1. Identify the DOE award number; name of recipient; project title; name of project director/principal investigator; and consortium/teaming members.

Award #: DE-SC0012705

Title: Systems Biology of *Rhodococcus opacus* to Enable Production of Fuels and Chemicals from Lignocellulose

Recipient: Washington University in St. Louis

PI: Gautam Dantas

Co-PIs: Tae Seok Moon; Marcus Foston

Project Period: 09/01/2014 - 08/31/2018

2. Display prominently on the cover of the report any authorized distribution limitation notices, such as patentable material or protected data. Reports delivered without such notices may be deemed to have been furnished with unlimited rights, and the Government assumes no liability for the disclosure, use or reproduction of such reports.

Many of the research disclosures in this report are unpublished and are patentable. Accordingly, we request that the contents herein are not publically disclosed without prior consultation of the authors/PIs.

3. Provide an executive summary, which includes a discussion of 1) how the research adds to the understanding of the area investigated; 2) the technical effectiveness and economic feasibility of the methods or techniques investigated or demonstrated; or 3) how the project is otherwise of benefit to the public. The discussion should be a minimum of one paragraph and written in terms understandable by an educated layman.

We have developed *Rhodococcus opacus* PD630 (hereafter *R. opacus*) as a chassis for lignocellulose valorization and biofuel production. Specifically, we completed adaptive evolution of *R. opacus* to increase tolerance and growth on mixtures of model lignin breakdown products (LBPs) and isolated adapted strains that showed up to 1900% growth improvement and up to a 225% increase in lipid titer. We identified genomic, transcriptomic, and lipidomic changes in adapted mutants that may increase phenolic tolerance and identified phenolic transporters and degradation pathways. We developed genetic tools for gene expression, gene insertions, gene knockouts, and targeted gene repression. These tools were applied to characterize aromatic transporters and aromatic degradation pathways. In addition to model LBPs, we have optimized a feedstock-independent thermochemical lignin depolymerization process for subsequent degradation by *R. opacus* and have determined the preferred phenolic substrates for *R. opacus* consumption. Overall, this project has contributed to the goal of renewable biofuel and chemical production from lignocellulose by increasing the efficiency of model LBP degradation and lipid production; identifying preferential conditions for lignin depolymerization; characterizing preferred LBP substrates; identifying degradation pathways; and developing tools for further characterization and engineering of *R. opacus*.

4. Provide a comparison of the actual accomplishments with the goals and objectives of the project.

Aim 1. Understand native genetic networks in *R. opacus* for metabolizing phenolic compounds.

Aim 1.1: Strain evolution and whole genome sequencing (WGS).

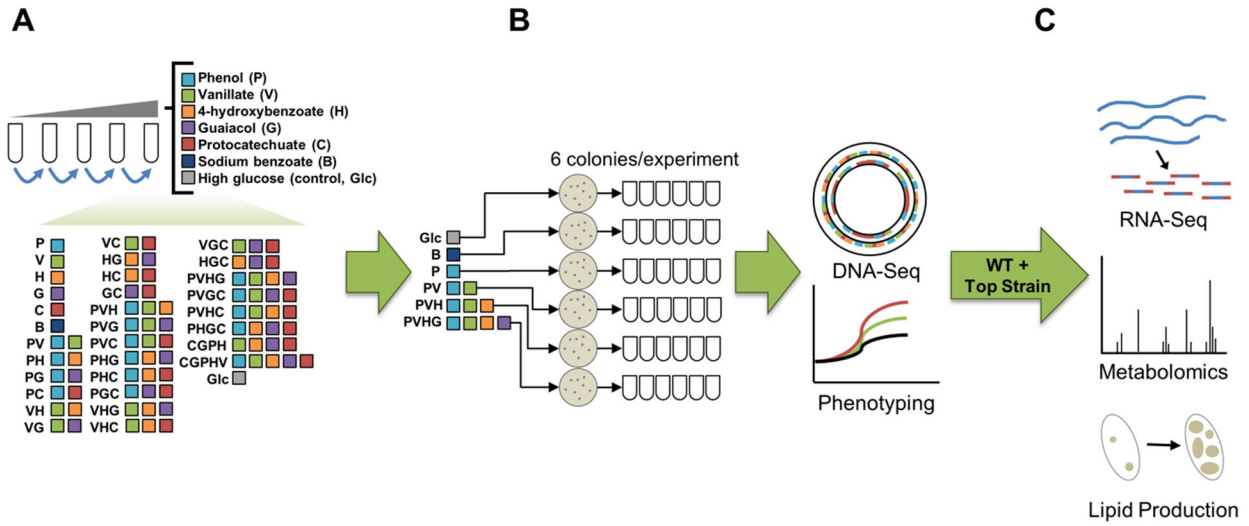


Figure 1. Experimental approach for adaptive evolution experiment[Pub2]. A) *R. opacus* PD630 was serially passaged on increasing concentrations of model lignin breakdown products and high glucose. B) Six colonies were selected from six adaptive experiments for confirmation of the increased growth phenotype before whole genome sequencing. C) The strain with the best growth phenotype on the combined mixture (PVHG6) and the WT strain were further characterized using transcriptomics and targeted metabolomics. Lipid titers were measured to determine the ability of the adapted strain to convert model lignin breakdown products into lipids.

To interrogate genomic targets involved with aromatic degradation and tolerance, we adaptively evolved *R. opacus* on glucose, the aromatic sodium benzoate (BEN), and all combinations of five phenolics including phenol (PHE), vanillate (VAN), 4-hydroxybenzoate (HBA), guaiacol (GUA), and protocatechuate (PCA; Fig 1A). These model compounds, excluding glucose, were chosen based on preliminary analysis of depolymerized lignin. During our adaption experiments, we serially passaged *R. opacus* on 33 different culture conditions with the number of subcultures ranging from 40 to 104 resulting in over 2,000 subcultures and thousands of potential evolved strains[2]. Passaging was terminated when adapted strains no longer achieved further increased growth phenotypes. From these adaption experiments, we selected 6 colonies each from the 4 adaption experiments that showed the most improved growth compared to wild type, as well as 6 colonies from both the sodium benzoate and glucose adaption experiments (see Fig 1A). Of these 36 isolates, 35 showed an improved growth phenotype when grown on the model lignin breakdown product (LBP) substrates (Fig 2). Adaptation and growth assays were originally planned to be conducted in 96-well plates, but this had to be modified to much larger volumes (50 mL glass tubes) due to inconsistent growth phenotypes on certain phenolic compounds. Despite these limitations, we still achieved high-throughput completion of our adaptations and growth assays.

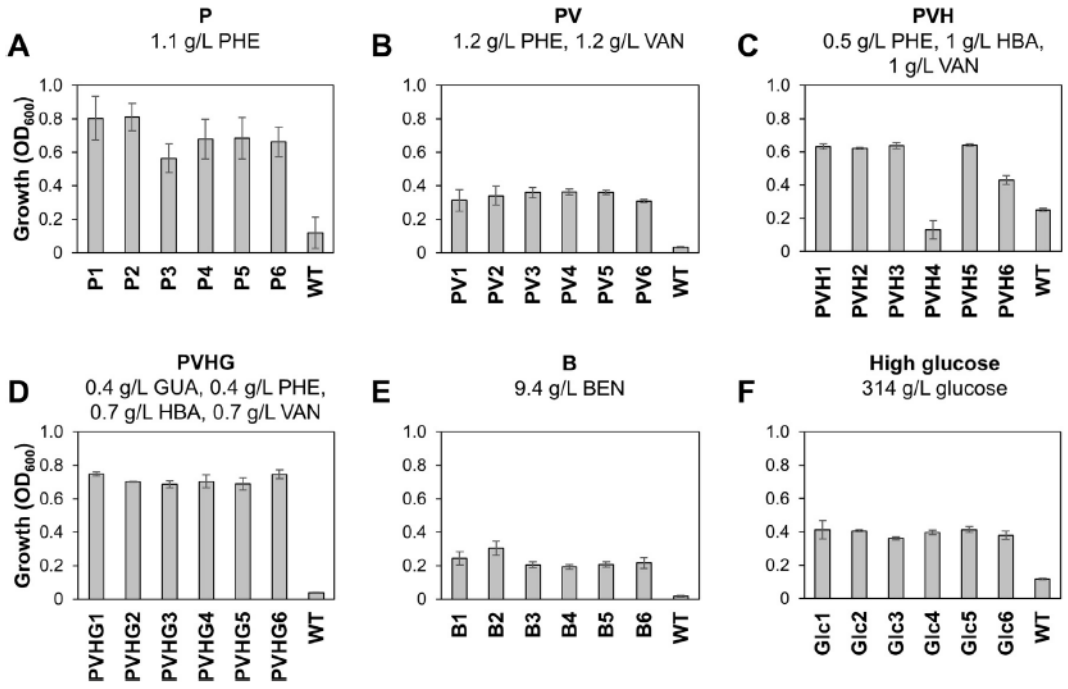
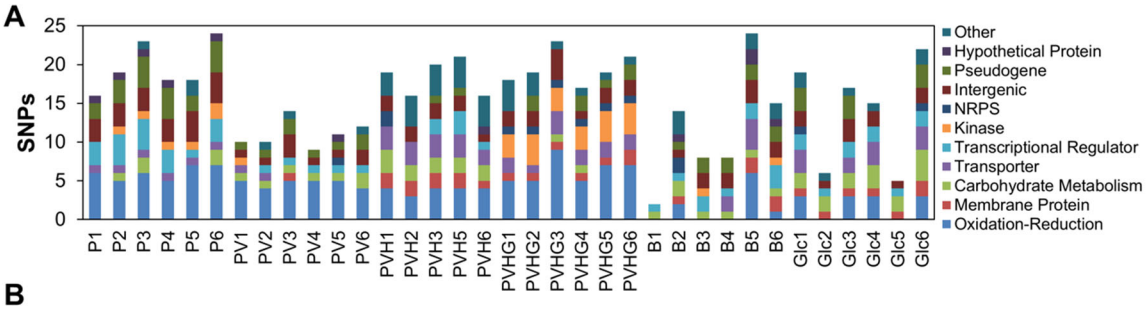


Figure 2. Growth of *R. opacus* strains using adaptation carbon sources[Pub2]. Bars represent the average OD_{600} values of three biological replicates and error bars represent one standard deviation. A) The growth of phenol (PHE) adapted strains (P1-6) compared to the WT strain. B) The growth of the PHE and vanillate (VAN) adapted strains (PV1-6) compared to the WT strain. C) The growth of the PHE, VAN, and 4-hydroxybenzoate (HBA) adapted strains (PVH1-6) compared the WT strain. D) The growth of the PHE, VAN, HBA, and guaiacol (GUA) adapted strains (PVHG1-6) compared to the WT strain. E) The growth of the sodium benzoate (BEN) adapted strains (B1-6) compared to the WT strain. F) The growth of the high glucose adapted strains (Glc1-6) compared to the WT strain.



B

Strains	Gene ID	Annotation	Location	Base Mean Expression
All P, All PV, PVH6, All PVHG	RS01890	superoxide dismutase	Chromosome	32,010
All B, All PVH	RS08140	phosphoenolpyruvate carboxykinase	Chromosome	30,198
P2, P3, P6, All PV	RS10095	class II fructose-bisphosphate aldolase	Chromosome	5,291
All P, All PV, All PVH, All PVHG	RS14825	cytochrome ubiquinol oxidase subunit I	Chromosome	23,629
PV3, All PVHG	RS32480	phosphotransferase	Chromosome	285
P5, PV3, PVH3, PVH5, B6	RS38570	FAD-binding oxidoreductase	Plasmid 1	60
P2, P3, P4, P5, P6, All PVHG	RS39015	serine/threonine-protein kinase PknK	Plasmid 3	354

Figure 3. Genome modifications in adapted strains[Pub2]. A) Categorization of non-synonymous SNPs and intergenic SNPs according to Gene Ontology (GO) terms organized by strain. B) Genes containing non-synonymous SNPs or intergenic SNPs that could alter gene expression that were found in strains across multiple adaptation experiments.

To examine mutations that arose during the adaptation that may be linked with improved aromatic tolerance and degradation phenotypes, we performed genome sequencing of the 35 improved strains. The genes affected by mutations (SNPs, deletions, insertions) were classified based on Gene Ontology into fundamental groups (Fig 3A). In strains adapted on phenolic compounds, we found an enrichment of mutations in genes involved in oxidation-reduction reactions. We identified genes that were consistently mutated across strains in different adaption conditions and the two genes that were mutated in the most strains, cytochrome ubiquinol oxidase subunit I and superoxide dismutase, are involved with oxidation-reduction (Fig 3B). Furthermore, a number of strains were found to have mutations affecting transcriptional regulators and kinases, which can have far reaching impacts on transcription and protein activities.

In addition to identifying and classifying SNPs, we also discovered that one or two of the nine total native *R. opacus* plasmids were missing or significantly reduced in size in all strains adapted on phenolic compounds[Pub2]. These are very large plasmids (~100kbp or larger) and the adaptive advantage to plasmid loss may be linked to a decreased cellular burden during cell replication. To interrogate this further, we determined the copy number of all the native plasmids using qPCR to characterize the native plasmid burden (Fig 4).

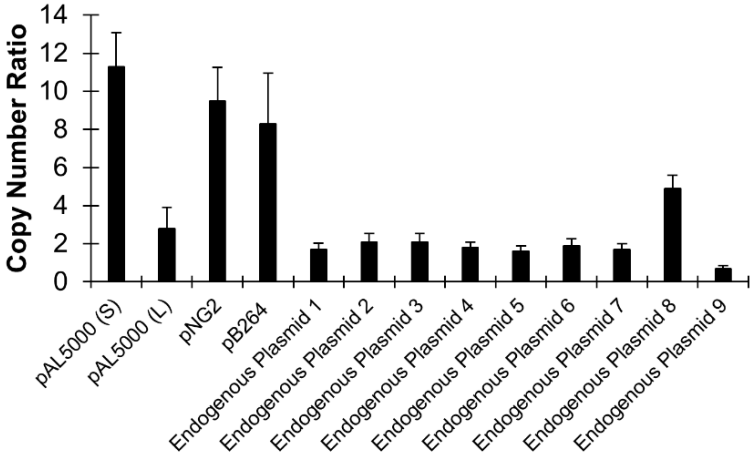


Figure 4. Copy number ratios of heterologous and endogenous plasmids in *R. opacus*[Pub4]. The numbers of copies of the four heterologous and nine endogenous plasmids relative to the single-copy chromosome were determined by qPCR. Values are averages of three replicates grown in minimal medium A, and error bars represent the propagated standard deviation.

Aim 1.2: High-throughput (HT) and mid-throughput (MT) growth/TAG assays.

From each adaptive experiment in Aim 1.1, we picked the evolved isolate with the best growth on a mixture of GUA, PHE, HBA, VAN, and BEN to find the strain with the best growth on a model lignin breakdown mixture. Narrowing down our mutants from this large adaption experiment to a single ideal strain was required for future analyses due to the previously mentioned switch from small volume (96-well plate) cultures to large volume (50 mL glass tubes). We found that isolate PVHG6, a strain adapted on GUA, PHE, HBA, and VAN, had the best growth profile reaching a final OD_{600} value that was 96% higher than the WT strain (Fig 5).

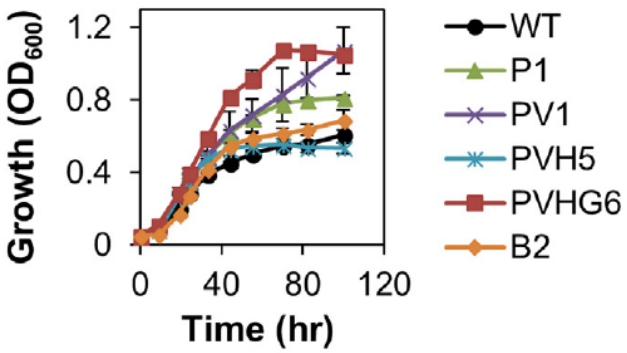


Figure 5. Growth on a mixture of lignin model compounds by WT and adapted strains[Pub2]. Growth of one strain from each selected aromatic adaptation experiment and the WT strain in a mixture of lignin model compounds (0.5 g/L each of GUA, PHE, HBA, VAN, and BEN; 2.5 g/L total aromatics). Points represent the average of three biological replicates and error bars represent one standard deviation. PHE = phenol, VAN = vanillate, HBA = 4-hydroxybenzoate, GUA = guaiacol, and BEN = sodium benzoate.

To determine if the adapted growth phenotype of PVHG6 also led to an increased lipid production, we optimized quantitative and semi-quantitative techniques to measure TAG concentrations in adapted and WT strains. A modified Folch lipid extraction technique was optimized to quantify total lipid content in the WT and PVHG6 strains (Fig 6A). To analyze samples more quickly, we optimized a Nile red fluorescence technique for rapid, semi-quantitative analysis of lipid productivity using 96-well plates and correlated it to lipid content values determined by Folch extraction (Fig 6B). In both cases, the adapted strains produced more lipids than the WT strain when grown on phenolic lignin model compounds indicating that lipid production is also selected for in adaption experiments.

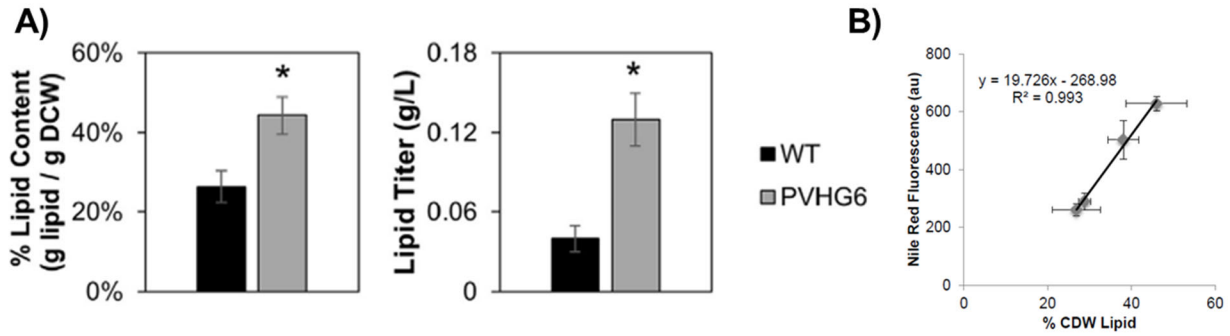


Figure 6. Conversion of lignin model compounds to lipids by *R. opacus*[Pub2]. A) Growth of wild type (WT) and best aromatic mixture adapted strain (PVHG6) in lipid accumulating conditions (low nitrogen) with 0.5 g/L of PHE, VAN, HBA, GUA, and BEN (2.5 g/L total aromatics) as carbon sources. Lipid content ((% CDW lipid) and lipid titer were determined by Folch extraction. Bars represent the average of three biological replicates and error bars represent one standard deviation. Asterisk (*) indicates that the difference between WT and PVHG6 values is statistically significant. B) Correlation between Nile red fluorescence and total lipid content (% CDW lipid). The Nile red fluorescence values represent the average of three technical replicates from two biological replicates for a total of six replicates. The % CDW lipid values are the average of two biological replicates. The error bars represent the standard error.

Aim 1.3: Transcriptional profiling of *R. opacus* strains in various conditions by RNA-Seq.

To identify transcriptional changes potentially related to phenolic tolerance and degradation, we performed RNA-seq on WT and PVHG6 in a number of growth conditions. We grew both strains on glucose, each individual aromatic compound, and a mixture of 5 model LBPs (GUA, PHE, HBA, VAN, and BEN). Using principal component analysis to cluster the transcriptomic profiles based on composition, we found significant differences not just between WT and PVHG6, but also differences between growth conditions for the same strain. (PERMANOVA; P = 0.002; Fig 7). It is presently unclear how these global transcriptome changes are occurring between WT and PVHG6 as there is not a strong link between observed genomic mutations and changes in the transcriptome. This topic is one that warrants a significant amount of further study, as it is likely very central to the improvements in growth and tolerance of PVHG6 on a mixture of aromatic compounds.

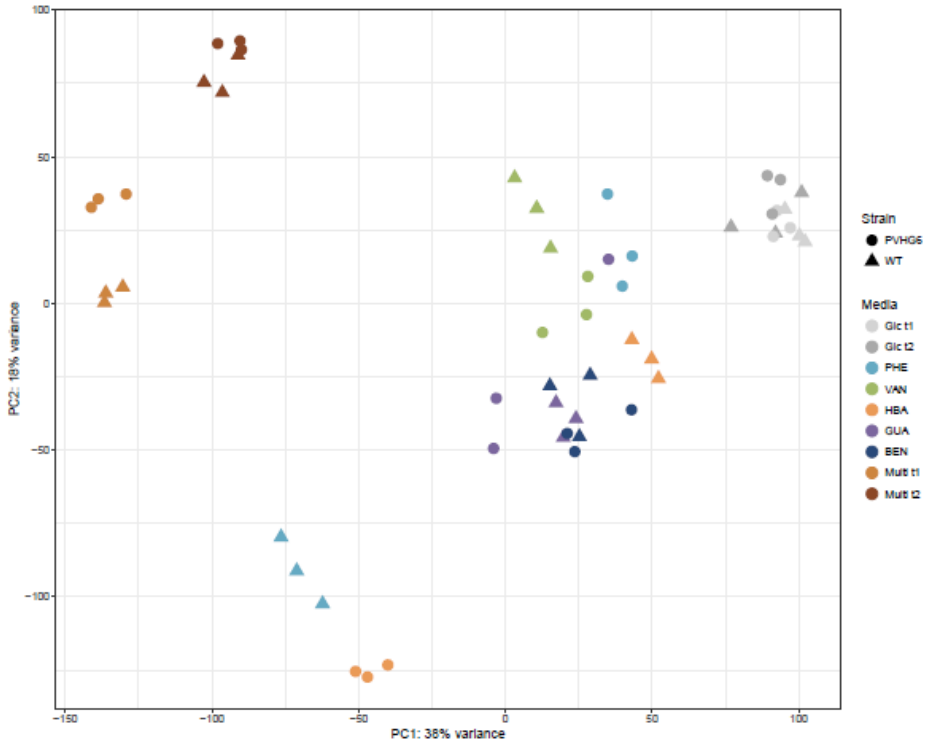


Figure 7. Principal component analysis of all WT and PVHG6 transcriptomes[Pub2]. Principal component analysis was performed on the transcriptomes of wild type (WT) and aromatic mixture adapted strain (PVHG6) using normalized counts of all genes in *R. opacus* except those located on Plasmids 1 and 2, which were often found to have been lost from the adapted mutants. Glc = 1 g/L glucose, Multi = 0.5 g/L of PHE, VAN, HBA, GUA, and BEN (2.5 g/L total aromatics). Other compounds listed in legend were used at a concentration of 0.5 g/L. Each symbol represents a biological replicate. Analysis was performed with normalized counts using the variance stabilizing transformation from the DESeq2 software package. PC1 = 1st principal component and PC2 = 2nd principal component. PHE = phenol, VAN = vanillate, HBA = 4-hydroxybenzoate, GUA = guaiacol, and BEN = sodium benzoate.

In order to elucidate the aromatic degradation funneling pathways in *R. opacus,* we used the transcriptomic profiles on the individual aromatic compounds to identify differentially upregulated genes and found clear patterns of gene upregulation for each (Fig 8). In general, the magnitude of upregulation of these genes is similar between WT and PVHG6, indicating that a further upregulation of funneling enzymes is likely not the source of the observed improvements in growth and tolerance of PVHG6.

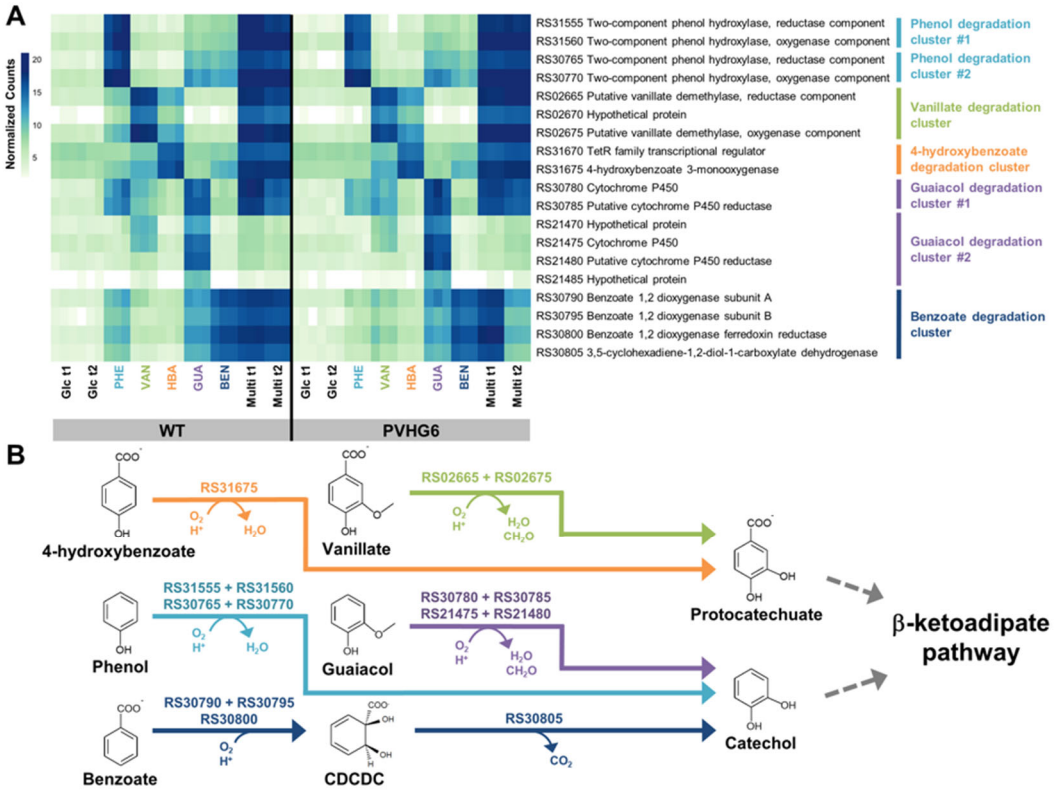


Figure 8. Activation of lignin model compound funneling pathways in *R. opacus*^{Pub6}. A) Heat map of normalized expression counts for upregulated aromatic degradation genes that convert (i.e. "funnel") lignin model compounds to the β-ketoadipate pathway during growth on individual lignin model compounds, a mixture of lignin model compounds, and glucose. Raw counts were transformed using the variance stabilizing transformation in DESeq2. Darker colors indicate higher normalized counts (see scale bar). Glc =1 g/L glucose, PHE =0.5 g/L phenol, VAN =0.5 g/L vanillate, HBA =0.5 g/L 4-hydroxbenzoate, GUA =0.5 g/L guaiacol, BEN =0.5 g/L sodium benzoate, and Multi =0.5 g/L of PHE, VAN, HBA, GUA, and BEN (2.5 g/L total aromatics). t1 represents cells harvested at early exponential phase, while t2 represents cells harvested at mid-exponential phase. Each square on the heat map represents a biological replicate. B) Pathway map showing identified genes involved in lignin model compound funneling pathways in *R. opacus*. Gene codes are from the NCBI reference sequence NZ_CP003949.1.

To identify how these same aromatics were utilized in the β-ketoadipate pathway (Fig 9B), the main aromatic degradation pathway in *R. opacus*, we again investigated our transcriptomic datasets (Fig 9A). We identified three primary β-ketoadipate pathway gene clusters. Similar to the funneling pathways, there were similar changes in expression between WT and PVHG6. We did, however, note that while the catechol degradation pathway goes from being off to on in the presence of relevant aromatics, the PCA degradation pathway is always lowly expressed. Additionally, we identified several ABC transporters that might control the uptake or export of aromatic compounds. We also identified two acetyl-CoA acetyltransferases, upregulated in PVHG6, that showed high positive amino acid sequence similarity to 3-oxoadipyl-CoA thiolases in *R. jostii* RHA1. The 3-oxoadipyl-CoA thiolases catalyze the last step in the β-ketoadipate pathway and upregulation might increase metabolic flux through the β-ketoadipate pathway, increasing aromatic degradation in PVHG6.

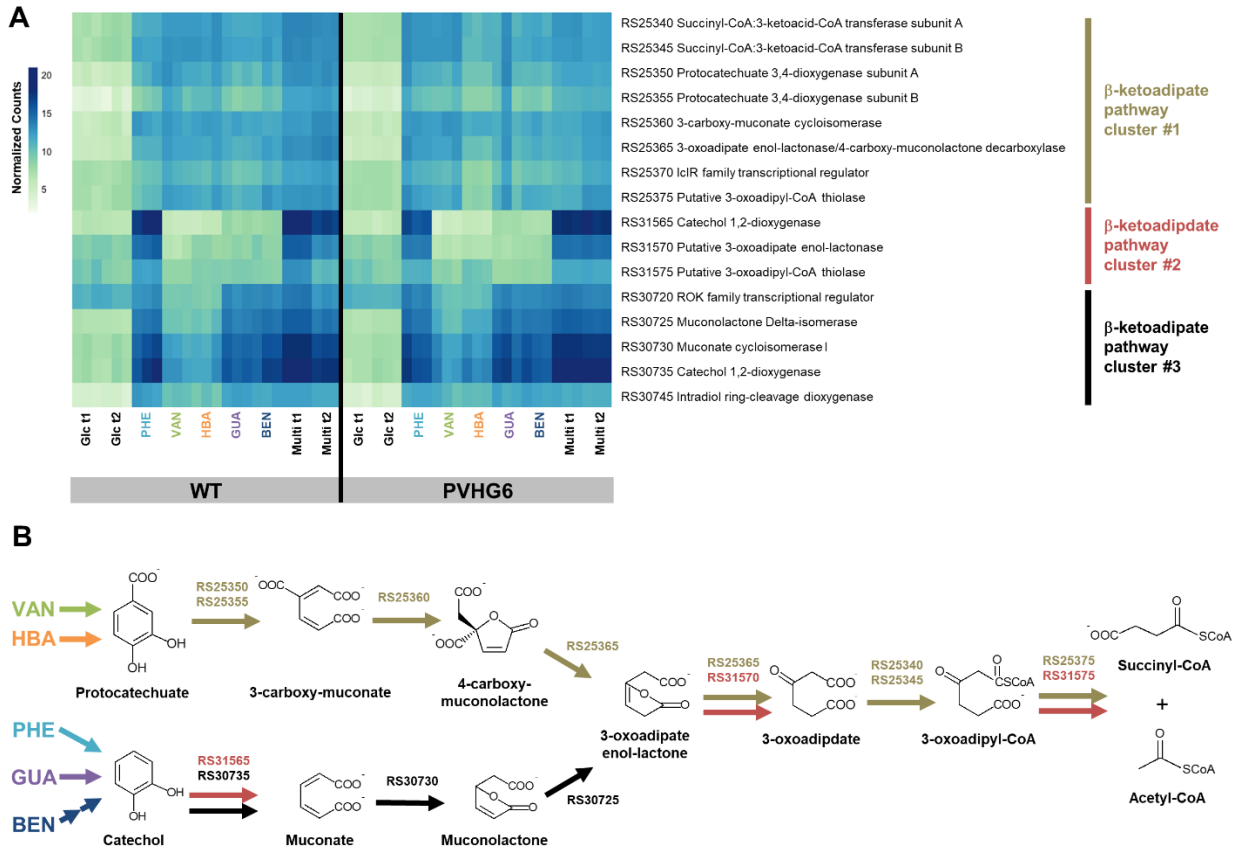


Figure 9. Activation of β-ketoadipate pathway gene clusters by lignin model compounds in *R. opacus*[Pub6]. A) Heat map of normalized gene expression counts for β-ketoadipate pathway gene clusters during growth on individual lignin model compounds, a mixture of lignin model compounds, and glucose. Raw counts were transformed using the variance stabilizing transformation in DESeq2. Darker colors indicate higher normalized counts (see scale bar). Glc =1 g/L glucose, PHE=0.5 g/L phenol, VAN =0.5 g/L vanillate, HBA =0.5 g/L 4-hydroxbenzoate, GUA =0.5 g/L guaiacol, BEN =0.5 g/L sodium benzoate, and Multi =0.5 g/L of PHE, VAN, HBA, GUA, and BEN (2.5 g/L total aromatics). t1 represents cells harvested at early exponential phase, while t2 represents cells harvested at mid-exponential phase. Samples from individual compound growth conditions were harvested at mid-exponential phase. Each heat map square represents a biological replicate. B) Pathway map showing genes involved in β-ketoadipate pathway gene clusters in *R. opacus*. Gene codes are from the NCBI reference sequence NZ_CP003949.1.

We also investigated the effects of phenol on genes involved with lipid biosynthesis in WT and our phenol adapted strains (evol33 and evol40) from our original pilot adaption experiment (Table 1). We found that TadA (LPD06283), a heparin-binding hemagglutinin-like protein required for normal lipid droplet morphology, was upregulated in all strains when grown on phenol. This suggests that lipid droplet morphology could be altered in the presence of a phenolic compound. Additionally, a fatty acid synthesis gene (FASI; LPD05549) and several stearoyl-CoA 9-desaturase genes (LPD00131, LPD03000, LPD05528, etc.[Pub6]) were upregulated in phenol for both WT and adapted strains. WT and adapted strains showed the same general patterns of upregulation or downregulation for most lipid biosynthesis genes, although the degree of fold change differed significantly, which could have implications for the final cellular lipid composition (Table 1).

Gene ID	Annotation / Strain:conditions	WT LowP	evol33 LowP	evol33 HighP	evol40 LowP	evol40 HighP
Lipid biosynthesis/metabolism						
LPD06283	Heparin-binding hemagglutinin (TadA)	4.67	4.76	3.77	4.58	3.14
LPD05549	Putative sterigmatocystin biosynthesis fatty acid synthase subunit alpha	6.01	5.04	5.84	5.04	5.82
LPD00131	Stearoyl-CoA 9-desaturase	11.80	5.10	6.66	6.52	6.24
LPD00130	Long-chain specific acyl-CoA dehydrogenase, mitochondrial	8.06	2.65	5.14	3.79	3.79
LPD03000	Stearoyl-CoA 9-desaturase	7.24	5.15	5.67	5.48	5.85
LPD05528	Stearoyl-CoA 9-desaturase	6.18	4.52	7.55	5.03	6.31
LPD05529	Stearoyl-CoA 9-desaturase electron transfer partner	5.80	4.23	6.93	4.79	5.85
LPD01195	Long-chain-fatty-acid--CoA ligase FadD19	−5.92	−2.13	−2.98	−4.41	−5.03
LPD06540	Long-chain specific acyl-CoA dehydrogenase, mitochondrial	−3.82	−2.20	−1.27	−2.50	−2.72
LPD04435	Long-chain-fatty-acid--CoA ligase	−2.34	−2.71	−3.93	−2.66	−3.51
LPD03408	Long-chain specific acyl-CoA dehydrogenase, mitochondrial	−3.76	−2.88	−2.76	−3.12	−3.06
LPD07841	Long-chain-fatty-acid--CoA ligase	−4.16	−3.66	−3.73	−3.96	−2.94
LPD03067	Short/branched chain specific acyl-CoA dehydrogenase, mitochondrial	−3.08	−3.71	−2.96	−3.10	−2.72
LPD03066	Short-chain specific acyl-CoA dehydrogenase, mitochondrial	−3.36	−3.78	−3.19	−3.35	−3.26

Table 1. Summarized log2 fold changes of selected lipid biosynthesis genes (over glucose condition) [Pub6]**.** The wild type (WT) and two phenol adapted mutants (evol33 and evol40) were grown in both glucose, low phenol (LowP), and high phenol (HighP). The log2 fold change over the glucose condition, either positive or negative, is given for the most highly up- or downregulated genes related to lipid biosynthesis and metabolism.

Aim 2. Interrogate phenolic tolerance mechanisms in *R. opacus* via high-throughput gain-of-function and loss-of-function screening.

Aim 2.1: Genomic library screening for genes involved in improved tolerance to phenolic compounds and increased TAG production.

A fast mechanism that an organism can utilize to adapt to a new condition is altering the expression level of different genes in its genome. For instance, we hypothesized that *R. opacus* would increase the expression level of its aromatic degradation enzymes and transporters in response to a phenolic compound, which it does in fact do (Fig 10). When fed phenol, *R. opacus* highly upregulates a number of genes including phenol hydroxylases, genes within the β-ketoadipate pathway, and transporters. However, constitutive overexpression of several of these genes on a heterologous plasmid did not seem to have any affect on the phenotype of the engineered strain. As the improved aromatic tolerance phenotype of the adapted strains likely consists of changes in expression of many genes spread across the genome, it was unlikely that the expression of a single gene, as is the goal of an expression library of small pieces of the genome, would lead to any relevant discoveries.

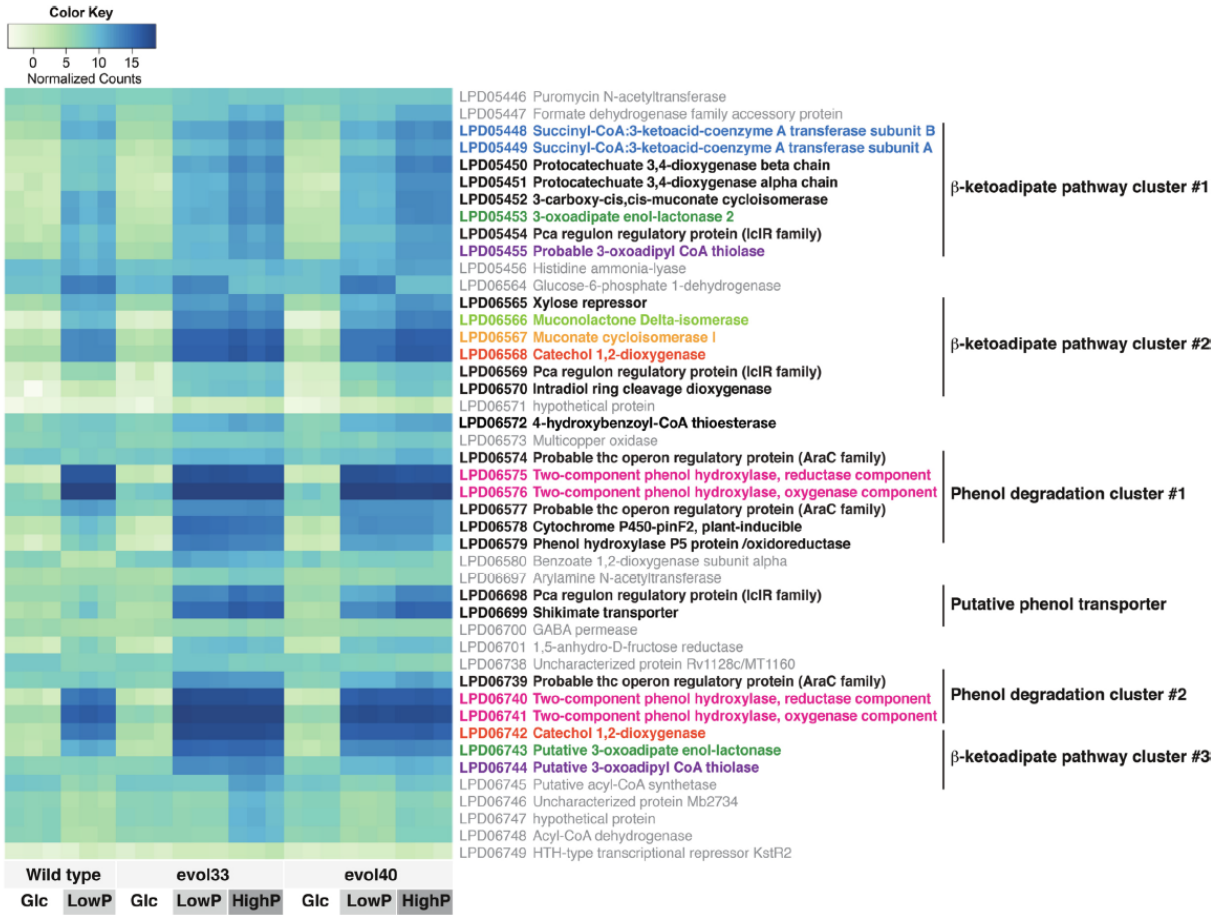


Figure 10. **Transcriptomic and genomic information of genes involved in phenol degradation and utilization**[Pub6]. Differential expression is shown for the phenol responsive degradation operons, β-ketoadipate pathway gene clusters, a putative phenol transporter gene and transcriptional regulator genes. Raw counts were normalized using variance stabilizing transformation in DESeq2 to fit in the range of 20 across all genes. Darker colors indicate higher normalized counts (as shown in Color Key). Glc, 1 g/l glucose; LowP, 0.75 g/l phenol; and HighP, 1.5 g/l phenol. Each square on the heat map represents a biological replicate.

Alternatively, we could investigate the role of both upregulated and downregulated genes through gene expression knockdown studies. To enable this kind of study, we optimized CRISPR interference (CRISPRi) for *R. opacus*. This first required the development of several basic genetic parts discussed in Aim 2.2 and 3.3, including a strong constitutive promoter, an inducible promoter (pBAD), additional antibiotic resistance markers, and a genome modification system for integration of a fluorescent reporter expression cassette integrated into the genome to simulate a native gene. Once these requirements were met, we could begin investigating different dCas9 proteins (a catalytically inactivated Cas9 nuclease that binds DNA and inhibits transcription) for gene repression efficiencies. The commonly used dCas9 from the AT-rich organism *Streptococcus pyogenes* was found to have no appreciable activity in *R. opacus*. Thus, we turned to alternate homologs of dCas9 and found that the inactivated Cas9 from *Streptococcus thermophilus* demonstrated appreciable levels of inducible repression up to ~50% (Fig 11).

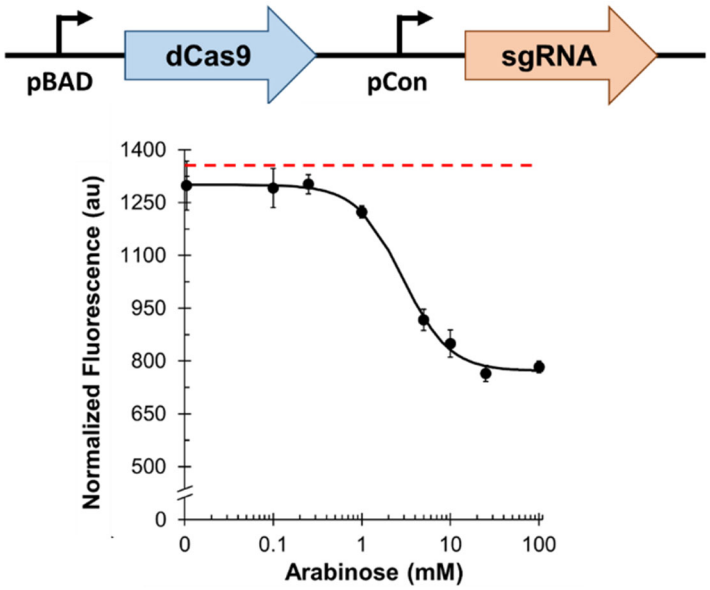


Figure 11. Tunable repression with CRISPRi in *R. opacus*[Pub4]. A codon-optimized version of dCas9 from *S. thermophilus* (dCas9Sth1) was placed under the control of the arabinose-inducible pBAD promoter. The strong pConstitutive promoter drives sgRNA transcription. The sgRNA was designed to target a constitutively expressed EYFP gene that was integrated using our genome modification platform into one of the neutral sites we identified in the chromosome. The y-axis represents normalized fluorescence at early stationary phase in the presence of 0, 0.1, 0.25, 1, 5, 10, 25, and 100 mM arabinose. The red dashed lines represent the positive control (EYFP only) value. Values are averages of three replicates and error bars represent one standard deviation.

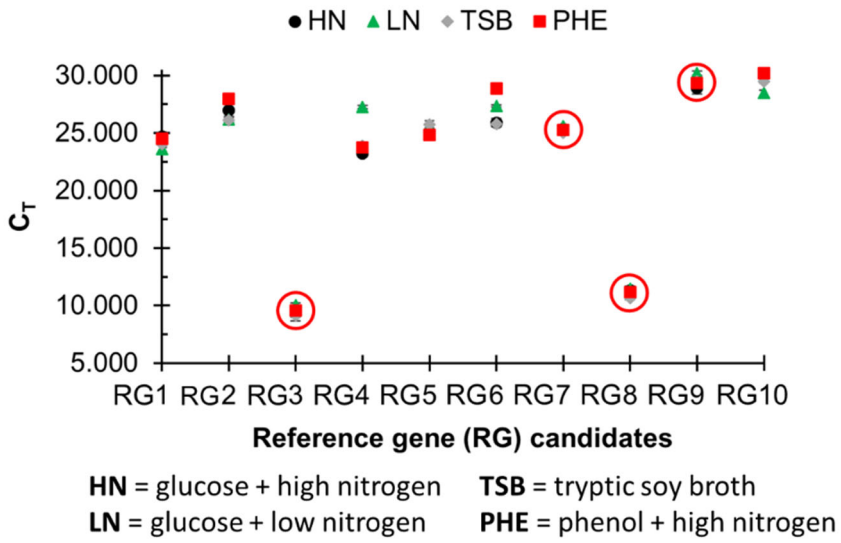


Figure 12. Cycle threshold (C_T) values for ten candidate reference genes in *R. opacus*[Pub3]. RNA was extracted from biological triplicates of *R. opacus* grown in four distinct growth conditions (HN [circle], LN [triangle], TSB [diamond], and PHE [square]. RT-qPCR was performed in technical triplicate on each biological replicate. Each point represents the average threshold cycle (C_T) value of all replicates for the listed gene and growth condition. Error bars represent one standard deviation. Red circles designate genes whose expression was identified as stable. RG3: PD630_ RS03840 23S rRNA, RG7: PD630_RS25530; ATP-dependent Clp protease ATP-binding subunit ClpX, RG8: PD630_RS01395; 16S rRNA, and RG9: PD630_RS37755; rRNA small subunit methyltransferase G.

Once a tool capable of targeted gene repression had been developed, the next step was to optimize a process for quickly quantifying native gene repression, as it does not tend to generate easily quantified phenotype changes like fluorescence reduction. One method previously discussed is RNA-Seq, but this quantifies the entire transcriptome and entails a high cost and time burden to acquire and process the data. An alternative approach to examine gene expression is reverse transcription quantitative PCR (RT-qPCR), which quantifies the amount of mRNA for a particular gene within the cell. A requirement for RT-qPCR is that you must have sets of reference genes that are stably expressed in order to normalize your data between samples. No such genes had been previously identified in *R. opacus*. Leveraging our transcriptomic data sets, we selected 10 candidate reference genes that were stable under the conditions tested. We then examined these candidate genes under four distinct growth conditions to test their stability (Fig 12). Several of these genes were confirmed as being stably expressed, and several statistical analyses were performed to identify pairs of these genes that could be used for data normalization when ribosomal RNA (rRNA) is either present or depleted.

Due to the length of time it took to develop the underlying technologies necessary for targeted gene repression and quantification of that repression, we were not able to examine the effect of the repression of native genes on *R. opacus* during this grant period. However, several genes were examined in Aim 2.2 via gene knockouts. Furthermore, CRISPRi is being employed in our follow-up project being funded by DOE for metabolic flux modification.

Aim 2.2: Random transposon mutagenesis and screening for genes involved in increased fitness in phenolic compound-added media by INSeq.

Originally, we proposed to perform a random transposon mutagenesis experiment, wherein a genetic element randomly inserts itself into the genome, leading to localized gene disruptions. However, we reassessed the value of this type of experiment after a deep analysis of the *R. opacus* genome annotation, which we improved by merging several different versions, found that many genes related to aromatic tolerance and lipid metabolism had multiple copies scattered across the genome. As random transposon mutagenesis only affects one region at a time, this technique would not allow us to alter all copies of a gene and thus reduce the likelihood of observing an effect. Furthermore, as discussed previously, the *R. opacus* transcriptome underwent vast changes between a sugar and phenolic feedstock, making a randomized deletion assay less likely to lead to insightful targets due to many small synergistic changes occurring throughout the genome.

To instead investigate gene targets identified by our omics analyses, we developed a site-specific genome editing platform for *R. opacus* based on two bacteriophage recombinases. These recombinases (Che9c60 and Che9c61) are GC-rich homologs of the RecT and RecE recombinases often used for genome recombineering and allow for a DNA donor template to undergo homologous recombination with the genome (Fig 13). This allows for both DNA deletion (i.e. gene knockouts) and gene insertions. For Aim 2.1, we used this genome modification platform to integrate a fluorescent reporter cassette into one of three neutral sites we identified in the genome using our transcriptomic data and knowledge of the *R. opacus* genomic architecture.

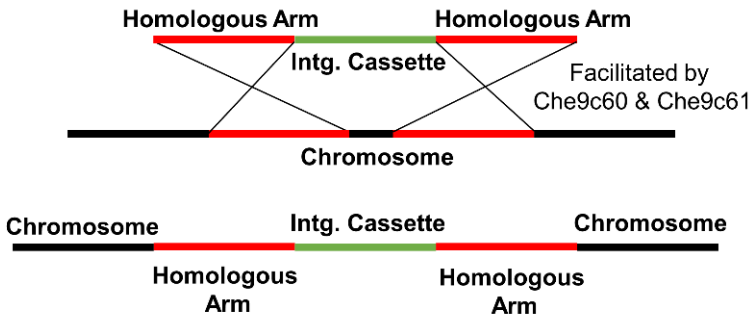


Figure 13. Schematic demonstrating methodology of genome modification platform[Pub4]. Double homologous recombination of a linearized double-stranded integration cassette (Intg. Cassette) is facilitated by two heterologous recombinases Che9c60 and Che9c61. Platform allows for gene knockouts and gene insertions.

The genome modification platform was applied to investigate three aromatic transporters and two core aromatic degradation pathways in *R. opacus* (Fig 14A). Based on our transcriptomic datasets, we identified that a shikimate transporter (ShikiT) was upregulated in the phenol growth conditions, while two additional transporters were upregulated in the benzoate and vanillate growth conditions (named BenT and VanT, respectively). We first used our tool to individually knockout these three transporters and then grow *R. opacus* on five individual lignin model compounds and a mixture of these compounds (Fig 14B). Each transporter knockout led to differential growth patterns, indicating that each of these proteins plays a unique role with preferential translocation of distinct aromatic monomers. ΔShikiT demonstrated significantly reduced growth on phenol and the mixture of aromatics. ΔVanT had a reduction in growth on PHE, VAN, and the mixture. Finally, ΔBenT was had reduced growth on several compounds, including PHE, VAN, GUA, BEN, and the mixture. None of these knockouts influenced growth on HBA, indicating that either its transporter is still yet to be found, or a number of these transporters can import it equally well.

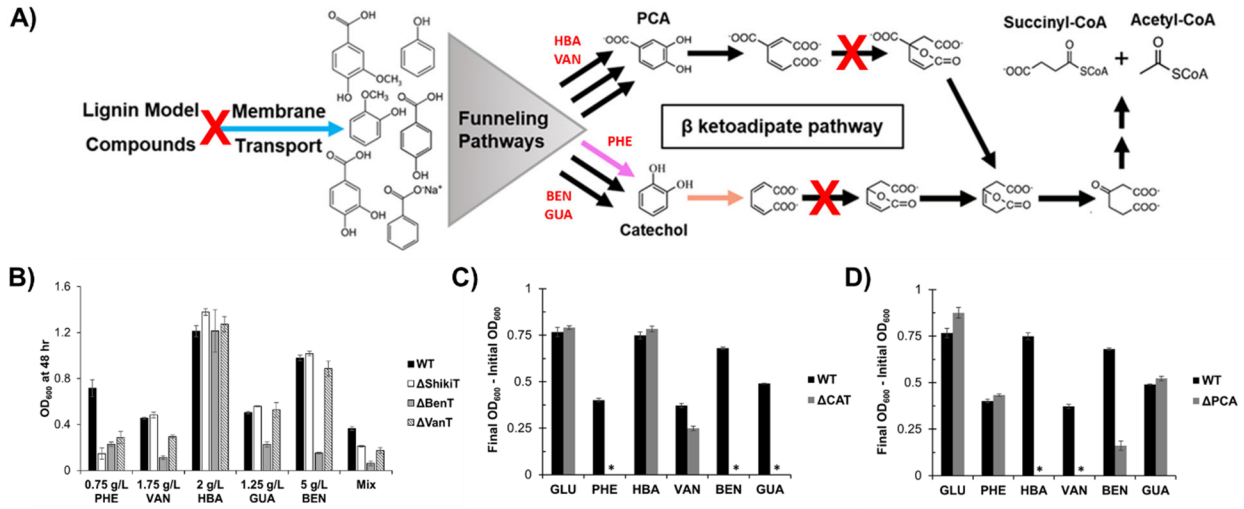


Figure 14. Gene knockouts of upregulated aromatic degradation genes and transporters in *R. opacus*[Pub2]. A) Schematic of aromatic degradation in *R. opacus*. Red x's represent gene knockouts (B-D). B) The final OD_{600} of wild type (WT) and transporter knockout strains after growth on the indicated compounds as a sole carbon source. B) Difference between initial OD_{600} and final

OD_{600} after 24 hours of growth for the knockout of the catechol branch of the β-ketoadipate pathway (ΔCAT) strain and the WT strain using different carbon sources. C) Difference between initial OD600 and final OD_{600} after 24 h of growth for the knockout of the protocatechuate branch of the β-ketoadipate pathway (ΔPCA) strain and the WT strain using different sole carbon sources. Asterisk (*) indicates that the difference between initial and final OD_{600} values is less than one standard deviation. Bars represent the average of three biological replicates and error bars represent one standard deviation. PHE = phenol, VAN = vanillate, HBA = 4-hydroxybenzoate, GUA = guaiacol, BEN = sodium benzoate, and Mix = 0.5 g/L of PHE, VAN, HBA, GUA and BEN (2.5 g/L total aromatics).

The degradation route of the five model lignin breakdown products (PHE, VAN, GUA, BEN, an HBA) through the two parallel branches of the β-ketoadipate pathway was investigated by knocking out each branch individually (Fig 14C and D). The catechol branch knockout (ΔCAT) eliminated the ability of *R. opacus* to grow on PHE, BEN, and GUA. Conversely, the protocatechuate branch knockout (ΔPCA) eliminated the ability of *R. opacus* to grow on HBA and VAN. These types of findings, knowing which metabolic intermediates each of these aromatic monomers is converted to, can provide insight into the chemical mechanisms of the upstream aromatic funneling pathways we identified in Aim 1.3.

Aim 2.3: Characterization of evolved strains (Aim 1) and strains with introduced variation (Aim 2) using targeted metabolomic profiling and high-/mid-throughput analysis of *R. opacus* TAG production.

To measure phenolic compound consumption by *R. opacus*, we first optimized a spectrophotometric assay to determine the concentration of a single aromatic compound using absorbance at 280 nm. Using this method, we examined the growth and consumption rates of WT and our best multi-compound adapted strain (PVHG6) on five model lignin breakdown products (PHE, VAN, HBA, GUA, and BEN; Fig 15). The initial concentrations of each compound were chosen based on the half maximal inhibitory concentrations (IC_{50}) of WT, which were determined by conducting a growth assays with a range of concentrations of each compound. Interestingly, despite PVHG6 demonstrating improved growth on a mixture of all 5 of these aromatic compounds, it only showed improved growth compared to WT on PHE and GUA, but reduced growth on HBA. Similarly, PVHG6 consumption of PHE and GUA increased, but it had decreased consumption of HBA relative to WT (Fig. 15).

To determine the concentrations of individual aromatics within a mixture of aromatics, we optimized a methyl chloroformate derivatization and GC-MS based protocol to quantify each compound independently of the others. On a mixture of PHE, VAN, HBA, GUA, and BEN, we found that PVHG6 demonstrated a faster and more complete consumption of the compounds relative to WT, but that both strains consumed the compounds in the same order. BEN and HBA were consumed first, followed by PHE and VAN, and finally GUA (Fig 16). Little is known about inhibition between aromatic pathways and further study could lead to engineering efforts to allow co-consumption.

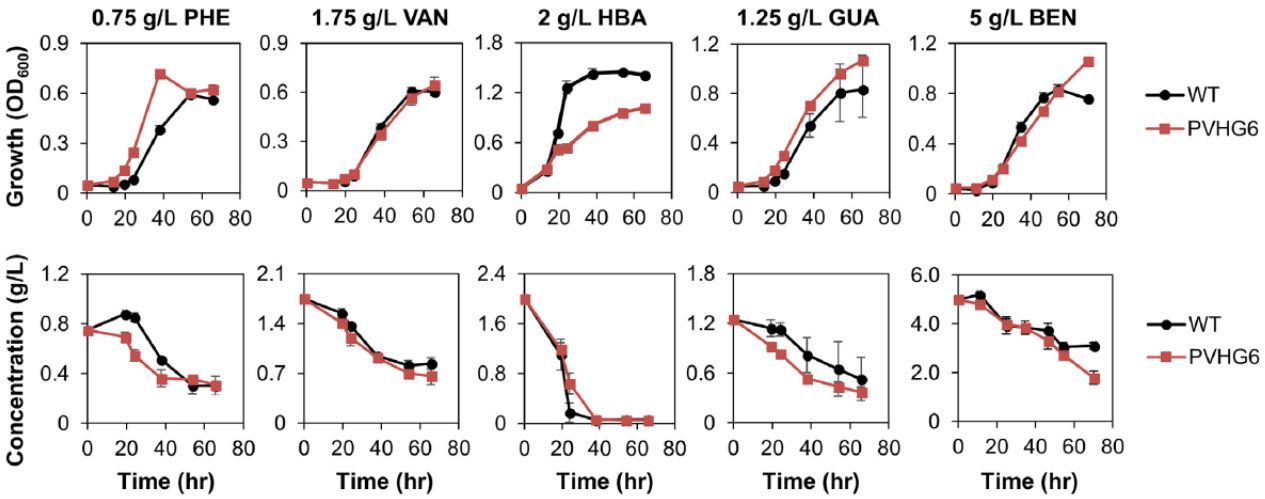


Figure 15. Growth and consumption of lignin model compounds by WT and adapted strains[Pub2]. Aromatic concentrations were estimated using UV absorbance based on our optimized spectrophotometric assay. For all plots, points represent the average of three biological replicates and error bars represent one standard deviation. PHE =phenol, VAN =vanillate, HBA =4-hydroxybenzoate, GUA =guaiacol, and BEN =sodium benzoate.

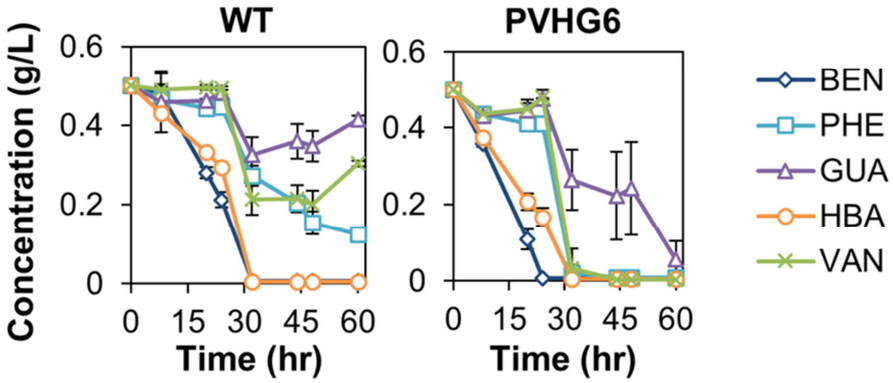


Figure 16. Aromatic consumption profiles of wild type (WT) and an aromatic adapted strain (PVHG6) [Pub6]. Strains were grown in a mixture of 0.5 g/L sodium benzoate (BEN), phenol (PHE), guaiacol (GUA), 4-hydroxybenzoic acid (HBA), and vanillate (VAN). Aromatic concentrations were measured by derivatizing culture supernatants using methyl chloroformate followed by GC-MS-FID.

To gain a deeper understanding of *R. opacus* lipid accumulation, we focused on analyzing the various lipid moieties of the WT and the phenol adapted strain (evol40) under different growth conditions. Utilizing liquid chromatography–mass spectrometry (LC-MS) and tandem MS (MS^n), we created a lipidomic library of >100 TAGs, phospholipids, and mycolic acids in WT *R. opacus*. We decided to examine several major classes of lipid species, rather than just TAGs, due to the cell membrane's structure and composition, which includes many non-TAG lipids, being very important for microbial stress tolerance. The cell membrane acts as a permeability barrier for the influx and efflux of different compounds, such as phenolics, and microbes respond to stressful growth conditions by modifying the structure of their cell membrane. Using our extensive lipidomic library, we performed several mass spectrometry analyses to identify how lipid species in our samples were changing.

To examine whether the lipidome of WT and evol40 changed at a global level and by lipid class, principal component analysis was performed. The global lipidome and each of the primary lipid classes (mycolic acids, triacylglycerols, and phospholipids) of each strain (WT and evol40) under each growth condition (glucose, low phenol, and high phenol) were examined (Fig 17). The lipid profiles of WT and evol40 generally cluster together when grown on glucose. However, there are large separations that occur between the two strains when grown in the low and high phenol growth conditions. This preliminary analysis demonstrates that phenol adaptation did alter the lipidome in *R. opacus* and that further study of these three different lipid classes is warranted.

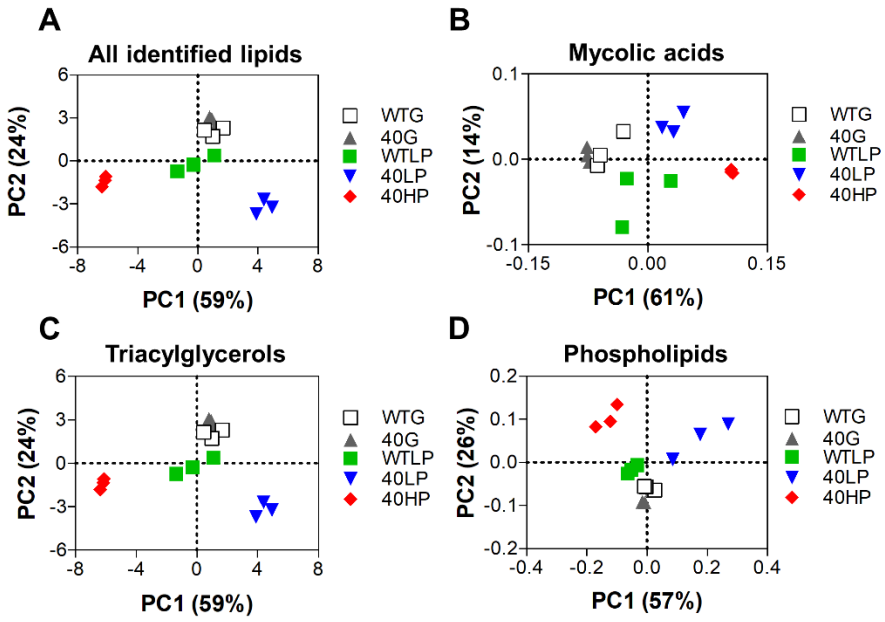


Figure 17. Principal component analysis of lipid types identified using LC/MS[Pub1]. A) Principal component analysis (PCA) of all identified lipids. B) PCA of mycolic acid species. C) PCA of triacylglycerol species. D) PCA of phospholipid species. Because triacylglycerol ion counts are >98% of total ion counts in each sample, the PCA plots for all lipid species (2A) and triacylglycerol species (2C) are almost identical. Each point represents one replicate. WTG = WT strain grown in 1 g/L glucose. 40G = evol40 grown in 1 g/L glucose. WTLP = WT strain grown in 0.75 g/L phenol. 40LP = evol40 strain grown in 0.75 g/L phenol. 40HP = evol40 strain grown in 1.5 g/L phenol. PC1 = 1st principal component; PC2 = 2nd principal component. Percent represents the amount of variance explained by each principal component.

One of the central themes of our proposal was the hypothesis that TAG production and composition was critical to phenolic tolerance. We investigated how the composition of TAGs changed based on whether WT or evol40 was grown on glucose, low phenol, or high phenol. We found that both strains demonstrated an increase in TAG saturation when phenol concentration increased, with a decreased prevalence of TAGs with 2 double bonds and an increase in TAGs with 0 or 1 double bond (Fig 18). Interestingly, evol40 at the high phenol growth condition demonstrated an even stronger trend towards TAG saturation. The carbon chain length of the TAGs significantly changed between conditions, but there was no consistent trend.

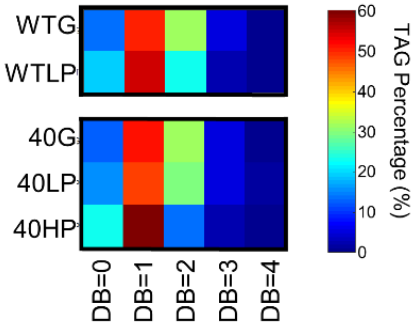


Figure 18. Triacylglycerol composition of WT and evol40 *R. opacus* strains using glucose and phenol as sole carbon sources[Pub1]. Distribution of TAG double bond (DB) numbers in the WT and evol40 strains in each growth condition. The DB number represents the total number of double bonds and cyclopropane units on acyl chains. Values in the heat map are the average of three replicates. See color bar for scale. WTG = WT strain grown in 1 g/L glucose. 40G = evol40 grown in 1 g/L glucose. WTLP = WT strain grown in 0.75 g/L phenol. 40LP = evol40 strain grown in 0.75 g/L phenol. 40HP = evol40 strain grown in 1.5 g/L phenol.

The inner cell membrane consists of phospholipids in *R. opacus*, and some bacteria are known to modify their phospholipid head groups and the number and type of double bonds in their chains in response to aromatic compounds. We compared the ratio of the two dominant phospholipid species in *R. opacus*, phosphatidylinositol (PI) and phosphatidylethanolamine (PE), between the glucose and phenol growth conditions. We observed that growth on phenol increased the ratio of PI to PE via an increase in PI production and led to an increase in phospholipid (PL) saturation in both WT and evol40, although to a much higher degree in evol40 at the high phenol condition (Fig 19). This increase in saturation when *R. opacus* is grown on phenol mirrors the shift observed in TAGs. The significant increase in PI species is a previously unknown response to phenolics and could represent a novel adaptation strategy.

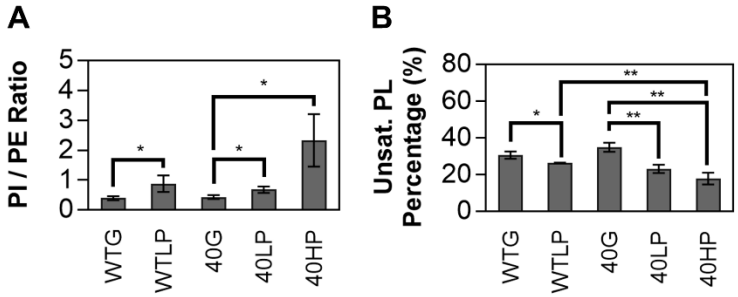


Figure 19. Phospholipid composition of *R. opacus* strains using glucose and phenol as sole carbon sources[Pub1]. A. Ratio of phosphatidylinositol (PI) to phosphatidylethanolamine (PE) total ion counts in the WT and evol40 strains using glucose and phenol as sole carbon sources. Bars represent the average of three replicates, error bars represent one standard deviation. WTG = WT strain grown in 1 g/L glucose (glucose). 40G = evol40 grown in glucose. WTLP = WT strain grown in 0.75 g/L phenol (low phenol). 40LP = evol40 strain grown in low phenol. 40HP = evol40 strain grown in 1.5 g/L phenol (high phenol).

R. opacus also produces an outer mycomembrane composed of mycolic acids, which has been shown to change in other actinomycetes in response to aromatic compounds. Similar to both the TAGs and phospholipids, the saturation of mycolic acids increased in both the WT and evol40 strains in the low phenol condition relative to the glucose condition (Fig 20A). However, mycolic

acid saturation increased even more significantly in evol40 in the high phenol growth condition relative to the glucose growth condition (Fig 20A). In the presence of phenol, there was a decrease in the abundance of mycolic acids with 3 double bonds and an increase in species with 1 or 2 double bonds. We also found that the carbon chain length (C number) of mycolic acids decreased in the adapted strain when grown on phenol compared to glucose, but that WT did not demonstrate a significant shift when compared to evol40 (Fig 20B).

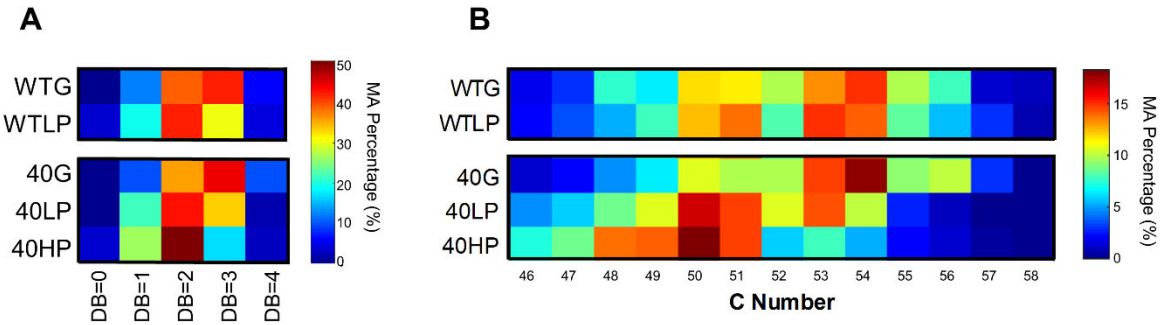


Figure 20. Mycolic acid composition of *R. opacus* strains using glucose and phenol as sole carbon sources[Pub1]. A) Heat map of double bond (DB) numbers for mycolic acids (MAs) in the WT and evol40 strains. The values shown in the heat map are the average of three replicates. See color bar for scale. The DB number represents the total number of double bonds and cyclopropane units on all acyl chains. B) Heat map of MA carbon (C) number distribution, defined as the total number of carbons located on acyl chains. The values shown in the heat map are the average of three replicates. See color bar for scale. WTG = WT strain grown in 1 g/L glucose (glucose). 40G = evol40 grown in glucose. WTLP = WT strain grown in 0.75 g/L phenol (low phenol). 40LP = evol40 strain grown in low phenol. 40HP = evol40 strain grown in 1.5 g/L phenol (high phenol). MA percentage is defined as the total ion counts of each category (DB number or C number) divided by the total ion counts of all mycolic acids detected in each sample.

Taken together, we observed a general increase in saturation of lipid species, an increase in the amount of total lipids (particularly PI), and a decrease in mycolic acid chain length when *R. opacus* was grown on phenol compared to glucose. Overall, this work represents the first lipidomic study of membrane and TAG lipids in *R. opacus* using phenol as a sole carbon source. These results suggest that the lipid metabolism of *R. opacus* is related to phenol tolerance by affecting the mycomembrane and phospholipid membrane composition and properties of TAG species during growth using phenol.

Aim 3. Identify preferred phenolic substrates in *R. opacus* from thermochemically depolymerized lignin and the governing metabolic regulation by combining subtractive metabolic profiling and transcriptomics.

Aim 3.1: Carbon source/feedstock generation and chemical profiling.

The primary underlying goal of this project was to feed actual lignin to *R. opacus* in order to produce bioproducts, in our case lipids. To facilitate this, we optimized the reaction conditions of a thermocatalytic lignin depolymerization process to generate higher concentrations of the resulting aromatic compounds (Fig 21). The addition of a palladium on carbon (Pd/C) catalyst drastically improved the yields of three of the main aromatic monomers under all temperature and pH conditions, while improving yield for two other monomers under some scenarios. Furthermore,

a higher temperature of 170 ºC versus 140 ºC, and a higher pH of 12 versus 9 led to a higher production of some aromatic monomers.

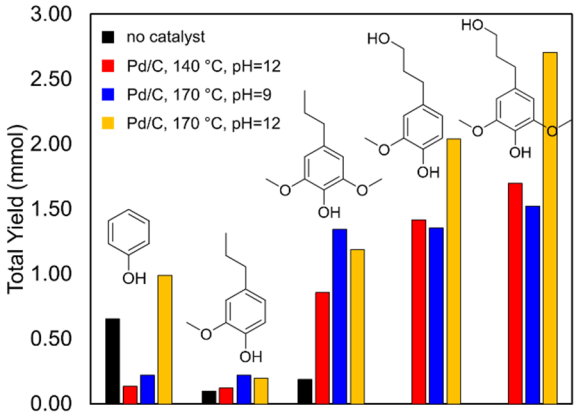


Figure 21. Effect of Pd/C catalyst and reaction conditions on aromatic monomer yield from poplar biomass. Poplar lignin was depolymerized at two temperatures (140 and 170 ºC) and at two pH (9 and 12) in the absence and presence of a Pd/C catalyst. Yields (mmol of carbon) were calculated using total organic carbon analysis of feedstock lignin and a GC-MS/methanizer-FID to identify and quantitatively profile specific compounds in the depolymerized products of actual lignin.

An important question though is whether *R. opacus* can actually utilize the aromatic monomers generated by our thermocatalytic depolymerization process. Using three distinct biomass sources (switchgrass, poplar, and pine), we further optimized the extraction process and thermochemical pretreatment to create lignin breakdown product mixtures that could be readily consumed by *R. opacus* (Fig 22A). The final process, which involved using the Pd/C catalyst with organosolv extracted lignin at 150 ºC, was demonstrated to be feedstock independent with *R. opacus* growing on all three biomass sources depolymerized using this process (Fig 22B). However, *R. opacus* grew the most rapidly on depolymerized switchgrass (LBP6). Using LBP6, we utilized the previously discussed methyl chloroformate derivatization and GC-MS assay to determine the aromatic compounds present in the mixture before and after *R. opacus* growth (Fig 22C). Based on this analysis, a number of compounds were consumed by *R. opacus*, particularly 4-alkyl phenols and 4-alkyl guaiacols. Optimizing this depolymerization process for *R. opacus* makes it possible to biologically valorize waste lignocellulose feedstocks into renewable products.

A)

Name/ Abbreviation	Lignin Source	Extraction Type	Depolymerization Conditions					
			Lignin Loading (mg)	Solvent	Catalyst	Temp (°C)	Time (hr)	Other Notes
LBP #1	Kraft Lignin	N/A	1000	H_2O (30 mL)	Pd/C 100 mg	150	24	500 psi H_2
LBP #2	Switchgrass	Organosolv (MeOH)	1000	H_2O (30 mL)	Pd/C 100 mg	150	24	500 psi H_2
LBP #3	Switchgrass	Organosolv (MeOH)	1000	3% wt NaOH in H_2O (30 mL)	Pd/C 100 mg	150	12	200 psi H_2, adjusted to pH = 7.5
No DMC LBP	Poplar	Organosolv (MeOH)	100	MeOH (3 mL)	CuPMO 50 mg	300	2.5	Solvent removed at -15 to -20 C
DMC LBP	Poplar	Organosolv (MeOH)	100	2:1 MeOH:DMC (3 mL)	CuPMO 50 mg	300	6	Solvent removed at -15 to -20 C
LBP #4 (oxidative)	Switchgrass	Organosolv (MeOH)	1000	H_2O (30 mL)	Pd/Al_2O_3	160		50 psi O_2
LBP #5 (alkaline pretreatment)	Switchgrass	Organosolv (MeOH)	3000	H_2O (20 mL)	NaOH	160		100 psi N_2, adjusted to pH = 7
LBP #6	Switchgrass Type 2	Organosolv (MeOH)	1000	H_2O (30 mL)	Pd/C 100 mg	150	24	500 psi H_2
LBP #7	Poplar	Organosolv (MeOH)	1000	H_2O (30 mL)	Pd/C 100 mg	150	24	500 psi H_2
LBP #8	Pine	Organosolv (MeOH)	1000	H_2O (30 mL)	Pd/C 100 mg	150	24	500 psi H_2

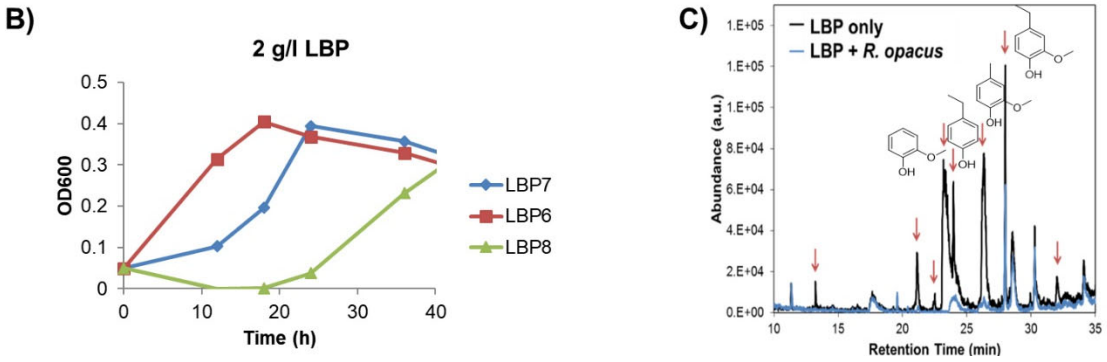


Figure 22. Identification of ideal lignin depolymerization conditions to support growth of *R. opacus*. A) Ten permutations of thermocatalytic reaction conditions and biomass sources were tested. These lignin breakdown products were fed to WT *R. opacus*, and the rows are color coded to signify growth or lack of growth (green = growth, red = no growth). B) Growth curves of *R. opacus* on LBP6-8. C) Subtractive metabolomics showing aromatic monomers consumed by *R. opacus* from LBP6.

Aim 3.2: Combining subtractive metabolic profiling with differential gene expression analysis to link phenolic metabolism with its underlying regulation.

Aromatic degradation pathway expression is very closely tied to the presence or absence of compounds degraded through those pathways. For instance, we previously discussed how each of the five aromatic compounds we tested have funneling pathways (Fig 8) that are only differentially upregulated in their presence. By taking multiple transcriptomic and metabolomic timepoints, we could relate how genes expression changes as a mixture of aromatic monomers was consumed (Fig 8, 9, and 16). As previously discussed, we observed that WT and PVHG6 consume aromatic compounds in a preferred order, with BEN and HBA being depleted first. Using RNA-seq, we identified differential regulation of the benzoate degradation cluster during two different time points of growth on the mixture of all aromatic compounds (before and after complete BEN consumption). Initially, benzoate degradation genes were upregulated when BEN was present, but they were quickly down regulated after benzoate was depleted in PVHG6 (Fig 23).

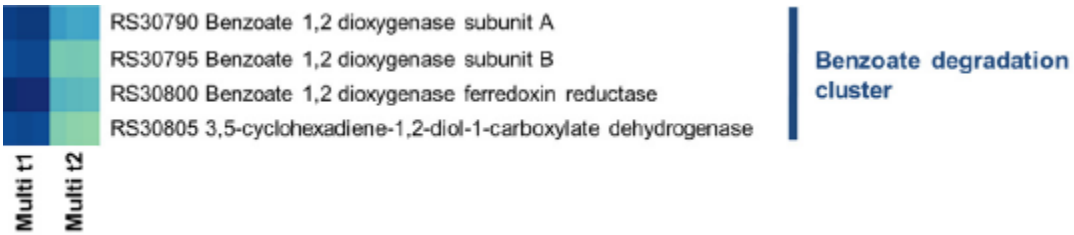


Figure 23. Decrease in benzoate funneling pathway expression after it is consumed from the aromatic mixture by *R. opacus*[Pub2]. (detail from Fig 8).

To further interrogate how *R. opacus* metabolizes a mixture of phenolic and sugar monomers, similar to the actual composition of whole depolymerized lignocellulose, we performed ^{13}C-metabolite fingerprinting (Fig 24). We found that *R. opacus*, unlike other organisms that demonstrate catabolite repression, is capable of co-consuming both glucose and phenol concurrently. This was determined based on the finding that when *R. opacus* was fed fully labeled ^{13}C (U-^{13}C) glucose (1 g/L vs. 5 g/L), the labelling pattern of the examined amino acids did not change when inspected at similar growth stages. This indicates that *R. opacus* is co-consuming both phenol and glucose and suggests that this host would be a good candidate for the simultaneous utilization of complete lignocellulose depolymerized substrates.

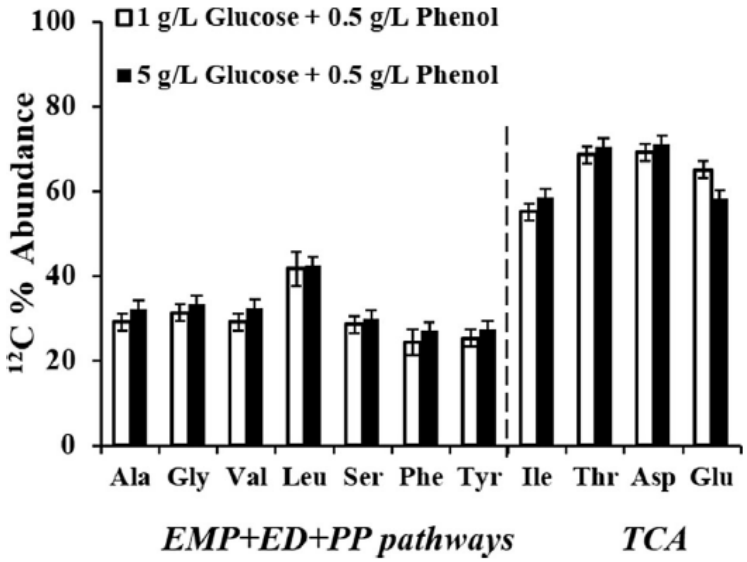


Figure 24. Demonstration of lack of phenolic and sugar monomer catabolite repression in *R. opacus*[Pub7]. The contribution of unlabeled phenol to synthesis of *R. opacus* proteinogenic amino acids was measured using GC-MS. White bar: ^{12}C % abundance from two tracer experiments using 0.5 g/L of phenol with 1 g/L fully labeled ^{13}C (U-^{13}C) glucose. Black bar: 0.5 g/L of phenol with 5 g/L of U-^{13}C glucose. The ^{12}C abundance of amino acids derived from the TCA cycle suggests that much of the carbon from phenol enters the TCA cycle (via the β-ketoadipate pathway). Significant ^{12}C % in amino acids derived from the EMP, the ED, or the PP pathway suggests that carbon from phenol also entered gluconeogenesis. The error bars represent standard deviations from GC-MS measurements.

Aim 3.3: Strain engineering of *R. opacus* to maximize TAG production.

Synthesizing the information gathered from the previous aims, we proposed to engineer *R.*

opacus to increase phenolic catabolism and improve TAG production. Instrumental to this goal is the construction of a more complete genetic toolbox, including fluorescent reporter proteins to tune expression levels, constitutive and inducible promoters for reliable gene expression, and antibiotic resistance markers and characterized plasmid backbones for gene overexpression. These types of fundamental genetic elements are critical to higher order genetic engineering, such as the genome modification platform and CRISPRi systems described in Aim 2.

The most easily measured output from a promoter is a reporter protein, particularly fluorescent proteins. Each organism has a different ideal fluorescent protein due to differences in background fluorescence levels created by native cellular proteins, metabolites, etc. To examine which reporters were best suited for *R. opacus*, we looked at six different proteins spanning three color classes (Fig 25). The fluorescence of CFP (cerulean), RFP (red), mCherry (red), GFP+ (green), sfGFP (green), and EYFP (yellow) were examined relative to the background fluorescence levels of *R. opacus*. The red fluorescent proteins, RFP and mCherry, demonstrated the highest levels of fluorescence over the background.

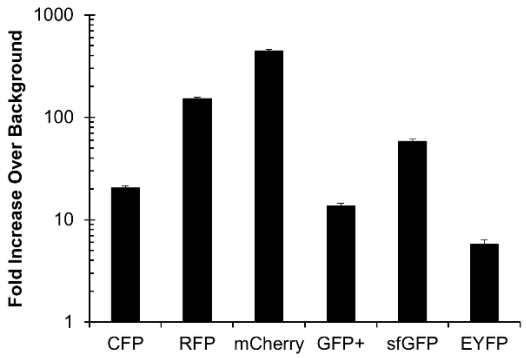


Figure 25. Screen of fluorescent reporter proteins[Pub5]. Six fluorescent reporters expressed under the same promoter and ribosome binding site (RBS) were examined. The fold increase over background represents the absorbance-normalized fluorescence of the reporter strain ($Fluo_{sample}/Abs_{600sample}$) divided by the absorbance-normalized fluorescence of the control strain containing the empty vector ($Fluo_{control}/Abs_{600control}$). Values represent the average of three replicates and error bars represent one standard deviation.

A key component of gene overexpression is a set of well-characterized promoters (Fig 26A). We examined several constitutive promoters to identify the strongest one and then performed saturation mutagenesis on it to create a library of different strength promoters (Fig 26B). This is a crucial asset for engineering multi-gene pathways, because each gene will have a different optimal expression level. However, it can be difficult to determine this ideal expression level using constitutive promoters, because many gene constructs would have to be made with promoters of varying strengths. One solution to this dilemma is chemically inducible promoters, whose expression directly correlates to the concentration of some inducer compound. We demonstrated that three distinct inducible promoters functioned in *R. opacus*, including an arabinose inducible pBAD, an acetamide inducible pAcetamide, and an aTc inducible pTet promoter (Fig 26C). We were able to increase the dynamic range of pTet from 5-fold to 67-fold by performing saturation mutagenesis on the TetR repressor proteins promoter and tune expression for *R. opacus*.

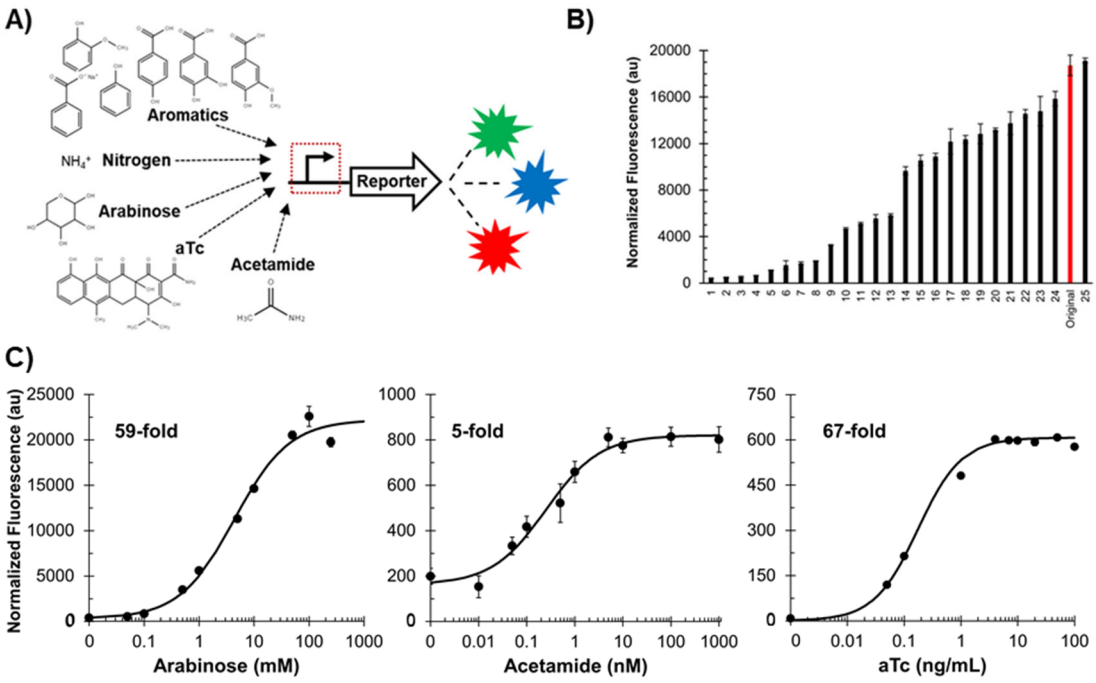


Figure 26. Summary of promoters developed in *R. opacus*[Pub4,6]. A) Three chemically inducible promoters were optimized for *R. opacus*, including pBAD (arabinose inducible), pAcetamide (acetamide inducible), and pTet (aTc inducible). Five metabolite sensors were developed that responds to different aromatic monomers and nitrogen. To facilitate promoter screening, six different fluorescent reporter proteins were screened for best activity in *R. opacus*. B) A constitutive promoter library spanning ~45-fold from the weakest to strongest promoter was created to facilitate a variety of expression levels. C) Transfer curves for pBAD, pAcetamide, and pTet with corresponding fold changes. Bars and points represent the average of triplicate fluorescence values and error bars represent standard deviation.

Another core element required for gene overexpression is either a self-replicating plasmid backbone to harbor DNA constructs or an ability to integrate constructs into the hosts genome. To better characterize the two plasmid backbones previously demonstrated in *R. opacus* (pNG2 and pAL5000), we performed qPCR to determine the number of copies of plasmid per copy of chromosome in *R. opacus* (Fig 4). Furthermore, we optimized a novel origin of replication for expression in *R. opacus* (pB264). pB264 has the advantage of being unstable once antibiotic selection pressure is removed, allowing for rapid curing of the plasmid. In Aim 2.2, we already described the development of a tool to integrate DNA into the *R. opacus* genome. As mentioned, we also identified 3 neutral or non-deleterious locations into which gene expression cassettes can be integrated without affecting *R. opacus* growth. To accommodate the increase in plasmid backbones and genomic integration sites, we developed two additional antibiotic selection markers (chloramphenicol and hygromycin B) for *R. opacus*.

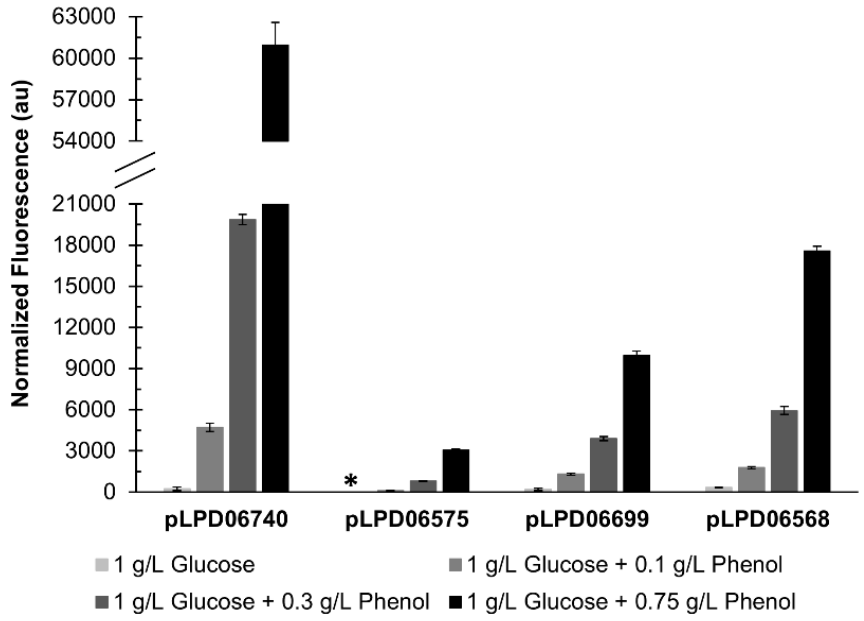


Figure 27. Four phenol sensors characterized in *R. opacus* using GFP^{Pub5}. The strains were grown with 1 g/L glucose and either 0, 0.1, 0.3, or 0.75 g/L phenol. Fold inductions (0.75 g/L phenol vs 0 g/L phenol) were 80-fold (LPD06568), 39-fold (LPD06699), and 247-fold (LPD06740). The fluorescence value of pLPD06575 at 1 g/L glucose was indistinguishable from the background level (*, within one standard deviation) and a fold change could not be calculated. Values represent the average of three replicates and error bars represent one standard deviation.

The original grant also proposed to utilize transcriptomic analysis in the presence or absence of phenolics (see Aim 1.3) to guide the discovery of interesting genetic elements, such as promoters and transcription factors. Using this information, we identified 4 promoters that when placed upstream of GFP demonstrated a dose-dependent response to phenol, with up to 247-fold upregulation in fluorescence (Fig 27). These same promoters were later shown to demonstrate differential response to a number of aromatic monomers, leading to a diverse set of aromatic sensors that could also help provide insight into metabolic flux (Fig 28). Additionally, genetic architectural analysis led to the identification of one-component transcriptional regulators that occur alongside all these aromatic sensors, providing understanding into gene regulation. These tools can eventually be applied to create dynamic genetic circuits, or those that only turn on in the presence of aromatic compounds. This type of approach reduces metabolic burden on the cell and has been demonstrated to improve product yields.

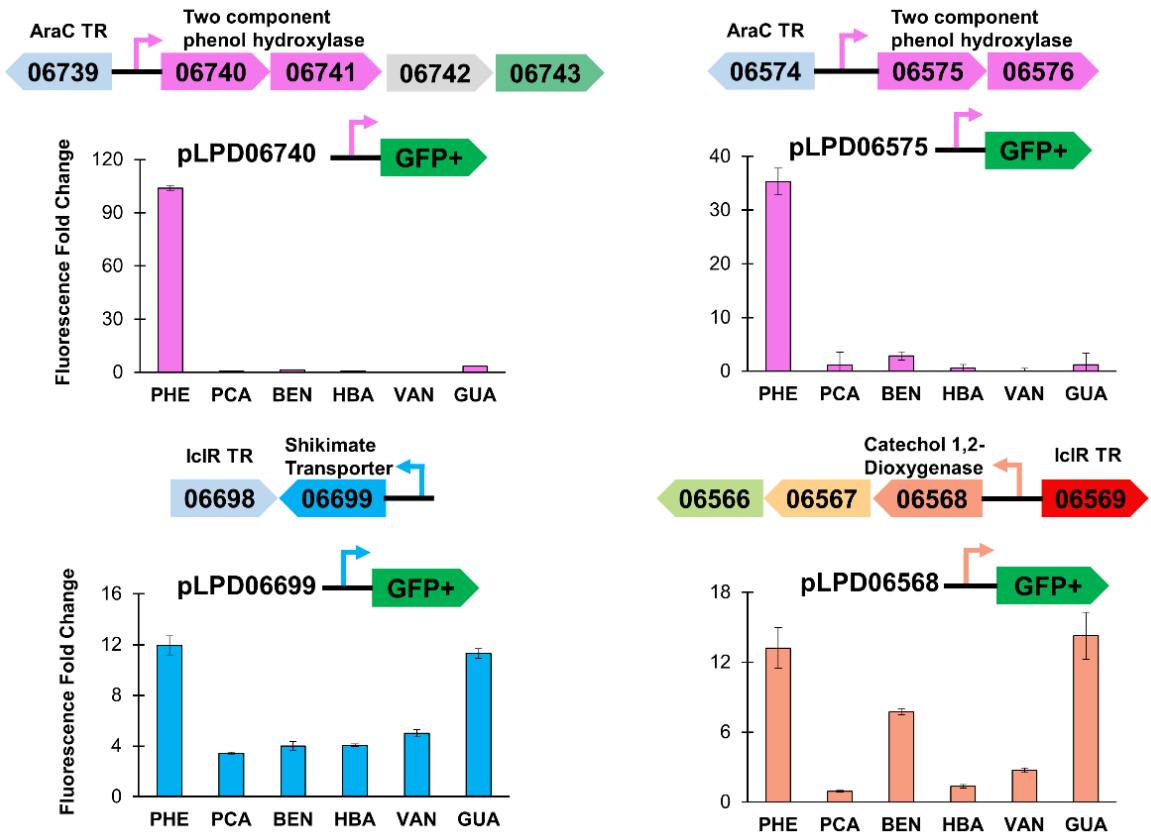


Figure 28. Differential response of native *R. opacus* promoters to distinct aromatic monomers[Pub5]. Native genomic context of aromatic degradation and transporter operons with annotated transcriptional regulators (TR). Promoters are represented as small arrows. Genes are represented as large arrows and are annotated with LPD gene numbers from the NCBI database (Refseq, NZ_CP003949.1). The upstream regions of LPD06568, LPD06575, LPD06699, and LPD06740 were cloned in front of GFP+. All cultures contained 1 g/L glucose (GLU) in addition to any additional aromatic monomers as a carbon source. The fluorescence fold change in the presence of 0.3 g/L phenol (PHE) or an equimolar amount of protocatechuic acid (PCA), sodium benzoate (BEN), 4-hydroxybenzoic acid (HBA), vanillic acid (VAN), or guaiacol (GUA) was determined at early stationary phase relative to the glucose only condition (GLU). Values represent the average of three replicates grown and error bars represent one standard deviation.

Another genetic tool developed based on transcriptomic data was an NH_4^+ (ammonium) sensitive promoter. This promoter was found to be repressed in the presence of NH_4^+, which is used as a nitrogen source by *R. opacus*. When placed upstream of GFP, this promoter demonstrated an 18-fold increase in fluorescence going from its repressed to expression state. We employed this promoter to create a cellular timer designed to activate at specific points in the cellular growth cycle (Fig 29). This was achieved by modulating the initial concentration of NH_4^+ in the growth media. As the NH_4^+ was depleted by *R. opacus*, the promoter would begin to turn on. We found that there was a highly linear correlation ($R^2 = 0.99$) between initial NH_4^+ concentration and the time it took to reach the half maximal expression level. A low-nitrogen condition promoter is also critical for our applications as lipid accumulation occurs under nitrogen limitation. Thus, we could employ this promoter, potentially with the aromatic sensors, to create complex genetic circuits that only turn on under precise growth conditions suited for the conversion of aromatics into lipids.

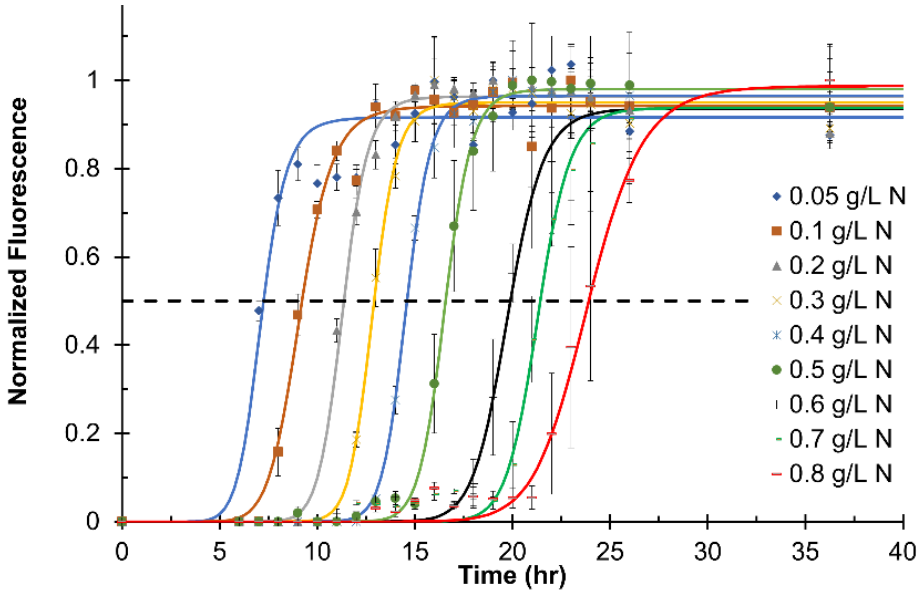


Figure 29. Dynamics of the nitrogen sensitive promoter pLPD03031 at different initial ammonium sulfate concentrations[Pub5]. *R. opacus* cultures containing pLPD03031-GFP were grown with 10 g/L glucose and either 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, or 0.8 g/L ammonium sulfate (N). As the nitrogen source is depleted over time due to cell growth, the pLPD03031 promoter activates and starts expressing GFP. The normalized fluorescence was scaled to the maximum value for each respective condition (scale 0 to 1). The dashed line represents the half-maximal fluorescence value. Values represent the average of three replicates and error bars represent one standard deviation. Continuous lines represent the fitted transfer functions.

Efforts to deploy these genetic tools for improved aromatic catabolism and TAG production were limited due to time constraints caused by having to substantially improve the genetic toolbox originally available to us. As previously mentioned, several aromatic catabolism genes and transporters were constitutively overexpressed in WT *R. opacus*, but no change in phenotype was observed. We hypothesize that this is due to these genes already being expressed at saturation. Future work will focus on utilizing all tools developed during the course of this project (promoters, plasmid, genome engineering, and CRISPRi) to further examine gene targets identified by our genomic, transcriptomic, and lipidomic analyses.

5. Summarize project activities for the entire period of funding, including original hypotheses, approaches used, problems encountered and departure from planned methodology, and an assessment of their impact on the project results. Include, if applicable, facts, figures, analyses, and assumptions used during the life of the project to support the conclusions.

The main goal of our project was to interrogate the metabolic networks and genetic regulation that control the utilization of and tolerance towards lignin-derived, aqueous-soluble phenolics in the triacylglycerol (TAG)-accumulating bacterium *Rhodococcus opacus* PD630. We specifically wanted to test the overarching hypothesis that TAG accumulation and composition is directly related to phenolic tolerance in *R. opacus*, representing a highly advantageous system in which our target molecule's production can improve growth on the inhibitory feedstock. Furthermore, we wanted to examine how different phenolic feedstocks, in both wild type and adapted strains, altered cellular lipid composition. The core component of this project focused on

the use of adaptive evolution, wherein a strong selection pressure in the form of increasing aromatic feedstock concentrations, was applied over many generations. Improved strains were isolated based on faster growth, aromatic tolerance, and TAG production. These adapted mutants were then examined at varying degrees, with some being characterized at the genomic, transcriptomic, metabolomic, and lipidomic levels. To further probe metabolic pathways, tools were developed for gene expression, gene knockouts, and targeted gene repression, in addition to ^{13}C-metabolite fingerprinting. In brief, we successfully adapted *R. opacus* to a multitude of aromatic mixtures representative of real lignin breakdown products, implemented a genetic toolbox for host engineering, performed a multi-omics analysis that provided novel adaptation strategies for microbial utilization of lignin, and improved production levels of our desired bioproduct (TAG).

Summary of Activities

- Successfully performed adaption of *R. opacus* to 32 conditions comprising 6 distinct aromatic compounds and mixtures of those compounds leading to the generation of several thousand mutants. Increases in growth rate and lipid accumulation of adapted strains compared to wild type ranged up to 1900% and 225%, respectively.
- Performed DNA-Seq analysis on 35 mutants and identified a number of shared gene targets between distinct evolutionary lineages. Of particular interest was the finding that changes in redox state may facilitate improved aromatic catabolism.
- Utilized RNA-Seq to identify distinct funneling pathways for each of the 5 aromatic compounds, in addition to many other transporters and catabolism enzymes of interest.
- Performed ^{13}C-metabolite fingerprinting to probe glucose and phenol metabolism in *R. opacus*. Determined that the Entner Doudoroff pathway is primary enzymatic path for glucose and that there exists no catabolite repression between a phenolic and sugar monomer when co-fed.
- Developed a suite of chemical and metabolite sensors and characterized plasmid backbones, fluorescent reporters, and antibiotic resistance markers for heterologous gene expression in *R. opacus.*
- Implemented CRISPR interference (CRISPRi) to targeted gene repression.
- Optimized RT-qPCR protocol for rapid quantification of gene expression level in *R. opacus* through the identification of stable reference genes.
- Optimized a Nile Red staining protocol for *R. opacus* such that total lipid concentrations could rapidly be semi-quantitatively determined.
- Developed a genome modification platform to perform gene knock-outs and knock-ins in *R. opacus*.
- Combined RNA-Seq and gene knockouts (enabled by genome modification platform) to characterize three aromatic transporters and confirm aromatic compound degradation routes.

- Utilized LC-MS and MS^n to create a *R. opacus* lipidomic library comprising >100 TAGs, mycolic acids, and phospholipids.
- Examined how carbon feedstock and phenolic adaptation alter lipid composition using MALDI-TOF, LC-MS, AFM, and NanoIR.
- Optimized a thermocatalytic lignin depolymerization protocol that produces lignin breakdown products best suited for *R. opacus*. Demonstrated that this process is feedstock independent, allowing *R. opacus* to catabolize aromatic monomers generated from poplar, switchgrass, and pine.
- Optimized an aromatic derivatization assay that enables the concentration of aromatic monomers to be determined via GC-MS rather than LC-MS (time and cost savings). Used this assay to perform subtractive metabolomics to identify which compounds *R. opacus* was consuming from mixtures of lignin model compound and real-world lignin breakdown products.
- Optimized a spectrophotometric assay to determine phenolic consumption rates using absorbance at 280 nm.
- Performed time course metabolomic assays to examine *R. opacus* aromatic compound preferences and to determine consumption rates.

Hypotheses tested

Hypothesis 1: *R. opacus* strains have evolved to alter the amount and/or composition of TAGs in response to phenolic (toxic solvent) compounds. In particular, **an increased amount of hydrophobic (increased aliphatic tail length, branching, and/or saturation) cell wall-associated lipid correlates positively with the phenolic tolerance level**.

TAGs in *R. opacus* not only serve as the primary carbon storage molecules but also provide substrates for the synthesis of long chain-length extracellular lipids that are components of its protective extracellular envelope. This cell envelope is believed to endow *R. opacus* with increased tolerance to toxic chemicals, including phenolics, although the mechanism by which it confers phenolic tolerance remains elusive. One method that *R. opacus* could adjust the properties of its outer envelope is through changing the structure of its lipids.

In order to understand changes in the lipid composition of *R. opacus*, we first needed a library of what lipid species were present due to the lipids of *R. opacus* having not been extensively studied. We performed high resolution (MS^n) analysis on fractionated lipid extracts to identify a total of 23 phospholipids, 53 mycolic acids, and 44 TAGs species. We then performed LC/MS of wild type (WT) and phenol adapted mutant (evol40) lipid extracts. To quickly examine whether any significant changes existed in lipid composition between the two strains when grown on either glucose or phenol, principal component analysis was performed (Fig 17). By geographically clustering data points based on similarity, we could conclude that while WT and evol40 were similar when grown on glucose, large differences in their lipidomes relative to each other and their glucose profiles existed when grown on either low or high phenol. Performing PCA analysis on

each of the primary lipid species (TAGs, mycolic acids, and phospholipids) confirmed that each of these exhibited compositional changes in WT and evol40 when grown on phenol compared to glucose.

To delve deeper into the actual changes in lipid composition that occurred, we first examined the TAGs. One key finding was that the saturation of TAG species increased between glucose and phenol growth conditions. The WT strain demonstrated a 14% decrease in TAG double bonds from glucose to low phenol, while evol40 demonstrated a 25% decrease between glucose and high phenol (Fig 18). Expanding our investigation to additional lipid species, we also saw a significant change in phospholipid and mycolic acid (MA) saturation and a shortening of mycolic acid carbon chain length (Fig 19 and 20). Phospholipids comprise the inner cellular membrane and have been demonstrated to alter their composition in response to inhibitory compounds. Mycolic acids comprise the outer mycomembrane and have been implicated in the high tolerance of mycobacteria (a related organism) to antibiotics, and they have been shown to change in other actinomycetes in response to aromatic compounds. In low phenol relative to glucose, the average mycolic acid double bond number decreased by 12% in the WT strain and by 16% in evol40 (Fig 20A). For evol40, this average decreased further by an additional 13% in high phenol relative to low phenol. Furthermore, in the WT strain the average carbon number of mycolic acids was similar in glucose and low phenol, while the average carbon number of mycolic acids decreased by ~2 carbons for evol40 in low phenol and 3 carbons in high phenol relative to glucose (Fig 20B). This suggests that the two strains remodel their mycomembranes differently in response to phenol and that potential alterations in mycomembrane composition could affect phenol tolerance.

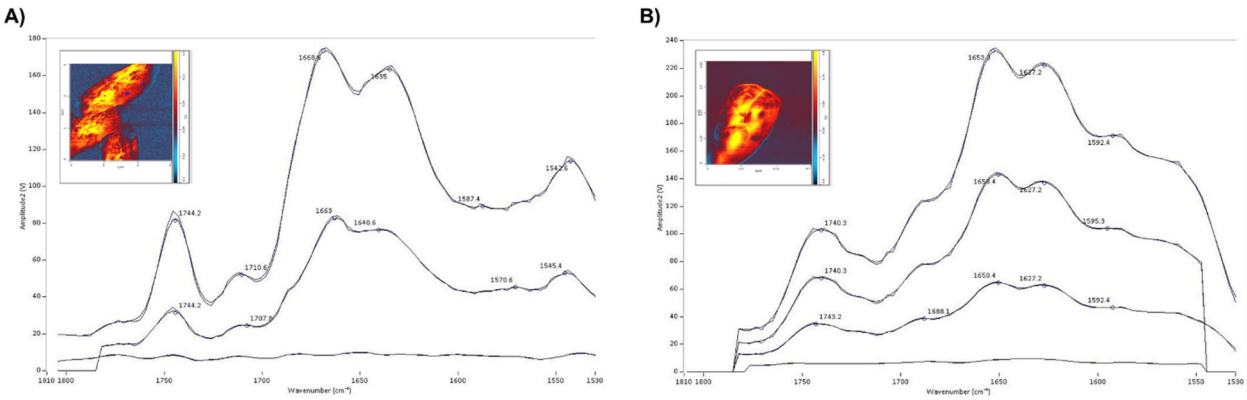


Figure 30. **Nano-IR analysis of *R. opacus***. WT *R. opacus* was grown on glucose (A) and phenol (B). Using Nano-IR, the chemical composition of individual cell surfaces was examined (resolution ~10 nm). Shifts in cell composition were observed between growth conditions.

Another method to investigate the outer cell membrane is Nano-IR, where an infrared (IR) laser is focused at the tip of an atomic force microscope (AFM) cantilever. This tool utilizes an AFM cantilever to examine the size, shape and roughness of the bacteria. However, while the AFM is fixed at a selected point, the laser wavelength is swept to acquire an IR spectrum (1530 to 1800 cm^{-1}), which allows identification of different chemical bonds and groups present. Using Nano-IR, we found preliminary data to suggest that WT cells have different surface cell compositional

profiles when grown on glucose and phenol (Fig 30), corroborating other data that the outer cell membrane undergoes a transition when exposed to phenolics.

In conclusion, we have found evidence that supports our hypothesis that *R. opacus* modulates the composition of its TAGs in response to phenolic compounds. However, TAGs are not the only lipid species that are altered, as we showed that the saturation and chain length of mycolic acids also significantly changed. The role of lipids in the tolerance mechanisms of *R. opacus* is more complicated than we initially hypothesized and warrants further study to conclusively determine the effects of the observed changes and to elucidate any additional changes in lipid structure that may be occurring.

Hypothesis 2: As *R. opacus* evolves to tolerate/utilize varying phenolic compounds, **the amount and/or composition of TAGs** produced will **depend on the carbon source**. Different phenolics will yield different intermediates which will result in varied evolutionary responses in the three types of FASs involved in lipid biosynthesis (e.g., changes in activity by direct enzyme mutation or altered regulation).

In addition to *R. opacus* altering the composition of its lipid profiles, another possibility is that under phenolic stress, *R. opacus* could increase the production of TAGs (or other lipids), which could then lead to an increase in the thickness of its extracellular envelope. To examine this, we used both Folch extractions and correlated Nile Red staining (Fig 6B) to estimate the amount of total lipids being produced when *R. opacus* was grown on glucose and phenol as sole carbon sources (Fig 31). The wild type cells had no change in lipid productivity when grown on either feedstock. However, the phenol adapted strains demonstrated a substantial increase in lipid productivity when grown on a low phenol concentration and even higher productivity when grown on the high phenol condition. Furthermore, we demonstrated that the type of aromatic feedstock can lead to differential amounts of lipid production (Fig 32). For instance, both wild type and phenol adapted strains produced more lipids when grown on HBA or benzoate than when grown on phenol.

TAGs were not the only lipid species that demonstrated a change in quantity based on carbon feedstock. To investigate the quantities of the two primary species of phospholipids, phosphatidylinositol (PI) and phosphatidylethanolamine (PE), in the lipid samples, we calculated their ratio (Fig19). There was a significant increase in the PI/PE ratio of 60 to 130% in the wild type and phenol adapted strain when going from the glucose to a low phenol growth condition. However, the phenol adapted strain demonstrated an even larger increase of 440% in the PI/PE ratio relative to the glucose condition when grown in a high phenol condition. PI is an uncommon phospholipid in bacteria and is more commonly found in eukaryotes such as yeast. In actinomycete strains, PI is an essential phospholipid for growth, and it is a precursor for PI-mannoside and other membrane components such as lipoarabinomannan. Modification of phospholipid head groups has been shown to increase tolerance to various compounds in *Escherichia coli*, including aromatic compounds. This work demonstrates the first study in bacteria where PI is increased in response to phenol and suggests that phospholipid head groups, specifically inositol head groups, could be related to phenolic tolerance in *R. opacus*.

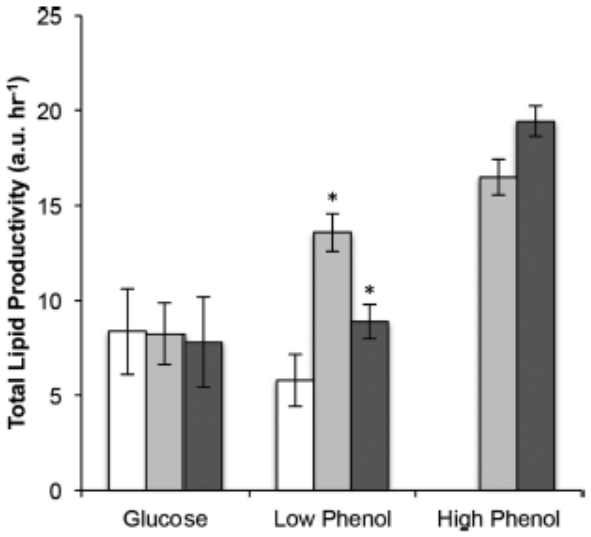


Figure 31. Characterization of lipid accumulation by WT and phenol adapted strains of *R. opacus*[Pub6]. Comparison of total lipid (Nile red fluorescence) productivity between WT and the phenol adapted strains (evol33 and evol40). Strains were grown in lipid accumulating conditions (low nitrogen). White bar = WT, gray bar = evol33 and black bar = evol40. Glucose = 1 g/l, Low Phenol = 0.75 g/l and High Phenol = 1.5 g/l. The data are the average of the arithmetic means of the Nile red fluorescence distribution obtained from three biological replicates. No growth was observed for the WT strain in 1.5 g/l phenol (<1 cell doubling in 10 ml cultures). Error bars represent one standard deviation with all staining and flow cytometry measurements performed on the same day.

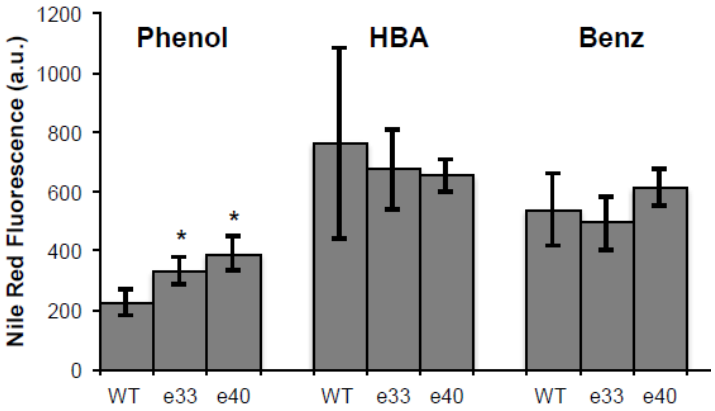


Figure 32. **Lipid accumulation using lignin model compounds[Pub6]**. Cultures were grown in lipid accumulating conditions (low nitrogen 0.05 g/L) using 1 g/L of total phenolics as a carbon source. HBA = 4-hydroxybenzoic acid, Benz = sodium benzoate, e33 = evol33 and e40 = evol40. Stars represent significantly higher lipid accumulation compared to WT. evol33 and evol40 show significantly higher lipid accumulation in phenol compared to WT. However, the adapted strains do not show increased lipid production, compared to that of WT, on other lignin model compounds, although all strains have higher lipid accumulation using hydroxybenzoic acid and sodium benzoate. Bars represent the average of five biological replicates, and error bars represent one standard deviation.

An indirect way that we examined the effect on lipid abundance and composition changes in *R. opacus* was through the use of atomic force microscopy (AFM) to examine cell wall adhesion and stiffness. Some preliminary analysis showed that the phenol evolved strains had a higher mean adhesion force than WT when grown on glucose (Fig 33). This change could indicate that

membrane hydrophobicity is shifted in the adapted strains even when a phenolic compound is not present.

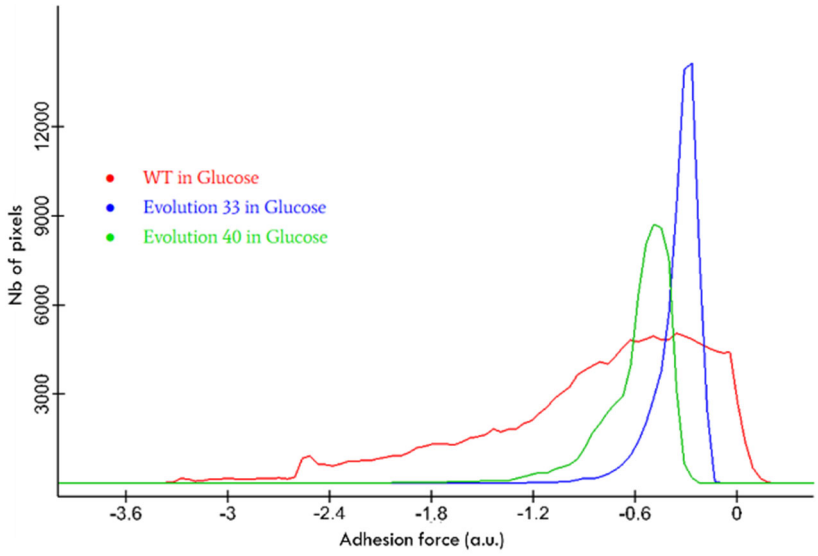


Figure 33. **AFM adhesion force measurements of *R. opacus* outer membranes**. An AFM cantilever was used to measure the adhesion force between the needle and outer membrane of R. opacus. The adhesion force of wild type (WT), and two phenol adapted strains (Evolution 33 and Evolution 40) were measured. A shift to higher adhesion forces was observed in the adapted strains.

An additional method to indirectly look at the effect of phenolic compounds on lipid composition is through RNA-Seq to look at relevant lipid synthesis gene's expression (Table 1). While it is difficult to make interpretations on the data due to the number of lipid related genes whose expression changes, there are many candidates we identified whose expression increases or decreases multiple folds when the cells are grown in glucose or phenol. These results support our more direct measurements showing that lipid composition significantly altered in *R. opacus* when the organism is exposed to phenolic compounds, and these changes could be key to its high levels of tolerance to these compounds.

In conclusion, our data supports that the abundance of TAGs increased in our phenolic evolved *R. opacus* strains when fed a phenolic compound. Furthermore, the amount of TAGs produced did vary based on aromatic feedstock, indicating that there is some correlation. However, TAGs were not the only lipid species to increase in quantity. Phosphatidylinositols, a precursor for phospholipids, drastically increased in our adapted strains when exposed to phenol. Similar to Hypothesis 1, there is a clear relationship between lipids and phenolic tolerance, but it is more complicated than we initially believed, and more work is required to fully investigate *R. opacus*' lipidome.

Hypothesis 3: **Increased utilization of phenolic compounds** by our evolved strains also **contributes to phenolic tolerance mechanisms** by *R. opacus*. This could be due to their enhanced control of catabolic pathways, including regulating the uptake of phenolic compounds through membrane transporters and balanced/increased metabolic flux resulting from evolved enzymes and regulatory proteins.

The ability to quickly catabolize an inhibitory compound is a key mechanism of tolerance for many organisms. The first dataset that we generated was through RNA-Seq, supporting the premise that *R. opacus* increases its utilization of phenolic compounds when exposed to them. When wild type (WT) and our phenol adapted strains (evol33 and evol40) were exposed to a low and high concentration of phenol, several genes annotated as transporters, phenol hydroxylases, and general aromatic degradation pathways (i.e. β-ketoadipate pathway) were very highly upregulated (Fig 10). Interestingly, the evolved strains had higher upregulation of some of these degradation genes compared to the WT at the same phenol concentration, signifying an adapted ability. Additionally, several genes including a transporter were highly upregulated in the evolved strains, but not in the WT. Further RNA-Seq experiments confirmed similar upregulation patterns for four other aromatic compounds (VAN, HBA, GUA, BEN). Additionally, we were able to identify the funneling pathways that convert these five aromatics into β-ketoadipate pathway precursors. (Fig 8). The consumption of aromatics can lead to the formation of toxic 1 carbon (C1) byproducts, such as formaldehyde. We observed upregulation (21- to 2600-fold) of several gene clusters involved in one carbon compound (C1) metabolism during growth on the aromatics VAN and GUA, which require a demethylation step that releases formaldehyde during catabolism.

To further investigate the role of transporters we found upregulated in the presence of different aromatic compounds, we utilized our newly developed targeted genome modification platform to knockout these genes. We then screened the knockout strains for their ability to grow on different aromatic compounds as a sole carbon source to glean insight into the specifies of these transporters (Fig 14). A putative shikimate transporter (RS31355; ShikiT) was significantly upregulated in PHE and the mixture of five aromatics relative to the glucose condition. When knocked out, *R. opacus* demonstrated significantly reduced growth on phenol and the mixture of aromatics. A major facilitator superfamily (MFS) transporter (RS33590; VanT) was moderately upregulated in response to PCA precursors VAN, HBA, and the mixture, but was minimally upregulated by CAT precursors PHE, GUA, and BEN. However, the growth of ΔVanT was reduced on PHE, VAN, and the mixture. Another MFS transporter (RS30810; BenT) located adjacent to the BEN degradation cluster was upregulated in BEN and the mixture. Growth of ΔBenT was significantly reduced on a number of compounds, including PHE, VAN, GUA, BEN, and the mixture. Supporting these findings of degradation being a tolerance mechanism, when the β-ketoadipate pathway was knocked out and *R. opacus* was co-fed glucose and an aromatic compound, the knockout strain was unable to grow in aromatic concentrations that the WT readily grew in.

Leveraging our large DNA-Seq data set, we also investigated genes that were frequent targets of mutation across distinct evolutionary lineages. While we expected to potentially see mutations in genes directly related to aromatic catabolism or lipid production, this was not the case. However, other interesting trends started to emerge. The most frequently mutated genes in *R. opacus* were related to redox reactions. One such example was superoxide dismutase (SOD), a highly conserved enzyme that protects cells from superoxide radicals ($O2^-$) by converting them into molecular oxygen (O_2) or hydrogen peroxide (H_2O_2). We also observed multiple mutations in all phenolic compound adapted strains affecting a cytochrome ubiquinol oxidase, which catalyzes the terminal step in one of the electron transport chains, which reduces O_2 to H_2O. The exact mechanistic contribution of these two mutated enzymes to the evolved aromatic phenotypes is not fully understood, but superoxide radicals are generated from O_2 by aromatic degradation enzymes to facilitate aromatic catabolism. Our observation that our multi-compound adapted strain, PVHG6,

had a 56% decreased SOD activity relative to the WT strain suggests a greater availability of superoxide radicals for aromatic degradation enzymes in PVHG6 (Fig 34).

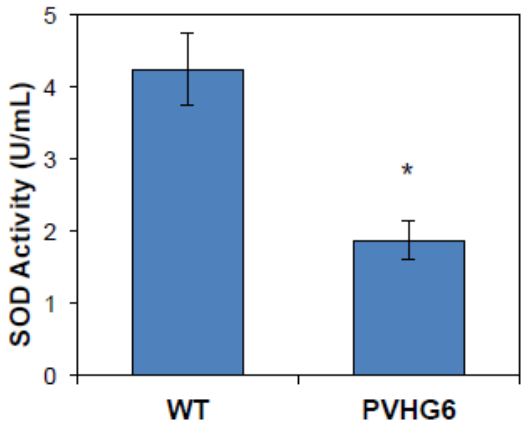


Figure 34. **Comparison of superoxide dismutase (SOD) activity from WT and PVHG6 cell lysates**[Pub2]. Measured relative SOD activity from WT and PVHG6 cell lysates using the Invitrogen Superoxide Dismutase Colorimetric Activity Kit. Bars represent the average of three biological replicates and error bars represent one standard deviation. Asterisk indicates significant difference in activity.

A final interesting insight provided by both our RNA-Seq and DNA-Seq data was that genes relevant to gluconeogenesis, which is the reverse of the EMP pathway, were upregulated in phenolic growth conditions and had DNA mutations across multiple evolutionary lineages. Our ^{13}C-metabolite fingerprinting (Fig 35) showed that glucose is exclusively catabolized through the Entner-Doudoroff pathway, which produces a molecule of NADPH, in addition to NADH and ATP, per every glucose molecule. A need for large quantities of NADPH, which is required for lipid synthesis, could explain this suggested metabolic route of catabolizing aromatics into TCA cycle metabolites and then utilizing gluconeogenesis. More research is needed to confirm that such metabolic fluxes are occurring.

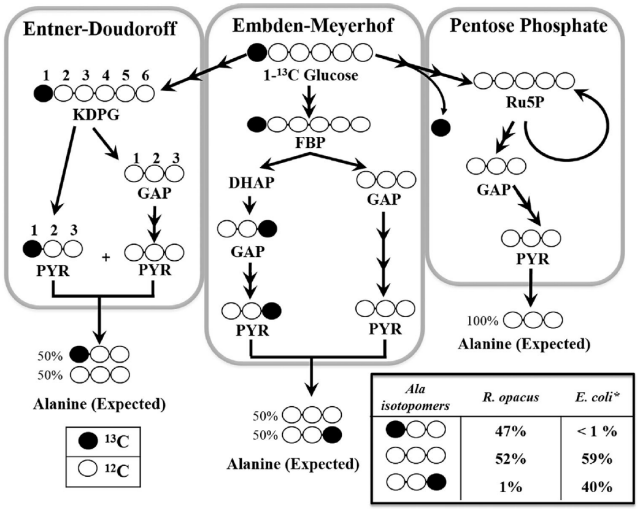


Figure 35. **1-^{13}C glucose catabolism via the ED, EMP, and oxidative PP pathways**[Pub7]. The relative amounts of three key isotopomers of alanine were calculated based on GC-MS data. Alanine from *R. opacus* was mainly labeled at its first position or unlabeled, signifying consumption through the ED pathway. Abbreviations: KDPG, 2-keto-3-deoxy-6-phosphogluconate; GAP, glyceraldehyde-3-phosphate; PYR, pyruvate; FBP, fructose-1,6-biphosphate; DHAP, dihydroxyacetone phosphate; Ru5P, ribulose-5-phosphate.

In summary, the compilation of this data rather conclusively suggests that consumption of aromatic compounds is critical to the high levels of tolerance demonstrated in *R. opacus*. When aromatic degradation pathways or transporters are knocked out, *R. opacus* demonstrates drastically reduced growth. Conversely, evolved strains show higher levels of tolerance by higher levels of upregulation of aromatic catabolism-related genes and by mutations that make aromatic degradation more favorable (altered redox state).

Problems encountered and departure from planned methodology

1) The ability to perform large-scale growth experiments and adaptations with *R. opacus* was limited by the finding that growth in small volume 96-well plates was not consistent on certain phenolic compounds. To eliminate these inconsistencies, we modified the assay protocol to a larger volume (10 mL in 50 mL glass tubes). These growth differences are likely attributed to aeration effects, which limit the amount of oxygen available for phenolic compound degradation. This change greatly hampered our ability to screen large numbers of mutants due to space and handling limitations. To supplement the decrease in mutants screened, we more deeply characterized several promising mutants, particularly PVHG6.

2) The published *R. opacus* genome annotation was not as high quality as originally believed, with many genes marked simply as hypothetical or even misannotated. To check for nucleotide sequence consistency, we aligned to several published assembled genomes. To increase the reliability of DNA-Seq/RNA-Seq interpretations, we merged and cross-referenced multiple genome annotations. Furthermore, we performed ^{13}C-metabolite fingerprinting to provide insightful information about how sugar and phenolics are catabolized by *R. opacus* (Fig 24). This led to several critical findings, such that *R. opacus* lacks catabolite repression between a phenolic and sugar monomer (can co-consume) and that the primary pathway for glucose is via the Entner Doudoroff pathway rather than the EMP pathway. Information such as this cannot be gleaned from genome annotation or even RNA-Seq, and helps better characterize the metabolism of *R. opacus*.

3) The proposed adaption on lignin breakdown products was hampered due to the lack of a reliable source of stable, depolymerized lignin products that were readily able to be consumed by *R. opacus* at the onset of the project. To remedy this, we spent considerable time optimizing a lignin depolymerization process for *R. opacus* and eventually found a feedstock-independent methodology (Fig 22). However, there was a concern that the depolymerized lignin products were not stable long-term due to compound oxidation, represented by a change in coloration, and the formation of precipitates over time (likely oligomeric polymers). If the breakdown products were not stable, we would have to continually produce more at relatively large volumes, in part due to the change from small volume to large volume growth assays, or risk skewing adaptation and growth assay results. As continual production was not feasible and due to the initial delay of a well optimized depolymerization methodology, the scope of the project was altered to not include a lignin

breakdown product adaptation. However, one of the other adapted strains showed great improvement over WT on the mixture of 5 model LBPs we had previously identified as being core constituents of real depolymerized lignin. This demonstrated the ability of *R. opacus* to improve on a mixture of aromatics and identified targets relevant to real depolymerized lignin. We are currently in process of examining relationship between lignin breakdown product age and growth by *R. opacus*, and there does not seem to be much of a negative effect on cell growth which is hopeful for future experiments.

4) The genetic tools available for gene expression and genome modification were not adequate for engineering lipid metabolism (Aim 3.3) or investigating gene roles in *R. opacus*. We spent considerable time developing new promoters (nitrogen and four aromatic sensors, a constitutive promoter library, and three chemical sensors), characterizing additional plasmid backbones and antibiotic resistance markers, and implementing a genome modification platform and system for targeted gene repression (Fig 4, 11, 13, and 25-29). We now possess three characterized heterologous plasmid backbones and can express at least two simultaneously. The aromatic and nitrogen sensors can also be employed in the future for dynamic regulation of heterologous pathways affecting lipid metabolism. Furthermore, we now have tools to examine native gene function through gene knockout experiments or by modulating pathway flux via targeted gene repression (CRISPR interference). While we now possess an impressive genetic toolbox, we did not have time to implement these tools towards lipid pathway engineering. However, we did characterize the role of two aromatic degradation pathways and three aromatic transporters (Fig 14).

5) The direction of Aim 2, particularly Aim 2.1 and 2.2, was shifted based on findings from Aim 1 to focus more on targeted gene knockouts and gene repression. There were several reasons for this change. First, after extensive review and updating of the genome annotation, it was found that there existed many copies of genes related to aromatic consumption and lipid metabolism (both identical copies and highly homologous copies) throughout genome. A transposon knockout library would likely be ineffective due to these many gene copies spread across the genome. Furthermore, preliminary experiments involving overexpression of genes identified by RNA-Seq as relevant to phenol transport and consumption did not lead to any improvement in cell growth or tolerance. We hypothesize that this is because these genes already undergo very high levels of upregulation in the presence of aromatic compounds and may be at saturation within the cell, which supports our hypothesis that rapid uptake and metabolism of aromatics is critical to cell tolerance. Additionally, one of our key findings was that mutations affecting genes related to the reduction and oxidation (redox) state within the cell could be key to improving *R. opacus* aromatic catabolism. A randomized expression library of small chunks of the genome would likely not be able to effectively alter complex traits such as redox state. To more effectively characterize targets identified in Aim 1, we instead developed a targeted genome modification platform to perform gene knockouts, implemented a targeted gene repression platform (CRISPRi) capable of knocking down gene expression, and optimized a RT-qPCR assay to quantify mRNA levels of single genes in *R. opacus*. Using these tools, we confirmed the degradation route of the 6 lignin model compounds we chose based on real depolymerized lignin

products and provided insight into the compound specificity of aromatic transporters identified from RNA-Seq.

6. Products Developed under the Award

Publications (+ Corresponding author)

1. WR Henson, F Hsu, G Dantas, TS Moon[+] and M Foston[+]. Lipid metabolism of phenol tolerant *Rhodococcus opacus* strains for lignin bioconversion. Under review.
2. WR Henson, T Campbell, D DeLorenzo, Y Gao, B Berla, SJ Kim, M Foston, TS Moon[+] and G Dantas[+]. Multi-omic elucidation of aromatic catabolism in adaptively evolved *Rhodococcus opacus*. ***Metab. Eng.*** 49, 69–83 (2018)
3. DM DeLorenzo and TS Moon[+]. Selection of stable reference genes for RT-qPCR in *Rhodococcus opacus* PD630. ***Sci. Rep.*** 8:6019 (2018)
4. DM DeLorenzo, AG Rottinghaus, WR Henson and TS Moon[+]. Molecular toolkit for gene expression control and genome modification in *Rhodococcus opacus* PD630. ***ACS Synth. Biol.*** 7, 727–738 (2018)
5. DM DeLorenzo, WR Henson and TS Moon[+]. Development of Chemical and Metabolite Sensors for *Rhodococcus opacus* PD630. ***ACS Synth. Biol.*** 6, 1973–1978 (2017)
6. A Yoneda, WR Henson, NK Goldner, KJ Park, KJ Forsberg, SJ Kim, MW Pesesky, M Foston, G Dantas[+] and TS Moon[+]. Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating *Rhodococcus opacus* PD630. ***Nucleic Acids Res.*** 44, 2240–2254 (2016)
7. WD Hollinshead, WR Henson, M Abernathy, TS Moon[+] and YJ Tang[+]. Rapid Metabolic Analysis of *Rhodococcus opacus* PD630 via parallel ^{13}C-Metabolite Fingerprinting, ***Biotechnol. Bioeng.*** 113, 91-100 (2016)
8. JA Barrett, Y Gao, CM Bernt, M Chui, AT Tran, MB Foston[+] and PC Ford. Enhancing aromatic production from reductive lignin disassembly: in situ O-methylation of phenolic intermediates. ***ACS Sustainable Chemistry & Engineering***. 4(12), 6877-6886. (2016)
9. CM Immethun, WR Henson, X Wang, D Nielsen and TS Moon[+]. Engineering Central Metabolism for Production of Higher Alcohol-based Biofuels, Chapter 1, p1-34 in ***"Biotechnologies for Biofuel Production and Optimization"*** (CA Eckert & CT Trinh, Ed). Elsevier. dx.doi.org/10.1016/B978-0-444-63475-7.00001-7, **Invited Book Chapter** (2016)
10. Y Gao, M Beganovic and M Foston[+]. Lignin conversion to fuels and chemicals. In Singh, S., Balan, V., ***Biomass Conversion Processes***. Nova Publishing Science Publishers, **Book Chapter** (2016)
11. A Hoynes-O'Connor and TS Moon[+]. Programmable genetic circuits for pathway engineering. ***Curr. Opin. Biotechnol.*** 36, 115-121. **Invited Review** (2015)

Invited Talks in which the DOE Project was discussed

1. TS Moon, Invited Speaker of CSHL Summer Course, Cold Spring Harbor Laboratory, Cold

Spring Harbor, NY, Aug. 1-2, 2018

2. MB Foston, Invited Speaker of 255th American Chemical Society Spring Meeting: "New Horizons in Sustainable Materials" sessions in the Division of Cellulose and Renewable Materials, 2018
3. TS Moon, Invited Speaker of Chemical and Biological Engineering Seminar Series, Northwestern University, Evanston, IL, Oct. 18-19, 2017
4. TS Moon, Invited Speaker of Chemical Engineering Seminar, Arizona State University, Tempe, AZ, Apr. 2-3, 2017
5. TS Moon, Invited Speaker of 2017 DOE Genomic Sciences Program Meeting, Arlington, VA, Feb. 5-8, 2017
6. MB Foston, Invited Speaker of 253rd American Chemical Society Spring Meeting: "2017 ACS Sustainable Chemistry & Engineering Lectureship Awards: Symposium in honor of Gregg Beckham" sessions in the Division of Industrial and Engineering Chemistry, 2017
7. TS Moon, Invited Speaker of ACS Best of BIOT Award Webinars, American Chemical Society, October 12, 2016
8. TS Moon, Invited Speaker of Biological Systems Engineering Seminar, Virginia Tech, Blacksburg, VA, Sep. 26-27, 2016
9. TS Moon, Invited Speaker of Bioinformatics and Computational Biology Graduate Program Seminar, Saint Louis University, St. Louis, MO, Aug. 29, 2016
10. TS Moon, Invited Speaker of 2016 Metabolic Engineering and Green Manufacturing in Microorganisms, Beijing University of Chemical Technology, Beijing, China, Jul. 3-4, 2016
11. TS Moon, Invited Speaker of Biological Sciences Department, KAIST, Daejeon, Korea, Jul. 1, 2016
12. TS Moon, Invited Speaker of Chemical and Biological Engineering Department, Seoul National University, Seoul, Korea, Jun. 28, 2016
13. TS Moon, Invited Session Speaker of AIChE Annual Meeting, Salt Lake City, UT, Nov. 8-13, 2015
14. TS Moon, Invited Speaker of Biomed. Eng. Dept Seminar, WashU, St. Louis, MO, Oct. 1, 2015
15. TS Moon, Invited Session Speaker of AIChE Annual Meeting, Atlanta, GA, Nov. 16-21, 2014
16. TS Moon, Invited Speaker of Indo-US Workshop on Synthetic & Systems Biology, New Delhi, India, Nov. 9-12, 2014

Conference Presentations

1. D DeLorenzo, WR Henson, A Rottinghaus and TS Moon, Development of a Genetic Toolkit in *Rhodococcus opacus* PD630 for reliable and predictable gene expression, *Oral Presentation, AIChE Annual Meeting*, Pittsburgh, PA, October 28 - November 2, 2018

2. D DeLorenzo, WR Henson, A Rottinghaus, YJ Tang, M Foston, G Dantas and TS Moon, Development of a Genetic Toolkit in *Rhodococcus opacus* PD630 for Lignin Valorization, *Poster Presentation, 2018 DOE Genomic Science Annual Meeting,* Tyson's Corner, VA, February 25-28, 2018

3. WR Henson, T Campbell, A Yoneda, Y Gao, D DeLorenzo, B Berla, SJ Kim, M Foston, TS Moon and G Dantas, Multi-omics analysis of *Rhodococcus opacus* strains evolved for optimized lignocellulose bioconversion, *Poster Presentation, 2018 DOE Genomic Science Annual Meeting,* Tyson's Corner, VA, February 25-28, 2018

4. WR Henson, D DeLorenzo, A Rottinghaus and TS Moon, Systems and synthetic biology of *Rhodococcus opacus* to enable conversion of lignin-derived aromatic compounds into lipids, *Poster Presentation, SEED2017*, Vancouver, British Columbia, Canada, June 20-23, 2017, *a rapid fire poster session with 1.5 min presentation*

5. WR Henson, T Campbell, Y Gao, SJ Kim, A Yoneda, B Berla, D DeLorenzo, M Foston, TS Moon and G Dantas, Systems and synthetic biology of *Rhodococcus opacus* to enable conversion of lignin-derived aromatic compounds into lipids, *Poster Presentation, 2017 DOE Genomic Science Annual Meeting,* Arlington, VA, February 5-8, 2017

6. WR Henson, D DeLorenzo and TS Moon, Systems and synthetic biology of *Rhodococcus opacus* to enable conversion of lignin-derived aromatic compounds into lipids, *Oral Presentation, ICBE*, San Diego, CA, January 8-11, 2017

7. WR Henson, D DeLorenzo, SJ Kim and TS Moon, Towards development of *Rhodococcus opacus* as a microbial cell factory: conversion of lignin-derived aromatic compounds into lipids, *Oral Presentation, AIChE Annual Meeting*, San Francisco, CA, November 13-18, 2016

8. WR Henson, A Yoneda, NK Goldner, B Berla, KJ Park, SJ Kim, KJ Forsberg, MW Pesesky, M Foston, G Dantas and TS Moon, Conversion of lignin-derived aromatic compounds into lipids by engineered *Rhodococcus opacus* strains, *Poster Presentation, Monsanto Fellows Symposium*, Chesterfield, MO, April 13-14, 2016.

9. WR Henson, SJ Kim, YJ Tang, M Foston, G Dantas and TS Moon, Conversion of lignin-derived aromatic compounds into lipids by engineered *Rhodococcus opacus* strains, *Oral Presentation, 251st ACS National Meeting*, San Diego, CA, March 13-17, 2016. Identified as the "Best Presentation" of the Session

10. A Yoneda, WR Henson, NK Goldner, KJ Park, KJ Forsberg, SJ Kim, MW Pesesky, M Foston, G Dantas and TS Moon, Conversion of lignin-derived aromatic compounds into lipids by engineered *Rhodococcus opacus* strains, *Poster Presentation, 2016 DOE Genomic Science Annual Meeting,* Tyson's Corner, VA, March 6-9, 2016

11. WR Henson, SJ Kim and TS Moon, Conversion of Aromatic Compounds to Lipids by Engineered *Rhodococcus* Strains, *Oral Presentation, AIChE Annual Meeting*, Salt Lake City, UT, November 8-13, 2015

12. WD Hollinshead, WR Henson, M Abernathy, TS Moon and Yinjie Tang, Elucidating the Metabolism of *Rhodococcus opacus* PD630 via Parallel ^{13}C-Metabolite Fingerprinting, *Oral Presentation, AIChE Annual Meeting*, Salt Lake City, UT, November 8-13, 2015

13. WR Henson, SJ Kim, S Srivastava, KJ Park and TS Moon, Development of a generalized

platform to produce value-added chemicals from lignocellulose, *Featured Oral Presentation, IBE Annual Conference*, St. Louis, MO, March 5-7, 2015

14. WR Henson, SJ Kim, S Srivastava, KJ Park and TS Moon, Development of a generalized platform to produce value-added chemicals from lignocellulose, *Oral Presentation, AIChE Annual Meeting*, Atlanta, GA, November 16-21, 2014

Collaborations Fostered

1. The team started collaborations with Dr. Yinjie Tang (WashU), Dr. Fuzhong Zhang (WashU), and Dr. Hector Garcia Martin (Lawrence Berkeley National Laboratory), leading to a new DOE project (DE-SC0018324; Systems Engineering of *Rhodococcus opacus* to Enable Production of Drop-in Fuels from Lignocellulose)
2. The team has been collaborating with Dr. Fong-Fu Hsu to understand lipid metabolism of *Rhodococcus opacus* strains for lignin bioconversion (a manuscript under review). Dr. Hsu is an expert in Mass Spectrometry.
3. The team has been collaborating with Dr. Laurene Tetard (University of Central Florida) to understand the cell membrane composition and properties of *Rhodococcus opacus* strains for lignin bioconversion.
4. Our team has been part of two Pacific Northwest National Laboratory Science Theme proposals at Environmental Molecular Sciences Laboratory, Science Theme proposals "Advanced Molecular Characterization of Solvo-Catalytic Lignin Disassembly (FY16) and "Integrated Mutli-omics Analyses for the Prediction of Metabolic Flux in Lignin Conversion" (FY18)

Submittal Form. The report must be accompanied by a completed electronic version of DOE Form 241.3, "U.S. Department of Energy (DOE), Announcement of Scientific and Technical Information (STI)." You can complete, upload, and submit the DOE F.241.3 online via E-Link. You are encouraged not to submit patentable material or protected data in these reports, but if there is such material or data in the report, you must: (1) clearly identify patentable or protected data on each page of the report; (2) identify such material on the cover of the report; and (3) mark the appropriate block in Section K of the DOE F 241.3. Reports must not contain any limited rights data (proprietary data), classified information, information subject to export control classification, or other information not subject to release. Protected data is specific technical data, first produced in the performance of the award that is protected from public release for a period of time by the terms of the award agreement.