsize=(9,4), width=0.7, color=color,
                                    edgecolor='black', alpha=0.7)
ax.set_ylabel('logFC', fontsize=12)
ax.set_xlabel('Accession', fontsize=12)
ax.tick_params(axis='x', which='major', labelsize=10)
ax.tick_params(axis='y', which='major', labelsize=11)
ax.patch.set_facecolor('w')

ax.figure.savefig('commonprot-all.png', dpi=300, bbox_inches='tight'
                                    )

# plot three dataset venn diagram overlap as a scatter plot
# isolate outliers and color red and label with acession number
# where necessary, use adjust_text package to label outliers

from adjustText import adjust_text
```

```
# make supblots
fig, ax = plt.subplots(1, 3, sharex=True, sharey=False, figsize=(14,
                                    4))
fig.patch.set_facecolor('xkcd:white')

for i in range(0,3):
    ax[i].set_xlim(-7,7)
    ax[i].set_ylim(-7,7)

    # plot y=x line
    lims = [
        np.min([ax[i].get_xlim(), ax[i].get_ylim()]),  # min of both
                                                    axes
        np.max([ax[i].get_xlim(), ax[i].get_ylim()]),  # max of both
                                                    axes
    ]

    ax[i].plot(lims, lims, 'k-', alpha=0.75, zorder=0, linestyle='--
                                        ', c='dimgray')
    ax[i].set_aspect('equal')
    ax[i].set_xlim(lims)
    ax[i].set_ylim(lims)

##

ax[0].scatter(hcf1.phba, hcf1.pcoum)
ax[0].set(xlabel='phba', ylabel='pcoum')
df0d = hcf1.query('phba >=0 & pcoum <=0')
df0u = hcf1.query('phba <=0 & pcoum >=0')
ax[0].scatter(df0d.phba, df0d.pcoum, color="red")
ax[0].scatter(df0u.phba, df0u.pcoum, color="red")
accessiond = df0d.index
xd = df0d.phba
yd = df0d.pcoum
accessionu = df0u.index
xu = df0u.phba
yu = df0u.pcoum
for label,x,y in zip(accessiond,xd,yd):
    ax[0].annotate(
        label,
        xy=(x,y),
        xytext=(-10,10), textcoords='offset points')
for label,x,y in zip(accessionu,xu,yu):
    ax[0].annotate(
        label,
        xy=(x,y),
        xytext=(0,10), textcoords='offset points')


ax[1].scatter(hcf1.phba, hcf1.ferul)
ax[1].set(xlabel='phba', ylabel='ferul')
```

```
df1d = hcf1.query('phba >=0 & ferul <=0')
df1u = hcf1.query('phba <=0 & ferul >=0')
ax[1].scatter(df1d.phba, df1d.ferul, color="red")
ax[1].scatter(df1u.phba, df1u.ferul, color="red")
accessiond = df1d.index
xd = df1d.phba
yd = df1d.ferul
accessionu = df1u.index
xu = df1u.phba
yu = df1u.ferul
for label,x,y in zip(accessionu,xu,yu):
    ax[1].annotate(
        label,
        xy=(x,y),
        xytext=(0,-15), textcoords='offset points')
textsd = []
for x,y,s in zip(xd,yd,accessiond):
    textsd.append(plt.text(x, y, s))
adjust_text(textsd, arrowprops=dict(arrowstyle="-", color='r', lw=0.
                                    75), ax=ax[1])


ax[2].scatter(hcf1.ferul, hcf1.pcoum)
ax[2].set(xlabel='ferul', ylabel='pcoum')
df2d = hcf1.query('ferul >=0 & pcoum <=0')
df2u = hcf1.query('ferul <=0 & pcoum >=0')
ax[2].scatter(df2d.ferul, df2d.pcoum, color="red")
ax[2].scatter(df2u.ferul, df2u.pcoum, color="red")
accessiond = df2d.index
xd = df2d.ferul
yd = df2d.pcoum
accessionu = df2u.index
xu = df2u.ferul
yu = df2u.pcoum
textsd = []
for x,y,s in zip(xd,yd,accessiond):
    textsd.append(plt.text(x, y, s))
textsu = []
for x,y,s in zip(xu,yu,accessionu):
    textsu.append(plt.text(x, y, s))
adjust_text(textsd, arrowprops=dict(arrowstyle="-", color='r', lw=0.
                                    75), ax=ax[2])
adjust_text(textsu, arrowprops=dict(arrowstyle="-", color='r', lw=0.
                                    75), ax=ax[2])

plt.savefig('venn-dge-compare-hcf', dpi=300, bbox_inches='tight')
```

Appendix E Chapter 4 Supplemental Information

This appendix describes supplemental information and supporting data for Chapter Four of this dissertation. The protocols described in this appendix are from Sluiter, et al. and NREL Laboratory Analytical Procedures (LAPs)/TP-510-42618 [1] and NREL LAP/TP-510-42625 [2].

E.1 Supplementary Methods: Compositional lignin analysis

E.1.1 Acid hydrolysis

Acid hydrolysis. Cell culture supernatant was separated from cells by centrifugation at 10,000 rpm for 10 minutes at 4 °C. The supernatant was lyophilized overnight in a 50 mL conical centrifuge tube. Samples were first hydrolyzed. Filtering crucibles were heated in a muffled furnace at 575 °C for a minimum of four hours and transferred to a desiccator to dry for a minimum of one hour. Dried crucibles were pre-weighed, placed back into the furnace at 575 °C, and ashed to constant weight. 300 mg of sample was weighed in a tared pressure tube and thoroughly mixed with 3 mL of 72% sulfuric acid. Samples were incubated in a water bath at 30 °C for one hour. Samples were mixed with a Teflon stir rod every 5 to 10 minutes without being removed from the water bath. After the one hour hydrolysis, the tubes were removed from the water bath, the acid was diluted to a 4% concentration with 84 mL of deionized water, and tubes mixed by inversion. Samples were autoclaved for one hour at 121 °C and then slowly cooled to near room temperature.

E.1.2 Protein content

Protein content was measured according to NREL Laboratory Procedure TP-510-42625. A complete amino acid profile was obtained according to AOAC982.30

E(a,b,c), CHP 45.3.05 (1995). A measurement of the total nitrogen content of each sample was obtained and values were converted to a percent dry weight basis. The amino acid and nitrogen weight percent values were used in equations to determine the nitrogen factor (NF or N-factor) (equation E.5), based on the plausible upper (k_1) and lower (k_2) limits calculated by equation E.3 and E.4, respectively.

$$k_A = \frac{\sum E_i}{\sum D_i} \tag{E.1}$$

where:

E_i = the grams of the ith amino acid per 100 grams of sample (dry weight basis)

D_i = the grams of nitrogen of the ith amino acid per 100 grams of sample (dry weight basis)

$$k_A = \frac{\sum E_i}{N} \tag{E.2}$$

where:

E_i = the grams of the ith amino acid per 100 grams of sample (dry weight basis)

N = the grams of nitrogen per 100 grams of sample (dry weight basis)

$$k_1 = avg\left(k_A, k_P\right) + 0.25\left(avg\left(k_A, k_P\right)\right) \tag{E.3}$$

$$k_2 = avg\left(k_A, k_P\right) - 0.25\left(avg\left(k_A, k_P\right)\right) \tag{E.4}$$

$$NF = avg\left(k_A, k_P\right) \tag{E.5}$$

$$\%\text{protein} = \%\text{nitrogen} \times NF \tag{E.6}$$

E.1.3 Acid insoluble lignin

The acid insoluble lignin was determined by gravimetric analysis. The autoclaved hydrolysates were vacuum filtered through the previously dried and preweighed crucibles and the filtrate was collected for soluble lignin analysis. Deionized water was used to quantitatively transfer the remaining solids in the pressure tube and into the filtering crucible. Solids were rinsed with 50 mL fresh deionized water and then dried at 105 °C for a minimum of four hours. The samples were cooled in a desiccator and weighed. The crucibles were placed into a muffled furnace at 575 °C for approximately 24 hours. The crucibles were removed from the furnace and transferred immediately to a desiccator to cool before being weighed. Ash content is described as acid insoluble residue (AIR) and is calculated from equation E.8. Acid insoluble lignin (AIL) content is calculated by equation E.9.

$$ODW = \frac{W_{\text{air dry sample}} \times \%\ \text{total solids}}{100} \tag{E.7}$$

$$\%\ AIR = \frac{(W_{\text{crucible + AIR}} - W_{\text{crucible}})}{ODW} \times 100\% \tag{E.8}$$

$$\%\ AIL = \frac{(W_{\text{crucible + AIR}} - W_{\text{crucible}}) - (W_{\text{crucible + ash}} - W_{\text{crucible}})}{ODW} \times 100\% \tag{E.9}$$

E.1.4 Acid soluble lignin content

The acid soluble lignin was analyzed by UV-vis spectrophotometry at 320 nm (ε_{320} = 30 L g-1 cm-1). The mass of acid soluble lignin (ASL) was calculated by

equation **??**.

$$\%\ ASL = \frac{UV_{abs} \times Volume_{filtrate} \times Dilution}{\varepsilon_{320} \times ODW \times Pathlength} \times 100\% \tag{E.10}$$

where:

UV_{abs} = average UV-vis absorbance for the sample at a specific wavelength (320 nm)

$Volume_{filtrate}$ = volume of filtrate (mL)

$Dilution$ = $\frac{Volume_{sample}+Volume_{\text{dilution solvent}}}{Volume_{sample}}$

ε_{320} = Absorptivity of lignin at 320 nm (L g^{-1} cm^{-1})

ODW = oven dry weight of samples (mg), see equation E.2

$Pathlength$ = pathlength of UV-vis cells (cm)

E.1.5 Structural carbohydrate content

Approximately 20 mL of the soluble hydrolyzed lignin fraction was neutralized with calcium carbonate to pH 5-6. The precipitate was allowed to settle, and the supernatant was decanted. The decanted liquid was filtered through a 0.2 μm filter. A Biorad Aminex HPX-87H column was used with an eluent of HPLC grade water at a flow rate 0.6 mL/min and temperature of 85 °C. Samples were run over 35 minutes and detected with a refractive index detector (RID).

E.1.6 Acetyl content

An aliquot of the soluble hydrolyzed lignin fraction was filtered through a 0.2 μm filter. A Biorad Aminex HPX-87H column was used with an eluent of 50 mM sulfuric acid at a flow rate of 0.6 mL/min and temperature between 55 and 65 °C. Samples were run over 50 minutes and detected with an RID.

Bibliography

[1] Sluiter, A., Hames, B., Ruiz, R., Scarlata, C., Sluiter, J., Templeton, D., and Crocker, D. Determination of Structural Carbohydrates and Lignin in Biomass. Technical Report: NREL/TP-510-42618. Issue Date: April 2008. url: https://www.nrel.gov/docs/gen/fy13/42618.pdf

[2] Hames, B., Scarlata, C., and Sluiter, A. Determination of Protein Content in Biomass. Technical Report: NREL/TP-510-42625. Issue Date: 23 May 2008. url: https://www.nrel.gov/docs/gen/fy08/42625.pdf

Appendix F *Thraustochyrium striatum* genome analysis

This appendix describes the genome extraction, sequencing, and annotation work performed on *Thraustochytrium striatum*, a marine protist, on behalf of Dr. Zheng Yi at Kansas State University. Briefly, *T. striatum* is a marine protist able produce omega-3 fatty acids and carotenoids. Recently, it was shown to grow in Kraft lignin hydrolysate [1, 2]. Lack of genome sequence or annotation has motivated the below work, which is on track to be published as a Genomes Announcement [3]. This appendix expands on the publication and provides details on methods.

F.1 Genome extraction for PacBio sequencing

Cells were cultured by members of the Yi laboratory [1]. Cells were pelleted in 50 mL centrifuge tubes, frozen on dry ice, and kept at -80 °C until use. Cells were then lyophilized overnight.

The DNA extraction protocol below is modified from a method for RNA extraction in Pine species [4]. I made my own tweaks to the protocol based on previous extractions I performed of this protist using other protocols. The base protocol was provided by the Clemson University Genome Institute (CUGI), which has since become Clemson University Genomics and Bioinformatics Facility (CUGBF). This work was done with help from Penny Xia.

Prepare extraction buffer by mixing all of the following except the last two components. Just before using the buffer, add the spermidine and BME. Listed concentrations are final concentrations:

- 100 mM Tris-HCl, pH 8.0

- 25 mM EDTA
- 2 M NaCl
- 2% (w/v) SDS
- 1% sodium metabisulfite
- 500 mg/L spermidine
- 0.1% (v/v) 2-mercaptoethanol (BME)

Follow the below protocol to extract high molecular weight DNA from the protist *Thraustochyrium striatum*. To prevent shearing from the force of pipetting, I suggest you use no smaller than 1000 μL pipette tips when possible, and cut the tip of the pipette off with a sterile razor. You may also use wide-bore pipette tips.

1. Grind lyophilized cell pellet to a fine powder with dry ice using a clean mortar and pestle.
2. Let the dry ice sublimate off in a chemical hood.
3. Transfer powder to a 50 mL centrifuge tube and add 10 mL of extraction buffer. Use 10 mL of extraction buffer to rinse the mortar and pestle and transfer to the centrifuge tube.
    - There may be foaming upon addition of the extraction buffer if the dry ice has not fully sublimated off.
4. Add 100 μL of Proteinase K (20 mg/mL). Mix by inverting.
5. Incubate overnight (>10 hours) at 50 °C. I mixed by gentle inversion after the first 2 hours.

6. Cool down at room temperature for 10-15 minutes.

7. In a chemical hood, add 1 volume (~20 mL) of chloroform and mix well by gentle inversion. Be gentle, as we are preserving the molecular weight of the DNA.

8. Centrifuge in a floor centrifuge at 3000 rpm and 4 °C for 20 minutes.

9. Collect the aqueous (top) layer and repeat the chloroform extraction twice for a total of 3 extractions.

10. Add 0.7 vol (~12 mL) of isopropanol and invert gently to mix.

11. Incubate at -20 °C for at least 1 hour. I left it overnight.

12. Centrifuge sample at 3000 rpm and 4 °C for 10 minutes. This will pellet the DNA as a white pellet at the bottom of the 50 mL centrifuge vial.

13. Pour off the supernatant. Add 10 mL of ice cold 70% molecular grade ethanol (do not pipette) and centrifuge as before.

14. Repeat the ethanol wash step with 5 mL of ice cold 70% ethanol and centrifuge as before.

15. Air dry the pellet for ~20 minutes. I left the tube upside down on a kimwipe in the chemical hood.

16. Add 500 μL of 10 mM Tris-HCl (pH 7.5). Leave at 4 °C overnight to resuspend and periodically shake *gently* to mix. Do not pipette.

The end result is 0.5 mL of very viscous, high molecular weight DNA. I did not add any sort of stabilizer nor did I try eluting with a Tris buffer. The DNA

Table F.1. Table of statistics for the PacBio assembly of *T. striatum*.

Assembly statistic	**Value**
Number of scaffolds	128
Total size of scaffolds (bp)	39,505,394
Longest scaffold (bp)	2,105,201
Shortest scaffold (bp)	587
No. of scaffolds > 1K bp	125 (97.7%)
No. of scaffolds > 10K bp	117 (91.4%)
No. of scaffolds > 100K bp	65 (50.8%)
No. of scaffolds > 1M bp	10 (7.8%)
No. of scaffolds > 10M bp	0 (0.0%)
Mean scaffold size (bp)	308,636
Median scaffold size (bp)	128,249
N50 scaffold length (bp)	869,324
L50 scaffold count (bp)	15
Scaffold %C	32.63
Scaffold %G	32.54

was stored at 4 °C and was subsequently lost in a fridge for a very long time, but remained in tact! Run with pulsed field gel electrophoresis to determine size of the extracted DNA. For suitable PacBio sequencing, fragments must be around 40 kb. I submit 1 μg of DNA for sequencing. The gDNA size was verified by pulsed field gel electrophoresis.

F.2 PacBio sequencing

Whole-genome sequencing was performed using PacBio single-molecule real-time (SMRT) sequencing. PacBio SMRTbell library creation and sequencing was performed by Dr. Brewster Kingham at the University of Delaware DNA Sequencing and Genotyping Center. SMRTbell libraries were sequenced on a PacBio RSII platform (P6/C4 chemistry), resulting in 90X coverage. Reads were assembled with MECAT2. Table F.1 summarizes the statistics of the finished scaffold-level assembly.

F.3 Genome annotation

PacBio sequencing reads were uploaded to GenSAS v6.0 for functional analysis. The FASTA sequence file resulting from PacBio was uploaded to the online GenSAS application and all software were run using default parameters.

The following software were accessed:

- Repeat masking: RepeatMasker v4.0.7, RepeatModeler v1.0.11
- Alignments: BLASTn v2.7.1, BLAT v35, DIAMOND v0.9.22
- Gene prediction: Augustus v3.3.1, GeneMarkES v4.33, GlimmerM v2.5.1
- Genomic features: tRNAScane-SE v1.2
- Consensus gene model: EvidenceModler
- Functional annotation: BLASTp v2.6.0, DIAMOND v0.9.22, InterProScan, Pfam v1.6, SignalP v4.1, TargetP v1.1

I tried using PASA for transcript alignment and RNAmmer to find rRNA sequences, but both software failed. Ultimately, the best gene sets were from GeneMarkES, an *ab initio* gene predictor. BUSCO scored these gene predictions as being 26.5% complete. The next best set was from the conserved gene set, provided by EvidenceModeler, which gave a BUSCO completeness score of 23.8%. BUSCO scores for Augustus and GlimmerM predictions were less than 10%. The masked GeneMarkES gene set was marked as the official gene set (OGS) for functional annotation.

The resulting annotation lists 11,269 genes and 11,262 proteins. The identified genes comprise 71.6% of the genome. 83% (9,307) of annotated proteins have Pfam

domains and 61% result in a homology match on BlastP. The average protein length is 630 amino acids and the average gene length is 2,514 base pairs.

F.4 Future work

This project would benefit from proteomic and/or transcriptomic data. Both would bolster the genome annotation and provide more information about relevant genes for metabolism of lignin-relevant substrates and production of carotenoids and lipids.

Bibliography

[1] Li X, Li M, Pu Y, Ragauskas AJ, Zheng Y. 2020. Black Liquor Valorization by Using Marine Protist *Thraustochytrium striatum* and the Preliminary Metabolic Mechanism Study. *ACS Sustainable Chemistry & Engineering* **8**:1786-1796.

[2] Li X, Li M, Pu Y, Ragauskas AJ, Zheng Y. 2020. Simultaneous depolymerization and fermentation of lignin into value-added products by the marine protist, *Thraustochytrium striatum*. *Algal Research* **46**.

[3] Yaguchi, A., Li, X., Xiao, R., Zheng, Y., Saski, C., and Blenner, M. The first draft genome and annotation of marine protist, *Thraustochytrium striatum*. *Genome Announ*. Manuscript in preparation.

[4] Brunner I, Brodbeck S, Buchler U, Sperisen C. 2001. Molecular identification of fine roots of trees from the Alps: reliable and fast DNA extraction and PCR–RFLP analyses of plastid DNA. *Molecular Ecology* **10**:2079-2087.

Appendix G Investigation of eccDNA from *Amaranthus palmeri* as an origin of replication

This appendix describes the work to establish functionality of a putative autonomic replication sequence (ARS) derived from extrachromosomal circular DNA (eccDNA) from *Amaranthus palmeri*.

G.1 Cloning eccDNA ARS and transforming into yeast

Work by Dr. William Molin and Dr. Chris Saski identified an eccDNA sequence that had potential to act as an ARS. I cloned the putative ARS from the eccDNA into non-replicative plasmid, pRS305, and replicative plasmid, pRS315, using primers referenced in G.2 with a standard SLIC reaction. Briefly, the CEN6 sequence from pRS315 and the ARS1 sequence from the eccDNA were PCR amplified. They were assembled using a three way SLIC assembly with a PCR linearized pRS305 backbone. The resulting plasmid was called pRS305+CS-ARS1 (Figure G.1A). To create pRS315+CS-ARS1, the native pRS315 ARS was removed and replaced with the eccDNA ARS (Figure G.1B). Ultimately, the resulting plasmids both are the same in sequence (Table G.1). All cloning was performed using NEB DH5α cells and verified by restriction digest and sequencing.

Table G.1. Table of strains and plasmids used to validate the eccDNA ARS.

Strain	**Genotype**
ATCC 208288	*MATα ura352 trp1 leu2Δ1 his3Δ200 pep4::HIS3 prb1Δ1.6R can1 GAL*
Plasmid	**Feature**
pRS305	pBR332 origin, AmpR, LEU2
pRS315	pRS305, CEN6/ARS
pRS305+CS-ARS1	pRS305 + CEN6 + CS-ARS1
pRS315+CS-ARS1	pRS315ΔCEN6/ARS + CEN6 + CS-ARS1

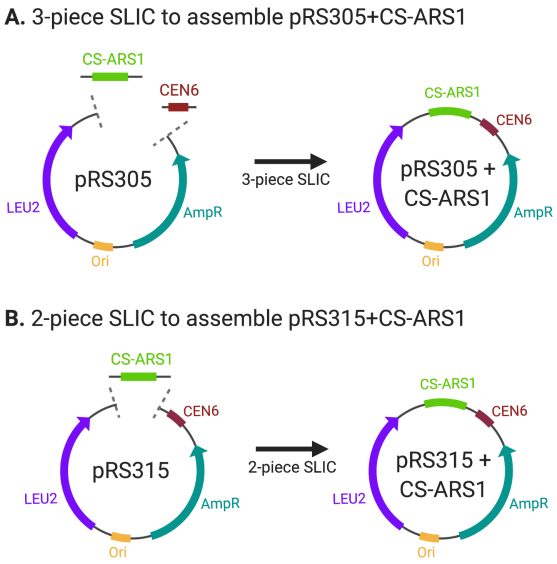


Figure G.1. Cloning strategy to create plasmids containing the eccDNA ARS.

Table G.2. Table of primers used to clone the eccDNA ARS.

Primer	**Sequence (5' → 3')**
167,312F_CEN-SLIC	AAAATAGTTTTTGTTTTCCGAAGATGTAA
	AAAGAGGATGGGATTTGTTACACTCA

Continued on next page

Table G.2 – *Continued from previous page*

Primer	**Sequence (5' → 3')**
167,467F_CEN-SLIC	AAAATAGTTTTTGTTTTCCGAAGATGTAA
	ATTAGCCATCCCCTTTTCTCAACTA
168,187R_SLIC	TATAGGTTAATGTCATGATAATAATGGTT
	TCTTAGCTTTCTCATCTTCCTCTGTACCC
	AA
pRS_ΔCEN-F	CTAAGAAACCATTATTATCATGACATTAA
	CCTATA
CEN-SLIC-F	CAGGTGGCACTTTTCGGG
pRS_ΔARS-R	TTTACATCTTCGGAAAACAAAAACT
ARS_cPCR_F	AAGAGGATGGGATTTGTTACACTCA
ARS_cPCR_R	CTTTCTCATCTTCCTCTGTACCCAA

Saccharomyces cerevisiae (ATCC 208288) cells were transformed in triplicate as previously described [3]. Briefly, yeast cells were grown in a YPD (10 g/L yeast extract, 20 g/L peptone, 20 g/L glucose) preculture overnight at 28 °C and 250 rpm. In a 250-mL baffled flask, 50 mL of pre-warmed YPD was inoculated to a final titer of 5 x 10^6 cells/mL. The culture was grown to a final titer of 2 x 10^7 cells/mL at 28 °C and 250 rpm. Cells were harvested by centrifugation at 3,000 x g for 5 minutes. The cell pellet was resuspended in 25 mL of sterile milli-Q water and centrifuged again three times before cells were resuspended in 1.0 mL of sterile water. Cell pellet was then resuspended in 360 μL freshly made transformation mix (240 μL PEG 3350 (50% w/v), 34 μL 1.0 M LiAc, 50 μL single-stranded salmon sperm DNA (2 mg/mL), 36 μL plasmid DNA plus sterile water). Cells were heat shocked at 42 °C for 40 minutes and resuspended in 1 mL of sterile milli-Q water. 200 μL of cells were plated on

YSC-Leu+2% glucose plates and grown at 28 °C for 2 days and colonies counted. pRS305 lacks an ARS and served as a negative control while pRS315 contains an ARS and served as a positive control.

To confirm plasmid retention, colonies from transformed plates were passaged on YSC-Leu+2% glucose plates three times. The passaged cells were used for a colony PCR using Q5 polymerase and primers ARS_cPCR-F and ARS_cPCR-R. The original colonies from the pRS305 transformation plate were used as templates, as they did not survive passaging. Positive controls using plasmids pRS305+CS-ARS and pRS315+eccARS and a negative control using wild type *S. cerevisiae* cells were performed. The PCR products were run on a 1% TAE gel with Genscript Ready-to-Use™ Plus 100 bp DNA Ladder.

G.2 Proving eccDNA ARS functionality in yeast

Figure G.2A and B demonstrates the transformation efficiencies for eccDNA ARS-containing plasmids as compared to the controls, pRS305 and pRS315. In G.2A, the gray data points indicate individual points where the black data point indicates the set average. The asterisks indicate statistical significance, as is listed in Figure G.3. The positive control, pRS315 is several orders of magnitude higher in efficiency than the pRS315 plasmid that has the eccDNA ARS swapped into it (pRS315+CS-ARS1). The efficiencies of the eccDNA ARS containing plasmids are statistically higher than efficiencies of the pRS305 plasmid, indicating presence of the eccDNA ARS acts as a functional, albeit weak, ARS in yeast.

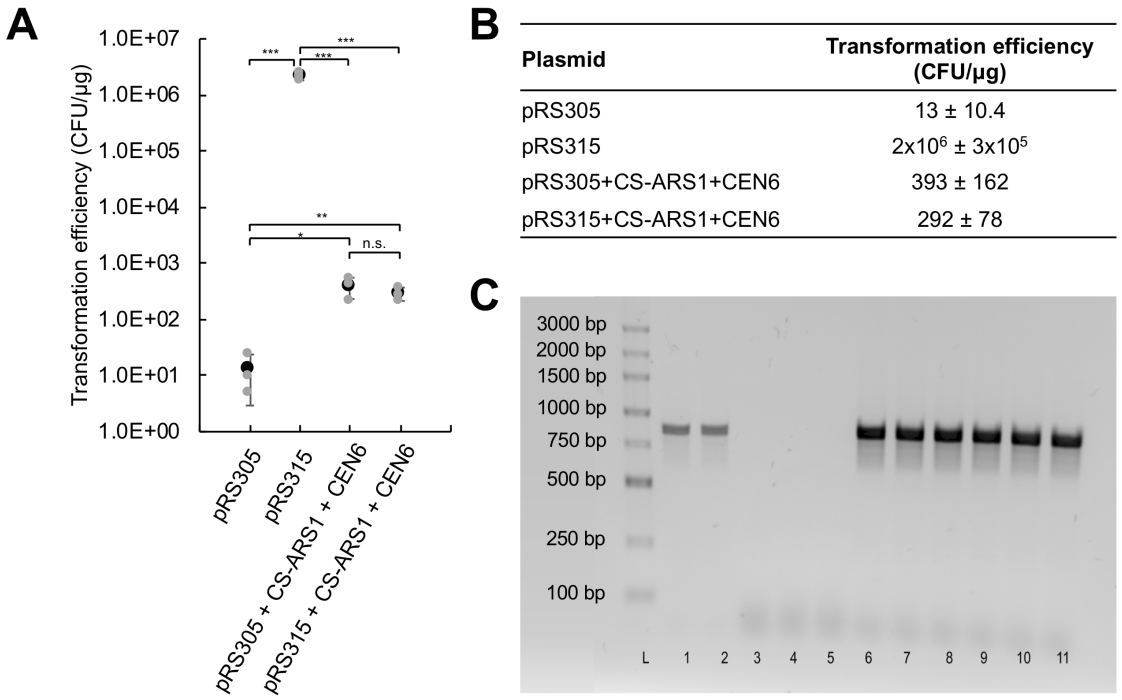


Figure G.2. Transformation efficiencies and cPCR results.

p-Values from Two-Tailed t-Test (α = 0.05)				
	pRS305	pRS315	pRS305+CS-ARS1+CEN6	pRS315+CS-ARS1+CEN6
pRS305	-	-	-	-
pRS315	0.00034	-	-	-
pRS305+CS-ARS1+CEN6	0.01531	0.00034	-	-
pRS315+CS-ARS1+CEN6	0.00362	0.00034	0.38227	-

Figure G.3. Tabulated p-values derived from a two-tailed t-test using a confidence interval of 95%.

To validate plasmid retention and stability, cells were passaged on selective plates three times. Passaged cells were used in a colony PCR to validate retention of the ARS sequence (Figure G.2C). Demonstration of bands in the cells transformed with plasmids containing CS-ARS1, and not in the wild type or cells transformed

with pRS315, indicates dividing plasmids are due to the eccDNA ARS1.

G.3 Future work

I could see this project going in a couple of different directions. The exact functions and mechanisms for the eccDNA ARS are unknown. A deeper level understanding of the function in the native plant host could be useful when trying to decontextualize the ARS and port it to a different organism. Additionally, the ARS could be shuffled with the native yeast ARS to try and increase transformation efficiencies. Perhaps understanding which sequences are key will enhance the understanding of the plasmid replication and retention in yeast. Alternatively, one could focus more on the putative DUE sequences and create a degenerate oligo library to similarly probe the DUE specifically. On a much broader scale, non-conventional organism such as *C. oleaginosus*, discussed throughout the main chapters of this dissertation, still lack replicating plasmids. An ARS sequence such as this could be useful in establishing one.

G.4 Acknowledgments

This work was performed in collaboration with Dr. Chris Saski at Clemson University and Dr. William Molin at the USDA. This appendix is based off Ref. 1, although Ref. 2 is a closely related computational counterpart. According to the Creative Commons CC BY license, I am not required to obtain special permissions since Ref. 1 is open access.

This material is based upon work supported by NIFA/USDA, under Project number SC-1700530, Technical Contribution No. 6871 of the Clemson University Experiment Station, and Project number 2017-026 from the Clemson University Re-

search Foundation.

Bibliography

[1] Molin, WT; Yaguchi, A; Blenner, MA; Saski, CA. Autonomous replication sequences from the *Amaranthus palmeri* eccDNA replicon enable replication in yeast. *BMC Res. Notes*. **2020**.

[2] Molin, WT; Yaguchi, A; Blenner, MA; Saski, CA. The eccDNA Replicon: A Heritable, Extra-Nuclear Vehicle that Enables Gene Amplification and Glyphosate Resistance in *Amaranthus palmeri*. *Plant Cell*. **2020**.

[3] Gietz, RD; Schiestel, RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, **2007**, 2, 31-34.