Winter 12-15-2019

Rhodococcus opacus PD630 Genetic Tool Development to Enable the Conversion of Biomass

Drew Michael Delorenzo
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

McKelvey School of Engineering

Department of Energy, Environmental, and Chemical Engineering

Dissertation Examination Committee:

Dr. Tae Seok Moon (chair)
Dr. Gautam Dantas
Dr. Marcus Foston
Dr. Himadri Pakrasi
Dr. Fuzhong Zhang

Rhodococcus opacus PD630 Genetic Tool Development to Enable the Conversion of Biomass

by

Drew Michael DeLorenzo

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2019

St. Louis, Missouri

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

Throughout my studies at Washington University, I have been fortunate to have a group of supportive friends, colleagues, and mentors. First, I would like to thank Dr. Tae Seok Moon for acting as my PhD advisor and committee chair. Second, I would like to acknowledge my committee members, Dr. Gautam Dantas, Dr. Marcus Foston, Dr. Himadri Pakrasi, and Dr. Fuzhong Zhang for their mentorship in crafting and conducting my dissertation research. I would also like to acknowledge my previous undergraduate research mentors, Dr. Jerry Hilbish and Dana Bethea, in addition to Dr. Claudia Benitez-Nelson for inspiring me to pursue a career in science.

None of this research would have been possible without my funding sources, including the National Science Foundation Graduate Research Fellowship [DGS-1143954] and the Department of Energy [DE-SC0012705; DE-SC0018324].

I would also like to thank my labmates, past and present, including Ray Henson, Cheryl Immethun, Tatenda Shopera, Ryan Lee, Austin Rottinghaus, Matt Amrofell, Rhiannon Carr, Charlie Johnson, Kenneth Ng, Soo Ji Kim, Anamika Chatterjee, Caroline Focht, and Matthew Leong, for their support, discussions (scientific or otherwise), and comradery. I am equally grateful to my friends, including Alison Schriro, J.B. Easley, Matt Flemer, Eugene Kim, Taylor Teshon, and Emily Purcell who have always provided counsel, laughs, and support when I needed it. To Katie, you are the beacon of light that helped me navigate the storm. Thank you for your love and constant support. Finally, I would like to express my gratitude to my parents for their unwavering support throughout graduate school and pushing me to pursue a PhD.

Drew DeLorenzo

Washington University in St. Louis
December 2019

Dedicated to my parents

Abstract of the Dissertation

Rhodococcus opacus PD630 Genetic Tool Development to Enable the Conversion of Biomass

By

Drew Michael DeLorenzo

Doctor of Philosophy in Energy, Environmental and Chemical Engineering

Washington University in St. Louis, 2019

Professor Tae Seok Moon, Chair

The discovery of fossil fuels facilitated a new era in human history and allowed many firsts, such as the mass production of goods, the ability to travel and communicate long distances, the formation of population dense cities, and unprecedented improvements in quality of life. Alternative sources of energy and chemicals are needed, however, as hydrocarbon reserves continue to deplete and the effects of burning fossils on the planet become better understood. Lignocellulosic biomass is the most abundant raw material in the world and a viable alternative to petroleum-derived products. The pre-treatment of lignocellulose (e.g., thermocatalytic depolymerization, enzymatic hydrolysis, pyrolysis, etc.) generates a range of products, including readily available sugars for microbial fermentation. One of the typically unused fractions of biomass is the structural component, referred to as lignin, that makes up 15 to 30% of the material and when depolymerized generates a heterogeneous mixture of toxic aromatic compounds. Generally, lignin is separated from the carbohydrate fraction and burned, but its utilization has been identified as a key factor in biorefinery profitability. One possible option for lignin valorization is to find a microbe that not only ferments lignocellulose-derived sugars into a valuable commodity, but also the lignin-derived aromatics.

Rhodococcus opacus PD630 (hereafter *R. opacus*) is a non-model, gram-positive bacterium that possesses desirable traits for biomass conversion, including consumption capabilities for both lignocellulose-derived sugars and aromatic compounds, significant accumulation of the biodiesel precursor triacylglycerol, a relatively rapid growth rate, and genetic tractability. Few genetic elements and molecular biology techniques, however, have been directly characterized in *R. opacus*, limiting its application for lignocellulose bioconversion. The goal of this dissertation is to greatly expand the genetic toolbox available in *R. opacus* in order to provide insight into its aromatic catabolism and to promote its use as a microbial chassis for the conversion of biomass-derived products into biofuels or other value-added products. The contributions developed as part of this dissertation include 1) the development of strong constitutive promoters for the overexpression of heterologous genes, 2) the development of chemical and metabolite sensors for tunable gene expression, 3) the characterization of native and endogenous plasmid backbones and resistance markers, 4) a heterologous T7 RNA polymerase expression platform for gene expression, 5) the demonstration of genetic logic circuits for programable gene expression, 6) a recombinase-based recombineering platform for gene knockouts and insertions, 7) a CRISPR interference (CRISPRi) platform for targeted gene repression, 8) the identification of stable reference genes for RT-qPCR applications, 9) insight into aromatic degradation through the β-ketoadipate pathway via gene knockouts, and 10) insight into the role of aromatic transporters via gene knockouts. Taken together, this work greatly advances the ability to engineer *R. opacus* for any desired application, in addition to providing understanding into its catabolism of aromatic compounds.

Chapter 1: Introduction

1.1 Need for renewable sources of energy and chemicals

Humanity has faced and overcome innumerable challenges throughout its collective history. The greatest challenge of the 21st century and beyond, however, will be responding to anthropogenic climate change. Since the discovery of fossil fuels and the subsequent onset of the industrial revolution in the mid-1700s, the relationship between humans and the biosphere has been significantly altered. The use of fossil fuels as energy-rich power sources facilitated a new era in which humans moved to population-dense cities, traveled and communicated rapidly across great distances, mass-produced goods, enhanced agricultural output, and improved their quality of life. This paradigm shift in human existence came at an unforeseen cost, however, as the burning of fossilized carbon sources led to the large-scale release of greenhouse gases, particularly carbon dioxide (CO_2). Compared to pre-industrial levels, the current global atmospheric CO_2 concentration has increased over 30%, or 100 parts per million (ppm), and is now stably maintained at over 400 ppm, with 75% of that change occurring since 1960.[1] Furthermore, the current atmospheric concentrations of carbon dioxide, methane, and nitrous oxide are at their highest levels in the last 800,000 years.[2] Emissions will be further exacerbated as the human population is projected to continue growing to 9.5 billion by 2050 and the standard of living in developing countries continues to improve.[3]

Increased levels of greenhouse gases have already led to many negative consequences across the globe and will continue to present an unparalleled challenge for years to come. The most direct outcome of an amplified concentration of greenhouse gases is global warming. Certain gases, such as CO_2 and methane, cause the atmosphere to behave similarly to a greenhouse,

wherein visible radiation from the sun can penetrate through the gaseous layer, but the resulting radiated thermal energy from the planet is prevented from escaping. This trapped thermal energy in turn leads to an increase in system temperature. As a result of this enhanced greenhouse effect, the three decades from 1983 to 2012 were the warmest 30-year period of the last 1400 years in the Northern Hemisphere.[2] The observed terrestrial warming would be even more significant if the oceans had not absorbed more than 90% of the additional thermal energy accumulated since 1971.[2] Over the past 250 years, the earth's atmosphere has warmed approximately 1 °C, ocean pH has increased by 26% due to the additional dissolved CO_2, Arctic sea ice is decreasing at a decadal rate of approximately 4%, and global mean sea level has risen about 0.2 meters.[2] If the overall planet temperature increases by over 2 °C, wide-scale global consequences including, but not limited to, even more substantial sea level rise from the melting of glaciers, decreased agricultural and fishery output, further ocean acidification, extreme weather (e.g., drought or heat waves), and higher rates of natural disasters (e.g., wildfires, hurricanes, monsoons) will all occur.[2] Based on leading emission scenarios calculated by the International Panel on Climate Change (IPCC), if global warming is to be limited to less than 2 °C over the next century, CO_2 levels must remain below 450 ppm. Achieving this would require bringing net CO_2 emissions down to 0 by 2100.[2] As such, an alternative supply of fossil fueled-derived energy and chemicals is sorely needed.

1.2 Sources of renewable energy and chemicals

Sustainable energy can be derived from multiple sources (e.g., solar, wind, geothermal, hydroelectric, or biomass), eliminating the need for fossil fuels to power electricity generation. A single technology, however, is not capable of fully displacing hydrocarbon deposits, meaning that a portfolio of energy sources will be required.[4] Furthermore, almost all sources of renewable energy are geographically or temporally limited, further necessitating the development of multiple

complementing technologies.[5] For example, solar panels can only generate electricity during the day, turbines are dependent on variable winds, and hydroelectric dams require a significant body of water. The stochastic availability of renewable energy sources can also lead to disruptions in the electrical grid, as conventional energy plants cannot quickly ramp their production up or down to meet demand.[5] To avoid blackouts, an excess of fossil fuel-derived energy is produced as a buffer against variable renewable energy sources, partially negating the desired effect of employing these sources. To make these forms of renewable energy available to the entire population, significant gains in long-range electricity transmission and energy storage (i.e., batteries) are required to deploy energy across electrical grids when needed.[5] In spite of these challenges, renewable energy made up 19.2% of global final energy consumption as of 2014 and is expected to continue to rise.[4]

An alternative to long-range electricity transmission or batteries is storing energy chemically through the production of energy-dense liquid fuels, such as biodiesel and bioethanol. Potential options include: electrochemically generating liquid fuels from CO_2 using a catalyst and a sustainable source of electricity, recycling waste oils (ex. cooking oil) into "drop-in" fuels for difficult to replace products (e.g., jet-fuel), and converting biomass, whether derived from food or energy crops, agricultural residues, or algae, into liquid fuels.[2-3, 6-7] Biomass is already the fourth-largest source of global energy, accounting for 10%–14% of the final energy consumption over the past decade, although this energy is typically extracted through the direct burning of material for heat.[8] Biomass, unlike other renewable energy sources, can also be used as a carbon feedstock to generate replacements for petroleum-derived products.

First-generation biofuels, or those generated from potential food products (e.g., starch, sugar, animal fats, and vegetable oil), are an increasingly-popular improvement over directly

burning biomass, with production reaching 106 billion liters worldwide in 2012 (82.6 billion liters of ethanol and 23.6 billion liters of biodiesel).[7] These types of fuels are not new; the inventor of the diesel engine himself advocated for farmers generating their own biofuel in areas without a source of petroleum.[9] There is concern, however, that as the human population continues to climb, using potential food goods for fuel could lead to shortages as the combined global demand for food and fuel outweighs the supply of arable land.[10-11] Furthermore, while first-generation biofuels may mitigate CO_2 emissions associated with fossil fuels, they still pose economic and environmental problems, such as the energy consumed related to irrigation, fertilizer production, crop cultivation, and transportation.[9-10, 12] Second-generation biofuels, or those produced from non-food biomass or lignocellulose, can alleviate many of the concerns garnered by first-generation biofuels.[13]

1.3 Lignocellulosic biomass as a sustainable carbon feedstock

Lignocellulose is the most abundant raw material in the world. Three primary sources exist for use as second-generation fuels and chemicals, including virgin biomass (e.g., common trees, shrubs, and grasses), agricultural and forestry wastes and residues (e.g., corn stover and sugarcane bagasse), and dedicated energy crops (e.g., poplar and switchgrass).[13-14] Lignocellulosic biomass, regardless of source, is comprised of three main fractions. Two of these fractions, cellulose and hemicellulose, are polymers predominantly composed of carbohydrates that can be hydrolyzed into easily fermentable sugars, predominantly glucose and xylose, and make up to 70-85% of the biomass dry-weight.[15] The third fraction, representing 15-30% of the biomass dry-weight, is the structural component, referred to as lignin, that imparts a “woody” characteristic to the plant, prevents polysaccharide degradation, and waterproofs the cell wall and vascular system.[16] Lignin’s structural rigidity and recalcitrance to degradation derives from it being a complex, cross-linked, and highly heterogeneous aromatic macromolecule.[17-18]

Up to a billion tons of lignocellulosic biomass is potentially available in the United States, deriving predominately from agricultural and forestry waste. If all of this biomass were utilized, current petroleum consumption could be reduced by at least 30%.[19] Additional study needs to conducted, however, to determine how much of these feedstocks should be used for energy and chemical production, as agricultural and forestry wastes are often used for other tasks, such as the maintenance of soil nutrient quality (e.g., re-tilling of corn stover into the soil).[3, 8] Furthermore, many industrial biomass by-products are currently already used for other needs and there may not be enough left over for energy production.[3, 8] This is the case in Germany, where all biomass waste is already fully utilized.[3, 8] Energy crops that perennially produce large quantities of biomass are a viable supplement to industrial, agricultural, and forestry wastes, as they can typically grow in nutrient-poor soil, have a lower water demand than food crops, and can be beneficial to the environment.[14] Approximately 13.6 million hectare of land that is not suitable for food production is available for switchgrass cultivation in the United States.[20] Regardless of source, biorefineries can convert biomass into a range of products by employing a number of distinct catalytic, thermochemical, and biological conversion processes to take advantage of the carbon and energy stored in this renewable material.[21-22]

Lignocellulose, unfortunately, does require more complex and costly processing to produce biofuels or products compared to first-generation feedstocks. Lignin provides a particular challenge as it resists degradation and generates toxic aromatic compounds when depolymerized.[23-25] Generally, the lignin fraction of biomass is separated from the carbohydrate fractions and is either discarded or burned to recover low-grade heat.[26] The economic viability of second-generation biorefineries, however, depends on the complete utilization and upgrading of lignin. According to the National Renewable Energy Laboratory (NREL), selling the lignin

fraction at the same price as the original feedstock for conversion into higher value by-products could reduce the minimum sugar transfer price by 12.5%.[27] This is significant considering that the production of sugar from corn stover makes up 69% of the suggested minimum ethanol sale price of $2.15, greatly eating into potential profits.[28] The global production of industrial-waste lignin is estimated at approximately 70 million tons per year and projections suggest that, by 2025, a single second-generation bio-alcohol refinery could generate approximately 20-25 million pounds per year of lignin as a by-product.[29-30] While the utilization of the carbohydrate fraction of biomass has increased by several orders of magnitude, with cellulosic bioethanol production projected to ramp up to 14 billion gallons per year by 2022, the question still remains what to do with the lignin fraction.[31-32]

1.4 Taking advantage of lignin

Lignin is the second most abundant polymer on the planet and can be separated from the carbohydrate fraction of lignocellulose in many ways (e.g., organic solvent, pulping, alkaline or acid treatment).[33-34] The method utilized, however, affects the composition of the final extract, so downstream processes need to be considered when choosing a technique.[33-34] As previously mentioned, lignin is an extremely heterogeneous polymer composed of aromatic subunits derived primarily from *p*-coumaryl, coniferyl, and sinapyl alcohols that have undergone pseudo-random enzymatic polymerization.[23, 34-35] The composition of the extracted lignin also depends on the original biomass source, as the ratios of these three monolignol precursors varies between species, genotypes, and even across tissue types, in addition to being a function of environmental factors during growth.[36] For example, softwood lignin consists mainly of coniferyl alcohols units (95%), while hardwood lignin predominantly contains coniferyl alcohol and sinapyl alcohol units in equivalent ratios.[37] The monolignols form several types of chemical bonds, but the most common

are aryl ether bonds (e.g., β-O-4, α-O-4, and 4-O-5).[16] The extreme diversity and variability in the composition of lignin makes commercial degradation difficult and widespread utilization of lignin challenging; however, it is the largest renewable source of aromatic compounds globally and is critical to displacing petroleum derived fuels and chemicals.

There are many different uses for lignin than span the gamut of valuations. Currently the most common and lowest value use, is burning the lignin for low-grade heat.[26] Burning biomass requires high initial temperatures (800 to 1200 ^{o}C) to start ignition, particularly when the material has not been pre-dried, but it is an economical way to generate heat.[38] Alternative options for electricity generation are the gasification of lignin into synthesis gas (syngas), which can be burned directly or converted into synthetic natural gas (SNG), or the production of liquid bio-oil through fast pyrolysis.[39-40] Gasification requires slightly lower temperatures than biomass burning, which reduces energy inputs and the production of undesirable nitrogen and sulfur oxides. Additionally, the final product can be used in efficient reciprocating engines or gas turbines for electricity generation.[38] Fast pyrolysis utilizes a much lower temperature than both burning and gasification (350 to 600 ^{o}C) and the resulting liquid product is more easily transported.[38] Bio-oil is typically not useful for anything other than burning though due to its extremely heterogeneous distribution of compounds that prevents effective chemical separations.[41]

Lignin can also be used as a carbon feedstock to produce value-added compounds. The most common method to upgrade lignin is recovering lignosulfonates formed during sulfite pulping of lignocellulose, which can then be used as plasticizers in concrete, additives in animal feed, or converted into the high-value food additive vanillin.[34, 42] There are several challenges associated with the production of lignosulfonates, the primary issue being that most industrially produced lignin is Kraft lignin, meaning that additional steps are required for sulfonation.

Additionally, lignosulfonates must be physically or chemically extracted and purified from the spent sulfite liquors collected during sulfite pulping, the degree of sulfonation is heterogeneous, and the molecular weights of the products can span orders of magnitude.[43] Alternatively, thermochemical and catalytic technologies (e.g., oxidative cracking, hydrogenolysis, and solvolysis) that depolymerize the entirety of lignin into its constituent aromatics can be applied.[44] Catalytic technologies can depolymerize lignin more selectively due to targeting specific chemical bonds, in addition to reducing the occurrence of secondary reactions due to the lower required input energy.[44] Enzymes and various microorganisms have also been investigated as a selective and low-cost method of lignin depolymerization.[45-46] In general, biological systems require mild conditions that avoid costs associated with the use of high temperatures and high pressures. Only a few bacteria (e.g., *Streptomyces* spp., *Rhodococcus* spp., and *Nocardia* spp.) and brown/white-rot fungi, however, have demonstrated an ability to depolymerize lignin, and their depolymerization rates are often considered too slow to be useful on an industrial scale.[24, 47-49]

Regardless of the method employed, a lignin breakdown mixture with a narrower distribution of aromatic compounds is more cost-effective for chemical separations and product upgrading, which helps facilitate the widespread adoption of lignin valorization. The lignin breakdown products, however, still generally contain so many compounds that performing chemical separations for specific products leads to isolation of only a small fraction of the total mixture. One solution is to utilize a process that can convert all the aromatic compounds in a depolymerized lignin mixture into a single compound that can then be processed into other chemicals using traditional methods. A biological catalyst (i.e., a microbe) can use its array of catabolic enzymes to "funnel" a mixture of compounds through the organism's metabolism into a single metabolite.[50] Combining a process such as thermocatalytic depolymerization of lignin with

a biological catalyst that can reduce lignin breakdown products into a single high-value compound has been referred to as a hybrid conversion strategy.[50-51]

1.5 Hybrid conversion of lignocellulose into value-added products

A hybrid conversion approach combines the best attributes of a depolymerization technology, such as a thermocatalytic reaction, with biological conversion technologies.[51] In a hybrid conversion approach, a process with advantageous reaction kinetics and selective product formation is applied for the initial lignin depolymerization. Microbial conversion is then applied downstream and funneling of the depolymerized lignin breakdown products to a single value-added product then occurs with advantageous selectivity.[52] Specially evolved organisms possess an array of upper degradation pathways than can convert a distribution of aromatics into common metabolic intermediates (e.g., protocatechuate and catechol), effectively acting as a "biological funnel".[50-51] These intermediates undergo further conversion to central metabolites (e.g., acetyl-CoA) that can be utilized to produce target compounds with high selectivity and concentration.

Hybrid conversion technologies have been previously implemented, but they have almost exclusively focused on cellulosic sugar utilization.[53-54] Recent work has begun to shift the focus to lignin utilization, with initial efforts focusing on screening different biological catalysts (e.g., *Pseudomonas putida* and *Rhodococcus* spp.) with lignin model compounds to characterize their aromatic degradation pathways and bioconversion abilities.[55-58] A number of organisms have now also been demonstrated to consume real lignin breakdown products, but which biological catalyst is best for conversion still remains an open question.[59-66]

1.6 Choosing *Rhodococcus opacus* PD630 as a biological catalyst

The choice of a biological catalyst is important, as the chassis needs to not only possess a sufficient number of catabolic pathways for the consumption of lignin-derived aromatics, and ideally hexose and pentose sugars, but also have a tolerance towards inhibitory compounds generated during biomass depolymerization. The pretreatment of lignocellulose generates an array of toxic compounds, such as aromatics from the lignin fraction or furfurals from the hemicellulose fraction, that conventional model microbes (e.g., *Escherichia coli*) are not able to effectively tolerate or consume.[15, 67-68] Furthermore, if the conversion process is to be profitable, the biological catalyst must be able to produce a valuable product in large quantities, which requires a high initial feedstock load.

The Actinomycetales bacterium *Rhodococcus opacus* PD630 (hereafter *R. opacus*) has many desirable traits for the microbial valorization of lignocellulose, most critically the ability to consume a wide variety of biomass-derived substrates. *R. opacus* has been screened on a broad range of compounds and demonstrated growth on oligosaccharides, alcohols (e.g., arabitol and mannitol), hexose sugars (e.g., glucose and galactose), carboxylates (i.e., organic acids), aromatics (e.g., phenol, guaiacol, and benzoate), and lignocellulose-derived by-products (e.g., furans).[69-71] Furthermore, *R. opacus* strains have been engineered to consume common biomass-derived pentose sugars (e.g., arabinose and xylose), in addition to the disaccharide cellobiose.[72-75] *R. opacus* has been further confirmed to consume several forms of treated biomass, including switchgrass pyrolysis oil[66], alkali-treated corn stover and poplar wood [48, 64, 76], and depolymerized kraft lignin[77]. Very recently, an engineered strain of *R. opacus* that secretes a heterologous laccase enzyme was demonstrated to depolymerize and grow on insoluble Kraft lignin in a consolidated bioprocess.[78] Providing further insight into the metabolism of *R. opacus*, an annotated reference

genome is available and several transcriptomic and proteomic datasets have been generated when cells were grown on various feedstocks (e.g., glucose, aromatics, and kraft lignin).[69, 78-79]

R. opacus not only possesses a diverse set of catabolic pathways, but it has also previously demonstrated a natural and adaptive tolerance towards inhibitory lignocellulose-derived products, including biomass pretreatment by-products (e.g., furans), organic acids, aromatics, and chlorinated and halogenated compounds.[69-70, 80] The tolerance mechanisms in *R. opacus* are not fully understood, but are partially attributed to its numerous catabolic pathways that consume inhibitory compounds, such as aromatics via a high flux β-ketoadipate pathway, which reduces the effective local concentration of these compounds.[18, 81] Furthermore, its ability to modulate its lipidome in response to a phenolic compound suggests membrane permeability contributes to tolerance.[82] *R. opacus* also demonstrates advantageous conversion rates as it can concurrently consume glucose via Entner–Doudoroff pathway, acetate via the glyoxylate shunt, and phenolic compounds via β-ketoadipate pathway, leading to improvements in productivity.[72, 83] It has also demonstrated osmotic tolerance to very high initial loadings of sugars feedstocks (up to 419 g/L glucose), which facilitates higher final product titers.[71, 84]

R. opacus can convert lignocellulose-derived feedstocks into a number of high-value products. *R. opacus* is primarily known for its ability to produce significant amounts of the biodiesel precursor triacylglycerol (TAG; up to 76% of its cellular dry weight when grown on gluconate).[80] This oleaginous ability in *R. opacus* is attributed to its type 1 FASI lipid synthesis mechanism, which is rare in prokaryotes and is normally found only in animals and fungus.[78] Bacteria commonly use the type 2 FASII system. The type 1 FASI system facilitates more efficient lipid synthesis through the use of a single multicomponent and multifunctional enzyme, compared to many single function enzymes in the type 2 FASII system.[78] The TAGs generated through this

efficient lipid synthesis pathway can then be extracted and readily converted into biodiesel through a transesterification reaction that generates three free fatty acids (FFAs) and a glycerol. The glycerol can be separated from the FFAs and fed back to the cell as a secondary carbon source.[85-86] The lipid metabolism has also been engineered in several ways to foster value-added product synthesis, including the overexpression of enzymes identified as reaction bottlenecks to increase lipid production and the re-routing of lipid species to produce wax esters, which have roles in the production of cosmetics, pharmaceuticals, and lubricants.[78, 87] Intermediates in the aromatic degrading β-ketoadipate pathway can also be diverted into high-value products, such as muconic acid or succinic acid.[88-90] Finally, *R. opacus* natively produces carotenoids, which can be used as dietary supplements to benefit human health.[91]

All of these natural and engineered abilities of *R. opacus* make it an ideal microbial biorefinery for the conversion of lignocellulosic material, particularly lignin, into fuels and chemicals. Many of the modifications performed in *R. opacus*, however, are commonly achieved using borrowed and generally uncharacterized genetic parts from related Actinobacteria.[72-74, 87, 92] Genetic parts (e.g., promoters, plasmid backbones, dynamic sensors, etc.) and molecular biology techniques (e.g., genome recombineering, targeted gene repression, gene expression quantification, genetic logic implementation, etc.) often taken for granted in model organisms such as *E. coli* are limited in *R. opacus*. The goal of this dissertation is to develop and apply a characterized genetic toolbox for advanced engineering in *Rhodococcus opacus* PD630 and to provide insight into its aromatic catabolism.

1.7 Summary and organization of dissertation

The primary objective of this dissertation is to expand the genetic toolbox for *Rhodococcus opacus* PD630 as a microbial platform for the conversion of lignocellulose into value-added

compounds. A subset of the tools developed herein will be applied to enable more sophisticated gene expression (i.e., genetic logic) and to provide insight into aromatic consumption and pathway regulation. In Chapter 2, several chemical and metabolite sensors were developed and characterized in *R. opacus* to facilitate tunable gene expression. Furthermore, multiple fluorescent proteins from three distinct color categories, which allows the simultaneous utilization of up to three reporters, were screened for the best signal over the background. One of the developed sensors was repurposed as a cellular timer to enable gene expression at a desired time point during cell culture. Finally, the activity of five native *R. opacus* promoters from genes related to aromatic catabolism were examined in response to six model lignin compounds, providing novel insight into regulatory mechanisms. This work was published in ACS Synthetic Biology.[93] In Chapter 3, additional genetic parts for strong, tunable gene expression were developed, including the development of novel IPTG-inducible promoters, four selective aromatic inducible promoters, and the first application of the T7 RNA polymerase expression system in *Rhodococcus* spp. Genetic parts from this chapter and Chapter 2 were then combined to demonstrate three different forms of genetic logic (AND, NAND, and IMPLY logic gates). This work was published in ACS Synthetic Biology. In Chapter 4, the genetic toolbox is further expanded to include a constitutive promoter library for variable gene expression, a new curable plasmid backbone for temporary gene expression, two newly characterized antibiotic resistance markers for selection, a system for targeted and tunable gene repression (e.g., CRISPR interference), three neutral sites in the genome for gene integration, and a platform to enable genome recombineering (i.e., gene knockouts and gene insertions). Furthermore, the copy number of 4 heterologous plasmids and 9 endogenous plasmids was determined through quantitative PCR (qPCR). This work was published in ACS Synthetic Biology.[94] In Chapter 5, a set of stably expressed reference genes was identified in *R.*

opacus for use as benchmarks to normalize mRNA expression data collected through reverse transcription quantitative PCR (RT-qPCR). RT-qPCR is critical for quantifying the repression of native genes via CRISPRi when no easily measurable metric is observable (i.e., fluorescence). This work was published in Scientific Reports.[95] In Chapter 6, three transporters and two catabolic pathways identified as being relevant to aromatic consumption via transcriptomic analysis were knocked out using parts developed in Chapter 4. The role of these genes was then characterized via growth assays. This work was published in Metabolic Engineering.[71] Chapter 7 summarizes the conclusions of this dissertation and discusses possible future directions for research efforts in *R. opacus*.

1.8 References

1. Prasad, P. V. V.; Thomas, J. M. G.; Narayanan, S., Global Warming Effects. In *Encyclopedia of Applied Plant Sciences (Second Edition)*, Thomas, B.; Murray, B. G.; Murphy, D. J., Eds. Academic Press: Oxford, 2017; pp 289-299.

2. *IPCC Climate Change 2014 Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; IPCC: Geneva, Swizterland, 2014, 2014; p 151.

3. Roth, A.; Riegel, F.; Batteiger, V., Potentials of Biomass and Renewable Energy: The Question of Sustainable Availability. In *Biokerosene: Status and Prospects*, Kaltschmitt, M.; Neuling, U., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2018; pp 95-122.

4. REN21. *Renewables 2016 Global Status Report, Renewable Energy Policy Network for the 21st Century*; 2016.

5. Yang, Y.; Bremner, S.; Menictas, C.; Kay, M., Battery energy storage system size determination in renewable energy systems: A review. *Renewable and Sustainable Energy Reviews* 2018, *91*, 109-125.

6. Ager, J. W.; Lapkin, A. A., Chemical storage of renewable energy. *Science* 2018, *360* (6390), 707-708.

7. Scarlat, N.; Dallemand, J.-F.; Monforti-Ferrario, F.; Banja, M.; Motola, V., Renewable energy policy framework and bioenergy contribution in the European Union – An overview from National Renewable Energy Action Plans and Progress Reports. *Renewable and Sustainable Energy Reviews* 2015, *51*, 969-985.

8. Anawar, H. M.; Strezov, V.; Strezov, V., *Renewable Energy Systems from Biomass*. 1st ed.; CRC Press: Boca Raton, 2018; p 278.

9. Lewis, P.; Karimi, B.; Shan, Y.; Rasdorf, W., Comparing the economic, energy, and environmental impacts of biodiesel versus petroleum diesel fuel use in construction equipment. *International Journal of Construction Education and Research* 2018, 1-15.

10. ElGalad, M. I.; El- Khatib, K. M.; Abdelkader, E.; El-Araby, R.; ElDiwani, G.; Hawash, S. I., Empirical equations and economical study for blending biofuel with petroleum jet fuel. *Journal of Advanced Research* 2018, *9*, 43-50.

11. Kalghatgi, G., Is it really the end of internal combustion engines and petroleum in transport? *Applied Energy* 2018, *225*, 965-974.

12. Elgowainy, A.; Han, J.; Ward, J.; Joseck, F.; Gohlke, D.; Lindauer, A.; Ramsden, T.; Biddy, M.; Alexander, M.; Barnhart, S.; Sutherland, I.; Verduzco, L.; Wallington, T. J., Current and Future United States Light-Duty Vehicle Pathways: Cradle-to-Grave Lifecycle Greenhouse Gas Emissions and Economic Assessment. *Environ. Sci. Technol.* 2018, *52* (4), 2392-2399.

13. Femeena, P. V.; Sudheer, K. P.; Cibin, R.; Chaubey, I., Spatial optimization of cropping pattern for sustainable food and biofuel production with minimal downstream pollution. *J Environ Manage* 2018, *212*, 198-209.

14. Heaton, E. A.; Dohleman, F. G.; Long, S. P., Meeting US biofuel goals with less land: the potential of Miscanthus. *Global Change Biol.* 2008, *14* (9), 2000-2014.

15. Balan, V.; Chiaramonti, D.; Kumar, S., Review of US and EU initiatives toward development, demonstration, and commercialization of lignocellulosic biofuels. *Biofuels, Bioproducts and Biorefining* 2013, *7* (6), 732-759.

16. Boerjan, W.; Ralph, J.; Baucher, M., Lignin biosynthesis. *Annu. Rev. Plant Biol.* 2003, *54* (1), 519-546.

17. Fengel, D.; Wegener, G., *Wood: chemistry, ultrastructure, reactions*. Walter de Gruyter: 1983.

18. Lin, S. Y.; Dence, C. W., *Methods in Lignin Chemistry*. Springer: 1992.

19. 2016 Billion-Ton Report. *U.S. DOE* July 2016.

20. Li, R.; Chen, J., Planning the next-generation biofuel crops based on soil-water constraints. *Biomass Bioenergy* 2018, *115*, 19-26.

21. Pu, Y.; Zhang, D.; Singh, P. M.; Ragauskas, A. J., The new forestry biofuels sector. *Biofuels, Bioproducts and Biorefining* 2008, *2* (1), 58-73.

22. Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J.; Hallett, J. P.; Leak, D. J.; Liotta, C. L., The path forward for biofuels and biomaterials. *Science* 2006, *311* (5760), 484-489.

23. Vanholme, R.; Demedts, B.; Morreel, K.; Ralph, J.; Boerjan, W., Lignin Biosynthesis and Structure. *Plant Physiology* 2010, *153* (3), 895-905.

24. Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD, Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 2007, *315* (5813), 804-807.

25. Ragauskas, A. J.; Beckham, G. T.; Biddy, M. J.; Chandra, R.; Chen, F.; Davis, M. F.; Davison, B. H.; Dixon, R. A.; Gilna, P.; Keller, M., Lignin valorization: improving lignin processing in the biorefinery. *Science* 2014, *344* (6185), 1246843.

26. Northey, R. A., Low-Cost Uses of Lignin. In *Emerging Technologies for Materials and Chemicals from Biomass*, American Chemical Society: 1992; Vol. 476, pp 146-175.

27. Montague, L.; Slayton, A.; Lukas, J. *Lignocellulosic biomass to ethanol process design and economics utilizing co-current dilute acid prehydrolysis and enzymatic hydrolysis for corn stover*; NREL Technical Report 2002: 2002.

28. Humbird, D.; Davis, R.; Tao, L.; Kinchin, C.; Hsu, D.; Aden, A.; Schoen, P.; Lukas, J.; Olthof, B.; Worley, M. *Process design and economics for biochemical conversion of lignocellulosic biomass to ethanol: dilute-acid pretreatment and enzymatic hydrolysis of corn stover*; National Renewable Energy Laboratory (NREL), Golden, CO.: 2011.

29. Nandanwar RA, Chaudhari AR, Ekhe JD, Utilization of Industrial Waste Lignin in Polymer Systems. *International Journal of Knowledge Engineering* 2012, *3* (1).

30. Gellerstedt G, Sjoholm E, Brodin I, The Wood-Based Biorefinery: A Source of Carbon Fiber? *The Open Agriculture Journal* 2010, *3*, 119-124.

31. Regalbuto, J. R., Cellulosic biofuels–got gasoline. *Science* 2009, *325* (5942), 822-824.

32. Kaparaju, P.; Serrano, M.; Thomsen, A. B.; Kongjan, P.; Angelidaki, I., Bioethanol, biohydrogen and biogas production from wheat straw in a biorefinery concept. *Bioresour. Technol.* 2009, *100* (9), 2562-2568.

33. Watkins, D.; Nuruddin, M.; Hosur, M.; Tcherbi-Narteh, A.; Jeelani, S., Extraction and characterization of lignin from different biomass resources. *Journal of Materials Research and Technology* 2015, *4* (1), 26-32.

34. Abdelaziz, O. Y.; Brink, D. P.; Prothmann, J.; Ravi, K.; Sun, M.; García-Hidalgo, J.; Sandahl, M.; Hulteberg, C. P.; Turner, C.; Lidén, G.; Gorwa-Grauslund, M. F., Biological valorization of low molecular weight lignin. *Biotechnol. Adv.* 2016, *34* (8), 1318-1346.

35. Mottiar, Y.; Vanholme, R.; Boerjan, W.; Ralph, J.; Mansfield, S. D., Designer lignins: harnessing the plasticity of lignification. *Current Opinion in Biotechnology* 2016, *37*, 190-200.

36. Gou, J. Y.; Park, S.; Yu, X. H.; Miller, L. M.; Liu, C. J., Compositional characterization and imaging of "wall-bound" acylesters of Populus trichocarpa reveal differential accumulation of acyl molecules in normal and reactive woods. *Planta* 2008, *229* (1), 15-24.

37. Shrotri, A.; Kobayashi, H.; Fukuoka, A., Chapter Two - Catalytic Conversion of Structural Carbohydrates and Lignin to Chemicals. In *Advances in Catalysis*, Song, C., Ed. Academic Press: 2017; Vol. 60, pp 59-123.

38. Roos, C. *Clean heat and power using biomass gasification for industrial and agricultural projects*; U.S. Department of Energy Clean Energy Application Center: 2010; p 64.

39. Naqvi, M.; Yan, J.; Dahlquist, E., Bio-refinery system in a pulp mill for methanol production with comparison of pressurized black liquor gasification and dry gasification using direct causticization. *Applied Energy* 2012, *90* (1), 24-31.

40. Kabalina, N.; Costa, M.; Weihong, Y.; Martin, A. R., Production of Synthetic Natural Gas from Refuse-Derived Fuel Gasification for Use in a Polygeneration District Heating and Cooling System. *Energies* 2016, *9* (12).

41. Mohan, D.; Pittman, Charles U.; Steele, P. H., Pyrolysis of Wood/Biomass for Bio-oil: A Critical Review. *Energy & Fuels* 2006, *20* (3), 848-889.

42. Pacek, A. W.; Ding, P.; Garrett, M.; Sheldrake, G.; Nienow, A. W., Catalytic Conversion of Sodium Lignosulfonate to Vanillin: Engineering Aspects. Part 1. Effects of Processing Conditions on Vanillin Yield and Selectivity. *Industrial & Engineering Chemistry Research* 2013, *52* (25), 8361-8372.

43. Aro, T.; Fatehi, P., Production and Application of Lignosulfonates and Sulfonated Lignin. *ChemSusChem* 2017, *10* (9), 1861-1877.

44. Zakzeski, J.; Bruijnincx, P. C. A.; Jongerius, A. L.; Weckhuysen, B. M., The Catalytic Valorization of Lignin for the Production of Renewable Chemicals. *Chem. Rev.* 2010, *110* (6), 3552-3599.

45. Chen, H.; Liu, J.; Chang, X.; Chen, D.; Xue, Y.; Liu, P.; Lin, H.; Han, S., A review on the pretreatment of lignocellulose for high-value chemicals. *Fuel Process. Technol.* 2017, *160*, 196-206.

46. Bugg, T. D. H.; Rahmanpour, R., Enzymatic conversion of lignin into renewable chemicals. *Curr. Opin. Chem. Biol.* 2015, *29*, 10-17.

47. Bugg, T. D.; Ahmad, M.; Hardiman, E. M.; Rahmanpour, R., Pathways for degradation of lignin in bacteria and fungi. *Nat Prod Rep* 2011, *28* (12), 1883-96.

48. Li, X.; He, Y.; Zhang, L.; Xu, Z.; Ben, H.; Gaffrey, M. J.; Yang, Y.; Yang, S.; Yuan, J. S.; Qian, W.-J.; Yang, B., Discovery of potential pathways for biological conversion of poplar wood into lipids by co-fermentation of Rhodococci strains. *Biotechnology for Biofuels* 2019, *12* (1), 60.

49. Gasser, C. A.; Hommes, G.; Schäffer, A.; Corvini, P. F.-X., Multi-catalysis reactions: new prospects and challenges of biotechnology to valorize lignin. *Applied microbiology and biotechnology* 2012, *95* (5), 1115-1134.

50. Linger, J. G.; Vardon, D. R.; Guarnieri, M. T.; Karp, E. M.; Hunsinger, G. B.; Franden, M. A.; Johnson, C. W.; Chupka, G.; Strathmann, T. J.; Pienkos, P. T.; Beckham, G. T., Lignin valorization through integrated biological funneling and chemical catalysis. *Proceedings of the National Academy of Sciences* 2014, *111*, 12013-12018.

51. Beckham, G. T.; Johnson, C. W.; Karp, E. M.; Salvachúa, D.; Vardon, D. R., Opportunities and challenges in biological lignin valorization. *Curr. Opin. Biotechnol.* 2016, *42*, 40-53.

52. Lynd LR, Van Zyl WH, McBride JE, Laser M, Consolidated bioprocessing of cellulosic biomass: an update. *Curr. Opin. Biotechnol.* 2005, *16* (5), 577-583.

53. Schwartz, T. J.; O'Neill, B. J.; Shanks, B. H.; Dumesic, J. A., Bridging the Chemical and Biological Catalysis Gap: Challenges and Outlooks for Producing Sustainable Chemicals. *ACS Catalysis* 2014, *4*, 2060-2069.

54. Sheldon, R. A., Green and sustainable manufacture of chemicals from biomass: state of the art. *Green Chemistry* 2014, *16*, 950-963.

55. Okamura-Abe, Y.; Abe, T.; Nishimura, K.; Kawata, Y.; Sato-Izawa, K.; Otsuka, Y.; Nakamura, M.; Kajita, S.; Masai, E.; Sonoki, T.; Katayama, Y., Beta-ketoadipic acid and muconolactone production from a lignin-related aromatic compound through the protocatechuate 3,4-metabolic pathway. *J. Biosci. Bioeng.* 2016, *121*, 652-658.

56. Mycroft, Z.; Gomis, M.; Mines, P.; Law, P.; Bugg, T. D. H., Biocatalytic conversion of lignin to aromatic dicarboxylic acids in Rhodococcus jostii RHA1 by re-routing aromatic degradation pathways. *Green Chemistry* 2015, *17*, 4974-4979.

57. Matera, I.; Ferraroni, M.; Kolomytseva, M.; Golovleva, L.; Scozzafava, A.; Briganti, F., Catechol 1,2-dioxygenase from the Gram-positive Rhodococcus opacus 1CP: Quantitative structure/activity relationship and the crystal structures of native enzyme and catechols adducts. *J Struct Biol* 2010, *170*, 548-564.

58. Sonoki, T.; Morooka, M.; Sakamoto, K.; Otsuka, Y.; Nakamura, M.; Jellison, J.; Goodell, B., Enhancement of protocatechuate decarboxylase activity for the effective production of muconate from lignin-related aromatic compounds. *J. Biotechnol.* 2014, *192 Pt A*, 71-77.

59. Sainsbury, P. D.; Hardiman, E. M.; Ahmad, M.; Otani, H.; Seghezzi, N.; Eltis, L. D.; Bugg, T. D., Breaking down lignin to high-value chemicals: the conversion of lignocellulose to vanillin in a gene deletion mutant of Rhodococcus jostii RHA1. *ACS chemical biology* 2013, *8* (10), 2151-2156.

60. Salvachúa, D.; Johnson, C. W.; Singer, C. A.; Rohrer, H.; Peterson, D. J.; Black, B. A.; Knapp, A.; Beckham, G. T., Bioprocess development for muconic acid production from aromatic compounds and lignin. *Green Chemistry* 2018, *20* (21), 5007-5019.

61. Becker, J.; Kuhl, M.; Kohlstedt, M.; Starck, S.; Wittmann, C., Metabolic engineering of Corynebacterium glutamicum for the production of cis, cis-muconic acid from lignin. *Microbial cell factories* 2018, *17* (1), 115.

62. Shi, Y.; Yan, X.; Li, Q.; Wang, X.; Xie, S.; Chai, L.; Yuan, J., Directed bioconversion of Kraft lignin to polyhydroxyalkanoate by Cupriavidus basilensis B-8 without any pretreatment. *Process Biochemistry* 2017, *52*, 238-242.

63. Bradfield, M. F.; Mohagheghi, A.; Salvachúa, D.; Smith, H.; Black, B. A.; Dowe, N.; Beckham, G. T.; Nicol, W., Continuous succinic acid production by Actinobacillus succinogenes on xylose-enriched hydrolysate. *Biotechnology for biofuels* 2015, *8* (1), 181.

64. Le, R. K.; Wells Jr, T.; Das, P.; Meng, X.; Stoklosa, R. J.; Bhalla, A.; Hodge, D. B.; Yuan, J. S.; Ragauskas, A. J., Conversion of corn stover alkaline pre-treatment waste streams into biodiesel via Rhodococci. *RSC Advances* 2017, *7* (7), 4108-4115.

65. Herrero, O. M.; Alvarez, H. M., Whey as a renewable source for lipid production by Rhodococcus strains: Physiology and genomics of lactose and galactose utilization. *European Journal of Lipid Science and Technology* 2016, *118* (2), 262-272.

66. Wei, Z.; Zeng, G.; Kosa, M.; Huang, D.; Ragauskas, A. J., Pyrolysis oil-based lipid production as biodiesel feedstock by Rhodococcus opacus. *Appl. Biochem. Biotechnol.* 2015, *175* (2), 1234-1246.

67. Pu, Y.; Zhang, D.; Singh, P. M.; Ragauskas, A. J., The new forestry biofuels sector. *Biofuels, Bioproducts and Biorefining* 2008, *2*, 58-73.

68. Beckham, G. T.; Johnson, C. W.; Karp, E. M.; Salvachua, D.; Vardon, D. R., Opportunities and challenges in biological lignin valorization. *Curr. Opin. Biotechnol.* 2016, *42*, 40-53.

69. Holder, J. W.; Ulrich, J. C.; DeBono, A. C.; Godfrey, P. A.; Desjardins, C. A.; Zucker, J.; Zeng, Q.; Leach, A. L. B.; Ghiviriga, I.; Dancel, C.; Abeel, T.; Gevers, D.; Kodira, C. D.; Desany, B.; Affourtit, J. P.; Birren, B. W.; Sinskey, A. J., Comparative and Functional Genomics of Rhodococcus opacus PD630 for Biofuels Development. *PLoS Genet* 2011, *7*, e1002219.

70. Kurosawa, K.; Laser, J.; Sinskey, A. J., Tolerance and adaptive evolution of triacylglycerol-producing Rhodococcus opacus to lignocellulose-derived inhibitors. *Biotechnology for Biofuels* 2015, *8* (76).

71. Henson, W. R.; Campbell, T.; DeLorenzo, D. M.; Gao, Y.; Berla, B.; Kim, S. J.; Foston, M.; Moon, T. S.; Dantas, G., Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus. *Metab. Eng.* 2018, *49*, 69-83.

72. Kurosawa, K.; Wewetzer, S. J.; Sinskey, A. J., Engineering xylose metabolism in triacylglycerol-producing Rhodococcus opacus for lignocellulosic fuel production. *Biotechnology for Biofuels* 2013, *6* (1), 134.

73. Hetzler, S.; Steinbüchel, A., Establishment of Cellobiose Utilization for Lipid Production in Rhodococcus opacus PD630. *Appl. Environ. Microbiol.* 2013, *79*, 3122-3125.

74. Kurosawa, K.; Plassmeier, J.; Kalinowski, J.; Ruckert, C.; Sinskey, A. J., Engineering L-arabinose metabolism in triacylglycerol-producing Rhodococcus opacus for lignocellulosic fuel production. *Metab. Eng.* 2015, *30*, 89-95.

75. Hetzler, S.; Broker, D.; Steinbuchel, A., Saccharification of Cellulose by Recombinant Rhodococcus opacus PD630 Strains. *Appl Environ Microbiol* 2013, *79* (17), 5159-66.

76. He, Y.; Li, X.; Ben, H.; Xue, X.; Yang, B., Lipid Production from Dilute Alkali Corn Stover Lignin by Rhodococcus Strains. *ACS Sustainable Chemistry & Engineering* 2017, *5* (3), 2302-2311.

77. Zhao, C.; Xie, S.; Pu, Y.; Zhang, R.; Huang, F.; Ragauskas, A. J.; Yuan, J. S., Synergistic enzymatic and microbial lignin conversion. *Green Chemistry* 2016, *18* (5), 1306-1312.

78. Xie, S.; Sun, S.; Lin, F.; Li, M.; Pu, Y.; Cheng, Y.; Xu, B.; Liu, Z.; da Costa Sousa, L.; Dale, B. E.; Ragauskas, A. J.; Dai, S. Y.; Yuan, J. S., Mechanism-Guided Design of Highly Efficient Protein Secretion and Lipid Conversion for Biomanufacturing and Biorefining. *Advanced Science 0* (0), 1801980.

79. Chen, Y.; Ding, Y.; Yang, L.; Yu, J.; Liu, G.; Wang, X.; Zhang, S.; Yu, D.; Song, L.; Zhang, H.; Zhang, C.; Huo, L.; Huo, C.; Wang, Y.; Du, Y.; Zhang, H.; Zhang, P.; Na, H.; Xu, S.; Zhu, Y.; Xie, Z.; He, T.; Zhang, Y.; Wang, G.; Fan, Z.; Yang, F.; Liu, H.; Wang, X.; Zhang, X.; Zhang, M. Q.; Li, Y.; Steinbuchel, A.; Fujimoto, T.; Cichello, S.; Yu, J.; Liu, P., Integrated omics study delineates the dynamics of lipid droplets in Rhodococcus opacus PD630. *Nucleic Acids Res* 2014, *42* (2), 1052-64.

80. Alvarez, H. M.; Mayer, F.; Fabritius, D.; Steinbuchel, A., Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630. *Arch Microbiol* 1996, *165* (6), 377-86.

81. Yoneda, A.; Henson, W. R.; Goldner, N. K.; Park, K. J.; Forsberg, K. J.; Kim, S. J.; Pesesky, M. W.; Foston, M.; Dantas, G.; Moon, T. S., Comparative transcriptomics elucidates

adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630. *Nucleic Acids Res* 2016, *44* (5), 2240-54.

82. Henson, W. R.; Hsu, F. F.; Dantas, G.; Moon, T. S.; Foston, M., Lipid metabolism of phenol-tolerant Rhodococcus opacus strains for lignin bioconversion. *Biotechnology for Biofuels* 2018, *11*, 339.

83. Hollinshead, W. D.; Henson, W. R.; Abernathy, M.; Moon, T. S.; Tang, Y. J., Rapid metabolic analysis of Rhodococcus opacus PD630 via parallel 13C-metabolite fingerprinting. *Biotechnol. Bioeng.* 2016, *113* (1), 91-100.

84. Kurosawa, K.; Boccazzi, P.; de Almeida, N. M.; Sinskey, A. J., High-cell-density batch fermentation of Rhodococcus opacus PD630 using a high glucose concentration for triacylglycerol production. *J. Biotechnol.* 2010, *147*, 212-218.

85. Herrero, O. M.; Moncalián, G.; Alvarez, H. M., Physiological and genetic differences amongst Rhodococcus species for using glycerol as a source for growth and triacylglycerol production. 2016, *162* (2), 384-397.

86. Goswami, L.; Tejas Namboodiri, M. M.; Vinoth Kumar, R.; Pakshirajan, K.; Pugazhenthi, G., Biodiesel production potential of oleaginous Rhodococcus opacus grown on biomass gasification wastewater. *Renewable Energy* 2017, *105*, 400-406.

87. Lanfranconi, M. P.; Alvarez, H. M., Rewiring neutral lipids production for the de novo synthesis of wax esters in Rhodococcus opacus PD630. *J Biotechnol* 2017, *260*, 67-73.

88. Vardon, D. R.; Franden, M. A.; Johnson, C. W.; Karp, E. M.; Guarnieri, M. T.; Linger, J. G.; Salm, M. J.; Strathmann, T. J.; Beckham, G. T., Adipic acid production from lignin. *Energy & Environmental Science* 2015, *8*, 617-628.

89. Kruyer, N. S.; Peralta-Yahya, P., Metabolic engineering strategies to bio-adipic acid production. *Curr. Opin. Biotechnol.* 2017, *45*, 136-143.

90. Sun, X.; Lin, Y.; Yuan, Q.; Yan, Y., Biological production of muconic acid via a prokaryotic 2,3-dihydroxybenzoic acid decarboxylase. *ChemSusChem* 2014, *7* (9), 2478-81.

91. Suwaleerat, T.; Thanapimmetha, A.; Srisaiyoot, M.; Chisti, Y.; Srinophakun, P., Enhanced production of carotenoids and lipids by Rhodococcus opacus PD630. 2018, *93* (8), 2160-2169.

92. Hetzler, S.; Steinbuchel, A., Establishment of cellobiose utilization for lipid production in Rhodococcus opacus PD630. *Appl Environ Microbiol* 2013, *79* (9), 3122-5.

93. DeLorenzo, D. M.; Henson, W. R.; Moon, T. S., Development of Chemical and Metabolite Sensors for Rhodococcus opacus PD630. *ACS synthetic biology* 2017, *6* (10), 1973-1978.

94. DeLorenzo, D. M.; Rottinghaus, A. G.; Henson, W. R.; Moon, T. S., Molecular Toolkit for Gene Expression Control and Genome Modification in Rhodococcus opacus PD630. *ACS synthetic biology* 2018, *7* (2), 727-738.

95. DeLorenzo, D. M.; Moon, T. S., Selection of stable reference genes for RT-qPCR in Rhodococcus opacus PD630. *Sci Rep* 2018, *8* (1), 6019.

Chapter 2: Development of chemical and metabolite sensors for *Rhodococcus opacus* PD630

An abridged version of this chapter was published as DeLorenzo, D.M., Henson, W.R., Moon, T.S. Development of Chemical and Metabolite Sensors for *Rhodococcus opacus* PD630. *ACS Synthetic Biology* (2017). 6, 1973–1978. Reprinted with permission.

Genetic elements for predictable, tunable, and quantifiable gene expression are a critical aspect for successfully engineering a host. In this chapter, multiple fluorescent reporter genes were screened to assist in quantifying gene expression in *Rhodococcus opacus* PD630. Furthermore, multiple genetic sensors that detect a variety of chemical and metabolite ligands were characterized in this host for tunable gene expression. These tools, such as the developed aromatic and ammonium sensors, could be utilized in the future to implement dynamic regulation. Additionally, one sensor was demonstrated to function as a delayed onset gene expression system or cellular timer. This chapter substantially adds to the *R. opacus* genetic toolbox and will enable future engineering efforts. I performed all experiments and wrote the manuscript.

2.1 Abstract

*Rhodococcus opac*us PD630 is a non-model, gram-positive bacterium that possesses desirable traits for biomass conversion, including consumption capabilities for lignocellulose-based sugars and toxic lignin-derived aromatic compounds, significant triacylglycerol accumulation, relatively rapid growth rate, and genetic tractability. However, few genetic elements have been directly characterized in *R. opacus,* limiting its application for lignocellulose bioconversion. In this letter, we report the characterization and development of genetic tools for tunable gene expression in *R. opacus*, including: 1) six fluorescent reporters for quantifying

promoter output, 2) three chemically inducible promoters for variable gene expression, and 3) two classes of metabolite sensors derived from native *R. opacus* promoters that detect nitrogen levels or aromatic compounds. Using these tools, we also provide insights into native aromatic consumption pathways in *R. opacus*. Overall, this work expands the ability to control and characterize gene expression in *R. opacus* for future lignocellulose-based fuel and chemical production.

2.2 Introduction

Lignocellulosic biomass represents an abundant and renewable resource that could displace petroleum feedstocks, but it requires the development of more efficient conversion processes for commercialization.[1] One approach uses microbe-based biorefineries to produce value-added products from plant material. For economic viability, the entire biomass (cellulose, hemicellulose and lignin) must be consumed.[1-3] The utilization of lignin (~15-30% of plant matter by dry weight) remains a challenge as it is a complex, cross-linked, and highly heterogeneous aromatic macromolecule; lignin is mostly burned for low-value process heat or discarded.[1, 3, 4] When lignin is catalytically depolymerized, it produces a wide range of toxic aromatics.[5] These compounds are not directly useful due to the extensive separation costs required to purify any desirable products. To refine this complex product stream, depolymerized lignin can be fed to biological catalysts (i.e., microbes) that use enzymatic pathways to "biologically funnel" aromatics into core intermediates.[3, 6] These intermediates can then be selectively converted into valuable, high titer compounds.[7] Additionally, microbial production of compounds, typically derived from petroleum feedstocks, could significantly diminish greenhouse gas emissions.[8-10]

The Actinomycetales bacterium *Rhodococcus opacus* PD630 (hereafter *R. opacus*) has many desirable traits for the microbial valorization of lignocellulosic biomass. The depolymerization of

lignocellulose, particularly lignin, generates a spectrum of growth inhibiting compounds.[1-3] *R. opacus* has previously demonstrated a high natural tolerance towards such products, including biomass pretreatment by-products (e.g., furfural and hydroxymethylfurfural [HMF]), polychlorinated biphenyls (PCBs), and aromatic and halogenated compounds.[11-13] This tolerance is partially attributed to numerous enzymatic consumption pathways that funnel aromatics into central metabolism via its high flux β-ketoadipate pathway.[4, 14] Intermediates in the β -ketoadipate pathway can be diverted into high-value products, such as adipic acid.[8-10] *R. opacus* can also concurrently consume glucose (via Entner–Doudoroff pathway), acetate (via glyoxylate shunt), and phenolic compounds (via β-ketoadipate pathway), improving productivity.[15, 16] Its high osmotic tolerance also allows for a high loading of sugars and high product titers.[17] *R. opacus* is primarily known for its ability to produce significant amounts of the biodiesel precursor triacylglycerol (TAG; up to ~78% of its cellular dry weight when grown on gluconate).[13] Both alkali lignin and biomass gasification wastewater have successfully been converted into TAG-based biofuels using *R. opacus.*[18, 19] These traits make *R. opacus* an ideal microbial biorefinery for the conversion of lignocellulosic material, particularly lignin, into fuels and chemicals.

Microbial tolerance to and utilization of lignocellulose degradation products can be further improved through rational engineering and adaptive evolution.[14, 20] Previous work has demonstrated the application of adaptive evolution to improve utilization of lignin monomers, such as phenol, in *R. opacus*.[14] Additionally, *R. opacus* has been engineered to consume other lignocellulose components, including xylose, arabinose, and cellobiose, but the modifications were achieved using borrowed and generally uncharacterized genetic elements from related Actinomycetales.[16, 21-23] Optimization of gene expression and dynamic regulation have both been demonstrated to significantly improve product titers, yields and productivity, but this process

requires well-characterized genetic tools to precisely control gene expression.[24-26] Promoters and dynamic sensors, often taken for granted in model organisms, are limited in *R. opacus*, but this work expands the available genetic toolbox for future engineering efforts.

To aid in quantifying gene expression and promoter characterization, we first characterized several fluorescent reporters (CFP, RFP, mCherry, GFP+, sfGFP, and EYFP). Next, we characterized chemically inducible promoters (pBAD, pAcet, and pTet; induced by unmetabolized chemicals in *R. opacus*) from other bacterial species and optimized the dynamic range of the pTet promoter by tuning repressor protein expression. Finally, we developed two classes of metabolite sensors that detect either nitrogen levels or aromatic compounds. A subset of these tools was applied to provide insights into native, aromatic consumption pathways in *R. opacus*. Together, these genetic tools provide the groundwork for the further development of *R. opacus* as a new chassis for fuel and chemical production from lignocellulose.

2.3 Results and discussion

2.3.1 Fluorescent reporters

Developing next generation synthetic biology tools in *R. opacus*, such as promoters and systems of gene control, requires easily discernable reporters. Previous work has tested *lacZ* and GFP in *R. opacus*.[27] However, *lacZ* requires cell lysis and an enzymatic assay, which does not facilitate real time measurement. Additionally, while GFP has been used, it is unknown whether this reporter is ideal in *R. opacus*. To determine which reporters generate the highest fluorescence relative to the background fluorescence (empty vector control strain), six reporter proteins with three distinct wavelength ranges (CFP, RFP, mCherry, GFP+, sfGFP, and EYFP) were tested under the same constitutive promoter and ribosome binding site (RBS; Figure 2.1). mCherry demonstrated the highest level of fluorescence relative to the empty vector control (~452-fold),

and cells were visibly red (Supplementary Figure A.1). In summary, we have demonstrated the outputs of multiple fluorescent reporters that span a wide range of excitation and emission wavelengths for quantifying expression (Supplementary Table A.8).

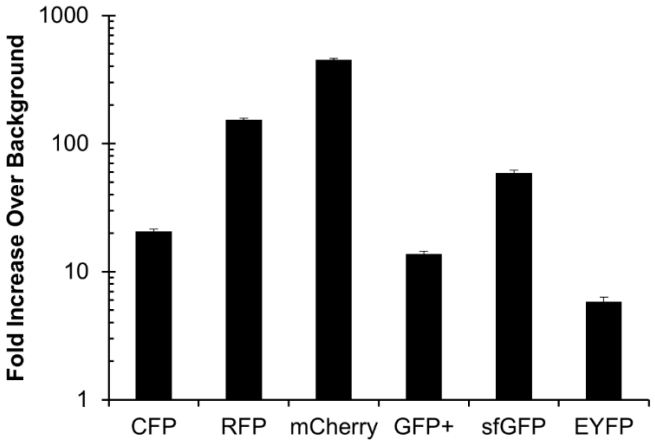


Figure 2.1. Fluorescent reporters in *R. opacus*. CFP, RFP, mCherry, GFP+, sfGFP, and EYFP were all expressed under the same promoter and ribosome binding site (RBS). Fold increase over background represents the absorbance normalized fluorescence of the reporter strain (Fluosample / Abs600sample) divided by the absorbance normalized fluorescence of the control strain containing the empty vector (Fluocontrol / Abs600control) (see Methods). Values represent the average of three replicates grown in minimal media A, and error bars represent one standard deviation.

2.3.2 Chemically inducible promoters

The ability to precisely regulate the quantity of gene product per cell is critical to synthetic control systems and metabolic engineering. Chemically inducible promoters allow an external, non-feedstock signal to modulate the level of transcription of the downstream gene. Presently, only one chemically inducible promoter has been previously described in *R. opacus*: the thiostrepton inducible TipA promoter.[(28)] Thiostrepton is an antibiotic that inhibits growth of *R. opacus*, and thus this promoter also requires a thiostrepton resistance marker. To further improve the selection of available genetic tools, we characterized three additional chemically inducible promoters in *R.*

opacus: an arabinose inducible promoter from *E. coli* (pBAD), an acetamide inducible promoter from *Mycobacterium smegmatis* (pAcet), and an anhydrotetracycline (aTc) inducible promoter from *E. coli* (pTet).

The arabinose degradation operon in *E. coli* is controlled by the arabinose inducible pBAD promoter, which is regulated by the AraC transcription factor and cAMP-cAMP receptor protein (cAMP-CRP) complex.[29] pBAD has been extensively studied for several decades due to its complex repression and induction mechanisms, and AraC homologs exist across bacterial species.[29] To test the application of this promoter in *R. opacus*, the pBAD promoter was placed in front of EYFP on the pAL5000 backbone and induced with arabinose (0 to 250 mM).[30] We observed a ~59-fold increase in fluorescence (Figure 2.2A) and found that arabinose had no effect on cell growth (measured by final culture absorbance at 600 nm; |logarithmic regression coefficient| < 0.01; Supplementary Figure A.2). These results suggest that arabinose does not support or inhibit growth at these concentrations, which is consistent with previous reports.[22]

The acetamide degradation enzyme, AmiE, in *M. smegmatis* is controlled by the pAcet promoter, the regulators AmiA, AmiC, and AmiD, and the transporter AmiS.[31] pAcet has been previously demonstrated as a constitutive promoter in *R. opacus*.[23] To improve characterization of this promoter, the entire acetamide degradation operon was expressed on pAL5000 (*amiA*, *amiC*, *amiD*, and *amiS*), and the *amiE* gene was replaced with GFP+. When acetamide was added to the culture (0 to 1000 nM), we observed a ~5-fold increase in fluorescence (Figure 2.2B), compared to a ~100-fold increase in *M. smegmatis*.[31] Acetamide at these concentrations was found to have no effect on final culture absorbance at 600 nm of the empty vector control strain, but led to a minor growth reduction in the strain harboring the pAcet construct at higher inducer concentrations (Supplementary Figure A.3). This result suggests that acetamide does not directly

affect *R. opacus* growth, but that the expression of GFP+ and the transcriptional regulators/transporter imposes burdens on the cells at higher induction levels. The discrepancy between the observed dynamic ranges in *R. opacus* and *M. smegmatis* reinforces the need for characterization of parts in new hosts.

The tetracycline (Tc) inducible pTet promoter is regulated by the tetracycline repressor (TetR), which controls the expression of the Tc export protein TetA in *E. coli*.[32] Anhydrotetracycline (aTc) is a derivative of Tc that exhibits reduced antibiotic activities and binds to TetR efficiently.[32] Ehrt et al. (2005) and Rock et al. (2017) developed pTet promoters for the Actinomycetales *M. smegmatis* and *M. tuberculosis*.[33, 34] Both of these promoters were screened in *R. opacus*, but they exhibited non-ideal traits, such as low inducibility and either leakiness or over-repression (data not shown). While both *Mycobacterium* species are members of the Actinomycetales order with *R. opacus*, there are clear differences between the organisms, which again reinforces the need to characterize genetic elements within each specific host.

To optimize an inducible pTet system for *R. opacus*, the pTet from Rock et al. (2017) was used as a starting point for the creation of a mutagenesis library.[34] Due to a significant growth defect and the inability to achieve appreciable induction (~2.5-fold), it was hypothesized that the TetR repressor was being too highly expressed. The -10 region of the constitutive TetR promoter underwent saturation mutagenesis to reduce transcription. After screening colonies for improvements in growth and dynamic range, the best candidate exhibited no growth defect relative to the empty vector control and demonstrated a ~67-fold increase in fluorescence when aTc was added to the culture (0.05 to 100 ng/mL; Figure 2.2C). The -10 region of the corresponding promoter changed from TATAAT to ACCTCT. aTc at these concentrations was found to have no

effect on final culture absorbance at 600 nm, suggesting that aTc does not support or inhibit growth at these concentrations (|logarithmic regression coefficient| < 0.01; Supplementary Figure A.4).

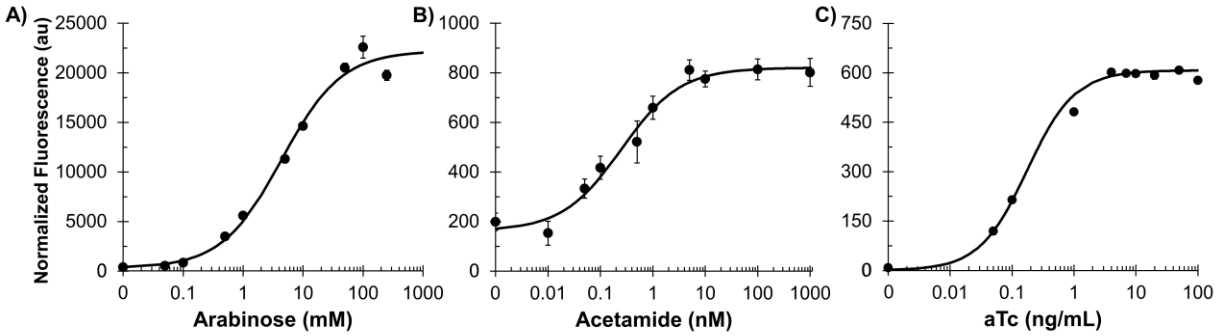


Figure 2.2. Chemically inducible promoters in *R. opacus*. **A)** Transfer curve for the arabinose inducible pBAD promoter driving EYFP production (~59-fold).[30] **B)** Transfer curve for the acetamide inducible pAcet promoter driving GFP+ production (~5-fold).[31] **C)** Transfer curve for the optimized aTc inducible pTet promoter driving mCherry production (~67-fold).[34] Values represent the average of three replicates grown in minimal media A with 1 g/L glucose (pBAD), minimal media A with 2 g/L glucose (pAcet), or minimal media B with 2 g/L glucose (pTet), and error bars represent one standard deviation. Lines represent the fitted transfer functions (see Supplementary Table A.4 for fitted parameters).

2.3.3 Nitrogen metabolite sensor

Chemically inducible promoters can precisely tune gene expression, but their use is limited in industrial scale fermentations due to prohibitive inducer costs. In contrast, metabolite sensors can dynamically regulate gene expression based on concentrations of metabolites. Dynamic pathway regulation has previously been demonstrated to improve final product titers, yields, and productivity.[25] One of *R. opacus*' potential products is triacylglycerol (TAG), a biodiesel precursor, and its production is increased when cells are grown in low nitrogen conditions. Lipid biosynthesis pathways can also be engineered to synthesize fatty acids of different chain lengths, saturation degrees, or chain branching types by introducing heterologous genes.[35] A nitrogen

sensor would allow heterologous genes to be expressed only under lipid production conditions, thus conserving cellular resources and improving product titers and productivity.

We tested seven *R. opacus* promoters with the hope of identifying a nitrogen responsive promoter (Supplementary Figure A.5). Although the corresponding seven genes were found to be highly transcribed under low nitrogen conditions (from RNA-sequencing),[(14)] RNA-sequencing was not performed under high nitrogen conditions in that study, requiring further tests. To this end, the upstream regions of these seven genes were placed in front of GFP+ and screened with either 0.05 or 2.0 g/L ammonium sulfate (Supplementary Figure A.5). The promoter region of LPD03031 (pLPD03031), which encodes a putative nitrite extrusion protein, was found to be induced ~18-fold in fluorescence under low nitrogen stress. To further characterize this promoter, a range of initial ammonium sulfate concentrations (0.05 to 0.8 g/L) with 10 g/L glucose were used to examine expression over time (Figure 2.3A). The response time of the promoter (time to reach the half maximum fluorescence) was linearly correlated (R^2 = 0.99; Figure 2.3B) with the initial ammonium sulfate concentration. In this way, the nitrogen sensor can be used as a cellular timer based on how much nitrogen is initially added to the culture. As the cells consume the available nitrogen, the promoter will eventually be induced once nitrogen stress occurs. A saturating amount of ammonium sulfate (2.5 g/L) prevented induction within the tested time-frame, demonstrating that the promoter is indeed responsive to low nitrogen levels and not to growth phase (Supplementary Figure A.6).

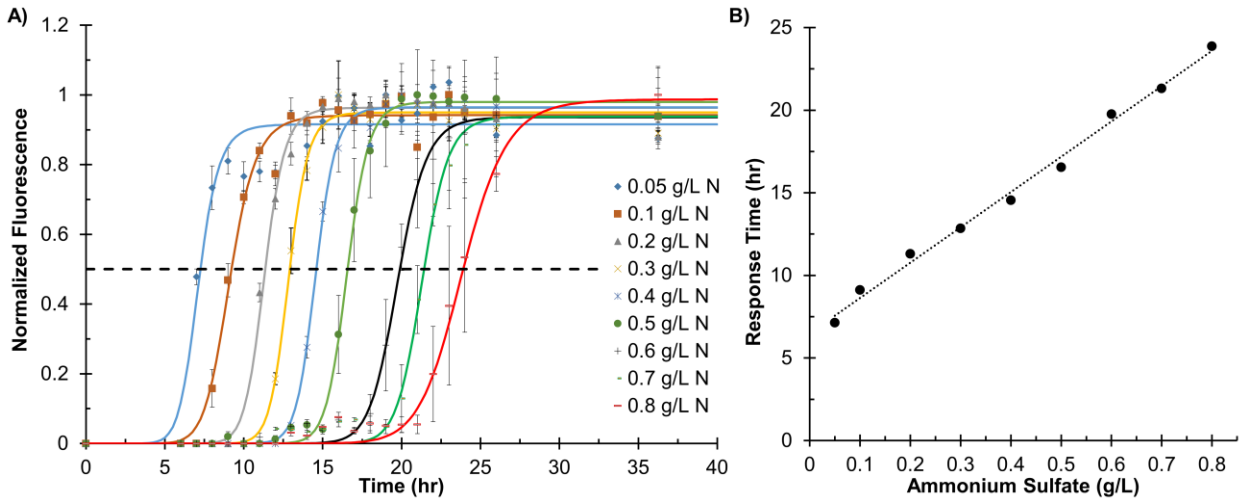


Figure 2.3. Nitrogen sensor in *R. opacus*. **A)** The upstream region of LPD03031 (putative nitrite extrusion protein) was cloned in front of GFP+. Cultures were grown with 10 g/L glucose and either 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, or 0.8 g/L ammonium sulfate (N). The normalized fluorescence (see Methods) was scaled to the maximum value for each respective condition (scale 0 to 1). Values represent the average of three replicates grown in minimal media B, and error bars represent one standard deviation. Continuous lines represent the fitted transfer functions (see Supplementary Table A.4 for fitted parameters). **B)** The response time (time required to reach the half-maximal fluorescence; dashed line in 3A) is directly correlated to the initial ammonium sulfate concentration (R^2 = 0.99; the dotted line represents linear fit).

2.3.4 Aromatic metabolite sensors and insights into aromatic consumption pathways

R. opacus is a promising candidate for the valorization of lignocellulose because of its ability to consume aromatic monomers. A constituent of lignocellulose is lignin (10-30% by dry weight), which when depolymerized produces a large distribution of aromatic compounds.[4, 6] Our previous transcriptomic analysis found that genes associated with aromatic transport and consumption were upregulated in the presence of phenol when compared to a glucose control.[14] To further investigate this upregulation of phenol-related consumption pathways, the upstream regions of four of these genes were placed in front of GFP+ (LPD06568 [putative catechol 1,2-dioxygenase], LPD06575 [putative small subunit of two-component phenol hydroxylase, copy 1], LPD06699 [putative shikimate transporter], and LPD06740 [putative small subunit of two

component phenol hydroxylase, copy 2]). The fluorescence was measured with the addition of 1 g/L glucose and either 0, 0.1, 0.3, or 0.75 g/L phenol (Figure 2.4). Fold inductions (0.75 g/L phenol vs. 0 g/L phenol) were 80-fold (pLPD06568), 39-fold (pLPD06699), and 247-fold (pLPD06740). The fluorescence output of pLPD06575 in the glucose-only condition was "completely off", meaning that it was indistinguishable from the background fluorescence (within one standard deviation), and thus a fold change relative to the "off" state could not be calculated.

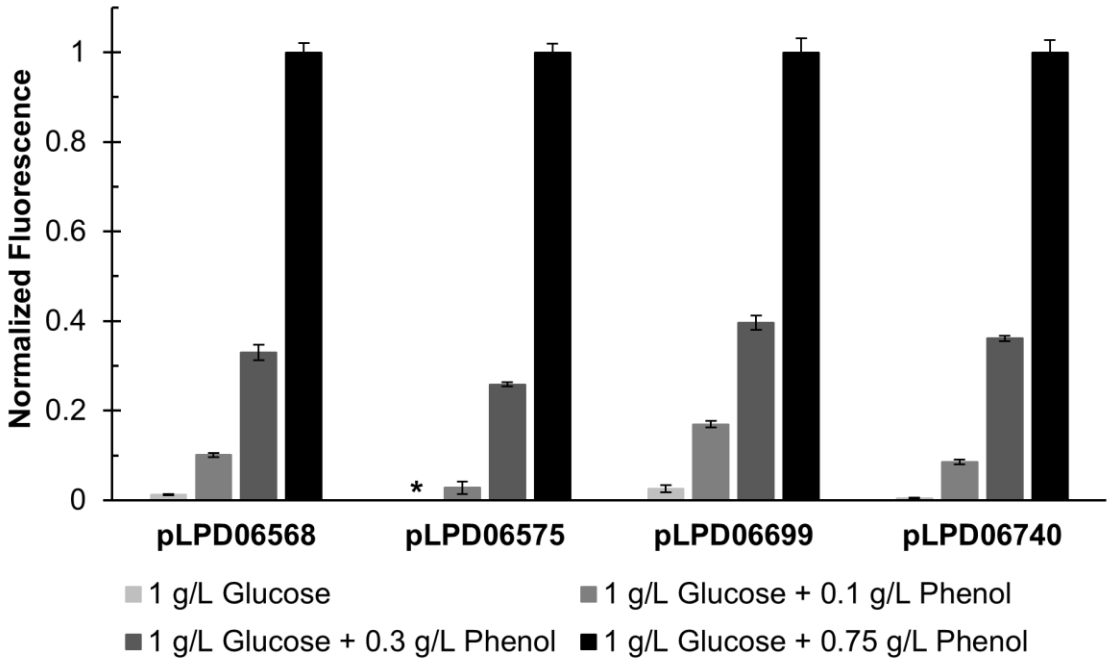


Figure 2.4. Four phenol sensors characterized using GFP+. The strains were grown with 1 g/L glucose and either 0, 0.1, 0.3, or 0.75 g/L phenol. Because increasing concentrations of phenol led to lengthening of the cell culture lag phase, early stationary phase fluorescence was measured at different time points (see Supplementary Table A.5 and Supplementary Figure A.7). 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 grown in minimal media A, and error bars represent one standard deviation.

The mechanisms of regulation of each of these characterized phenol sensors are not well understood. Aromatic metabolic pathways and related genes (Figure 2.5A) in Actinobacteria are in general regulated by numerous transcriptional regulators (TRs) that interact directly with the aromatic compounds and the DNA.[36, 37] These regulators frequently occur in close genomic context to their primary targets (Figure 2.5B). Some of these transcriptional regulators can act as activators or repressors, depending on the location of the binding site within the promoter region.[36] Additionally, some regulators have been demonstrated to possess multiple substrate binding sites and the ability to bind to promoters across various enzymatic pathways to create a complex network of interconnected regulation that fosters hierarchical consumption of compounds and catabolite repression.[36, 37]

It was demonstrated in *Rhodococcus erythropolis* CCM2595 that the promoter of the two-component phenol hydroxylase operon (*pheA2* [small subunit] and *pheA1* [large subunit]) is activated by the AraC family regulator PheR when it binds to phenol.[37] Both LPD06575 and LPD06740 (copy 1 and 2 of putative small subunit of two-component phenol hydroxylases) and LPD06574 and LPD06739 (copy 1 and 2 of putative AraC family regulators) show high nucleotide and amino acid sequence similarity to *pheA2* and *pheR*, respectively (Figure 2.5B and Supplementary Table A.6). The promoters pLPD06575 and pLPD06740 also share high sequence conservation with the *pheA2* upstream region (Supplementary Table A.6). It was also demonstrated that the three gene *catABC* operon is repressed by the adjacent IcIR family regulator CatR in *R. erythropolis* CCM2595,[37] which all exhibit high nucleotide and amino acid sequence similarity to the four gene cluster LPD06566 to LPD06569 (Figure 2.5B and Supplementary Table A6). The promoter pLPD06568 also shares high sequence conservation with the *catABC* operon upstream region (Supplementary Table A.6). The regulatory mechanism of LPD06699, encoding

a shikimate transporter, is unknown, but an IcIR family regulator (LPD06698) is located adjacent to it.

We sought to examine how a functionally diverse set of aromatic monomers, typically found in depolymerized lignin, would affect each of these promoters (Figure 2.5). Each construct was tested in the presence of either 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). All conditions also contained 1 g/L glucose (GLU). Interestingly, the tested promoters exhibited differential expression based on the aromatic added, and the response generally corresponds to the original annotated gene function. For example, both phenol hydroxylase promoters (pLPD06575 and pLPD06740) were substantially induced only in response to phenol. Comparatively, the promoter for the shikimate transporter (pLPD06699), which is expected to aid in general aromatic transport, was upregulated in the presence of all tested compounds. The catechol degradation operon promoter (pLPD06568) was most substantially upregulated by compounds known to be metabolized thorough the catechol branch of the β-ketoadipate pathway (PHE, BEN, and GUA) and minimally affected by compounds (PCA, HBA, and VAN) that are degraded via the parallel protocatechuic acid branch.[38-40] This work represents the first demonstration of compound-specific, inducible aromatic consumption pathways in *R. opacus*.

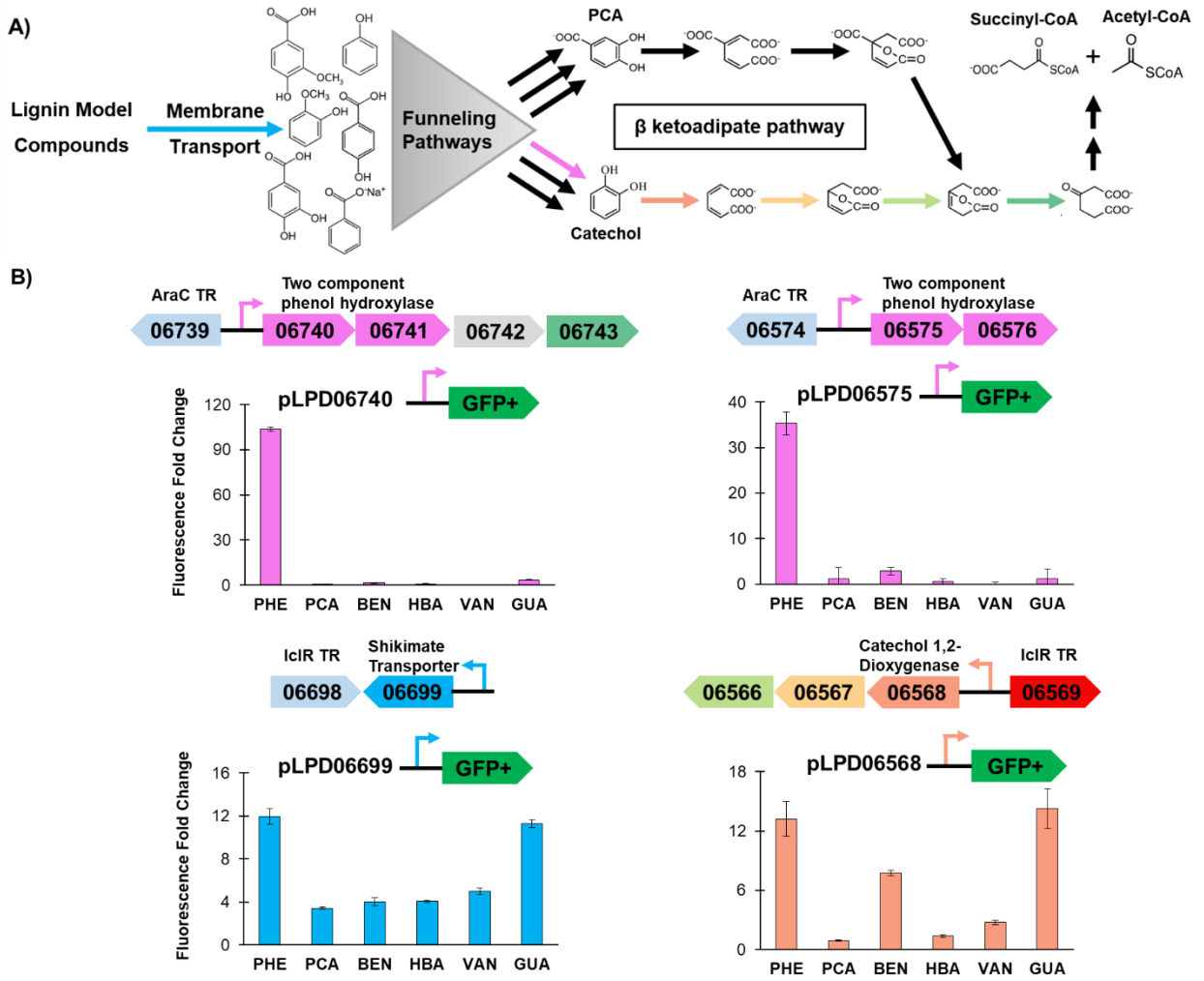


Figure 2.5. Differential response of native *R. opacus* promoters to distinct aromatic monomers. **A)** Schematic of lignin model compound conversion to central metabolites via aromatic funneling enzymes and the central β-ketoadipate pathway. Arrow colors correspond to those of genes and promoters in 5B, and black arrows are for conversion steps whose genes are not shown in 5B. **B)** 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 (See Supplementary Table A.7 and Supplementary Figure A.7) relative to the glucose only condition (GLU). Bar chart colors correspond to functional steps in 5A and promoters in 5B. Values represent the average of three replicates grown in minimal media B, and error bars represent one standard deviation.

2.4 Conclusion

R. opacus is an ideal candidate for conversion of lignocellulose, particularly lignin, due to its natively high tolerance to depolymerization products and ability to consume a broad range of functionally diverse aromatic monomers.[11-14] To develop *R. opacus* as a new chassis for lignocellulose conversion, we first expanded the suite of available fluorescence reporters (CFP, RFP, mCherry, GFP+, sfGFP, and EYFP). To date, the tool to express genes in *R. opacus* has been limited to constitutive promoters or a thiostrepton inducible promoter.[16, 28] In this work, we characterized three chemically inducible promoters (pBAD, pAcet, and pTet) in *R. opacus* and demonstrated expression differences between highly related species, reinforcing the need for re-characterization of genetic tools. The pTet promoter was engineered to reduce growth defects caused by overexpression of the TetR repressor and to maximize dynamic range (2.5-fold to 67-fold). Additionally, two classes of metabolite sensors were developed. The first is a nitrogen sensor that is induced when *R. opacus* cells undergo nitrogen stress. The second is a set of promoters that are induced in the presence of phenol or other aromatic monomers. These aromatic sensors were applied to provide insights into native aromatic consumption pathways in *R. opacus*. The tools developed herein set a foundation for more controlled expression of genes for metabolic engineering in *R. opacus* and a step towards enabling this organism to become industrially relevant.

2.5 Methods

2.5.1 Strains, plasmids and growth conditions

All plasmids were assembled in *E. coli* DH10B[41] using either Gibson or Golden Gate assembly methods.[42, 43] Kanamycin (20 μg/mL) was added as appropriate. Electrocompetent *E. coli* DH10B and *R. opacus* cells were transformed as previously described (See Supplementary

Materials and Methods A).[41, 44] Plasmid DNA was isolated using Zyppy Plasmid Miniprep Kit (Zymo) and PCR products were extracted from electrophoresis gels using Zymoclean Gel DNA Recovery Kit (Zymo). The TetR promoter sequence underwent saturation mutagenesis at the -10 region using degenerate oligonucleotides and blunt end ligation.[45] All genetic part sequences, plasmids, and strains are listed in Supplementary Table A.1-3, respectively. Enzymes were purchased from New England Biolabs and Thermo Fischer Scientific.

R. opacus was cultured in one of two defined minimal media, as previously described (see Supplementary Materials and Methods A), at 30°C and 250 rpm using 50 mL glass culture tubes.[15] Unless noted otherwise, 2 g/L glucose was used as the carbon source and 1 g/L ammonium sulfate was used as the nitrogen source. For all cultures, a single colony was transferred from a tryptic soy broth (TSB) agar plate into 2.5 mL of minimal media and incubated for ~24 hrs, and then subcultured into 10 mL of media to produce a larger quantity of seed culture. For the nitrogen sensor, the seed culture was centrifuged at 3000 relative centrifugal force (rcf) and re-suspended in nitrogen free media. All inductions and growth experiments were performed in triplicate at the 10 mL scale, with an initial OD_{600} of ~0.2. Inducers include arabinose (0.05 to 250 mM), acetamide (0.01 to 1000 nM), anhydrotetracycline (aTc; 0.05 to 100 ng/mL), and phenol (0.1, 0.3, and 0.75 g/L). The nitrogen sensor was induced with filter sterilized (0.22 μM filter) ammonium sulfate (0.05 to 0.8 g/L). Fluorescence measurements were performed at ~20 hrs after induction, unless noted otherwise. All chemicals were purchased from Sigma-Aldrich.

2.5.2 Fluorescence measurements

Cell fluorescence and absorbance were measured in black 96-well plates (Greiner Bio-One flat bottom, chimney well, μclear) using a Tecan Infinite 200 Pro plate reader. The excitation and emission wavelengths for fluorescent reporters are listed in Supplementary Table A.8.

Fluorescence measurements were normalized using the following equation: $Fluo_{norm} = \frac{Fluo_{sample}}{Abs600_{sample}} - \frac{Fluo_{control}}{Abs600_{control}}$, where $Fluo_{norm}$ is the normalized fluorescence, $Abs600_{sample}$ is the test strain absorbance at 600 nm, $Fluo_{sample}$ is the test strain fluorescence, $Fluo_{control}$ is the empty vector control strain fluorescence, and $Abs600_{control}$ is the empty vector control strain absorbance at 600 nm. For transfer functions, the Hill equation was fitted to the data using the Microsoft Excel solver by minimizing the root mean square error (see Supplementary Materials and Methods A).

2.6 Supporting information

Supplementary Tables A.1-8, Supplementary Figures A.1-8, and Supplementary Materials and Methods A can be found in Appendix A.

2.7 References

1. Pu, Y., Zhang, D., Singh, P. M., and Ragauskas, A. J. (2008) The new forestry biofuels sector, Biofuels, Bioproducts and Biorefining 2, 58-73.

2. Balan, V., Chiaramonti, D., and Kumar, S. (2013) Review of US and EU initiatives toward development, demonstration, and commercialization of lignocellulosic biofuels, Biofuels, Bioproducts and Biorefining 7, 732-759.

3. Beckham, G. T., Johnson, C. W., Karp, E. M., Salvachua, D., and Vardon, D. R. (2016) Opportunities and challenges in biological lignin valorization, Curr. Opin. Biotechnol. 42, 40-53.

4. Lin, S. Y., and Dence, C. W. (1992) Methods in Lignin Chemistry, Springer.

5. Pandey, M. P., and Kim, C. S. (2011) Lignin Depolymerization and Conversion: A Review of Thermochemical Methods, Chemical Engineering & Technology 34, 29-41.

6. Fengel, D., and Wegener, G. (1983) Wood: chemistry, ultrastructure, reactions, Walter de Gruyter.

7. Lynd LR, Van Zyl WH, McBride JE, Laser M. (2005) Consolidated bioprocessing of cellulosic biomass: an update., Curr. Opin. Biotechnol. 16, 577-583.

8. Vardon, D. R., Franden, M. A., Johnson, C. W., Karp, E. M., Guarnieri, M. T., Linger, J. G., Salm, M. J., Strathmann, T. J., and Beckham, G. T. (2015) Adipic acid production from lignin, Energy & Environmental Science 8, 617-628.

9. Kruyer, N. S., and Peralta-Yahya, P. (2017) Metabolic engineering strategies to bio-adipic acid production, Curr. Opin. Biotechnol. 45, 136-143.

10. Sun, X., Lin, Y., Yuan, Q., and Yan, Y. (2014) Biological production of muconic acid via a prokaryotic 2,3-dihydroxybenzoic acid decarboxylase, ChemSusChem 7, 2478-2481.

11. Kurosawa, K., Laser, J., and Sinskey, A. J. (2015) Tolerance and adaptive evolution of triacylglycerol-producing Rhodococcus opacus to lignocellulose-derived inhibitors, Biotechnology for Biofuels 8. 76

12. Holder, J. W., Ulrich, J. C., DeBono, A. C., Godfrey, P. A., Desjardins, C. A., Zucker, J., Zeng, Q., Leach, A. L. B., Ghiviriga, I., Dancel, C., Abeel, T., Gevers, D., Kodira, C. D., Desany, B., Affourtit, J. P., Birren, B. W., and Sinskey, A. J. (2011) Comparative and Functional Genomics of Rhodococcus opacus PD630 for Biofuels Development, PLoS Genet 7, e1002219.

13. Alvarez, H. M., Mayer, F., Fabritius, D., and Steinbuchel, A. (1996) Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630, Archives of microbiology 165, 377-386.

14. Yoneda, A., Henson, W. R., Goldner, N. K., Park, K. J., Forsberg, K. J., Kim, S. J., Pesesky, M. W., Foston, M., Dantas, G., and Moon, T. S. (2016) Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630, Nucleic Acids Res 44, 2240-2254.

15. Hollinshead, W. D., Henson, W. R., Abernathy, M., Moon, T. S., and Tang, Y. J. (2016) Rapid metabolic analysis of Rhodococcus opacus PD630 via parallel 13C-metabolite fingerprinting, Biotechnol. Bioeng. 113, 91-100.

16. Kurosawa, K., Wewetzer, S. J., and Sinskey, A. J. (2013) Engineering xylose metabolism in triacylglycerol-producing Rhodococcus opacus for lignocellulosic fuel production, Biotechnology for biofuels 6, 134.

17. Kurosawa, K., Boccazzi, P., de Almeida, N. M., and Sinskey, A. J. (2010) High-cell-density batch fermentation of Rhodococcus opacus PD630 using a high glucose concentration for triacylglycerol production, J. Biotechnol. 147, 212-218.

18. K. Le, R., Jr, T. W., Das, P., Meng, X., J. Stoklosa, R., Bhalla, A., B. Hodge, D., S. Yuan, J., and J. Ragauskas, A. (2017) Conversion of corn stover alkaline pre-treatment waste streams into biodiesel via Rhodococci, RSC Advances 7, 4108-4115.

19. Goswami, L., Tejas Namboodiri, M. M., Vinoth Kumar, R., Pakshirajan, K., and Pugazhenthi, G. (2017) Biodiesel production potential of oleaginous Rhodococcus opacus grown on biomass gasification wastewater, Renewable Energy 105, 400-406.

20. Ling, H., Teo, W., Chen, B., Leong, S. S., and Chang, M. W. (2014) Microbial tolerance engineering toward biochemical production: from lignocellulose to products, Curr. Opin. Biotechnol. 29, 99-106.

21. Hetzler, S., and Steinbüchel, A. (2013) Establishment of Cellobiose Utilization for Lipid Production in Rhodococcus opacus PD630, Appl. Environ. Microbiol. 79, 3122-3125.

22. Kurosawa, K., Plassmeier, J., Kalinowski, J., Ruckert, C., and Sinskey, A. J. (2015) Engineering L-arabinose metabolism in triacylglycerol-producing Rhodococcus opacus for lignocellulosic fuel production, Metab. Eng. 30, 89-95.

23. Hetzler, S., Broker, D., and Steinbuchel, A. (2013) Saccharification of Cellulose by Recombinant Rhodococcus opacus PD630 Strains, Applied and environmental microbiology 79, 5159-5166.

24. Lee, M. E., Aswani, A., Han, A. S., Tomlin, C. J., and Dueber, J. E. (2013) Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay, Nucleic Acids Res 41, 10668-10678.

25. Xu, P., Li, L., Zhang, F., Stephanopoulos, G., and Koffas, M. (2014) Improving fatty acids production by engineering dynamic pathway regulation and metabolic control, Proc Natl Acad Sci U S A 111, 11299-11304.

26. Hoynes-O'Connor, A., and Moon, T. S. (2015) Programmable genetic circuits for pathway engineering, Current opinion in biotechnology 36, 115-121.

27. Hanisch, J., Waltermann, M., Robenek, H., and Steinbuchel, A. (2006) The Ralstonia eutropha H16 phasin PhaP1 is targeted to intracellular triacylglycerol inclusions in Rhodococcus opacus PD630 and Mycobacterium smegmatis mc2155, and provides an anchor to target other proteins, Microbiology 152, 3271-3280.

28. Dong, L., Nakashima, N., Tamura, N., and Tamura, T. (2004) Isolation and characterization of the Rhodococcus opacus thiostrepton-inducible genes tipAL and tipAS: application for recombinant protein expression in Rhodococcus, FEMS microbiology letters 237, 35-40.

29. Hahn, S., and Schleif, R. (1983) In vivo regulation of the Escherichia coli araC promoter, J Bacteriol 155, 593-600.

30. Immethun, C. M., DeLorenzo, D. M., Focht, C. M., Gupta, D., Johnson, C. B., and Moon, T. S. (2017) Physical, chemical, and metabolic state sensors expand the synthetic biology toolbox for Synechocystis sp. PCC 6803, Biotechnol. Bioeng. DOI: 10.1002/bit.26275

31. Roberts, G., Muttucumaru, D. G., and Parish, T. (2003) Control of the acetamidase gene of Mycobacterium smegmatis by multiple regulators, FEMS Microbiol. Lett. 221, 131-136.

32. Gossen, M., and Bujard, H. (1993) Anhydrotetracycline, a novel effector for tetracycline controlled gene expression systems in eukaryotic cells, Nucleic Acids Res 21, 4411-4412.

33. Ehrt, S., Guo, X. V., Hickey, C. M., Ryou, M., Monteleone, M., Riley, L. W., and Schnappinger, D. (2005) Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor, Nucleic acids research 33, e21.

34. Rock, J. M., Hopkins, F. F., Chavez, A., Diallo, M., Chase, M. R., Gerrick, E. R., Pritchard, J. R., Church, G. M., Rubin, E. J., Sassetti, C. M., Schnappinger, D., and Fortune, S. M. (2017) Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform, Nature microbiology 2, 16274.

35. Lennen, R. M., and Pfleger, B. F. (2013) Microbial production of fatty acid-derived fuels and chemicals, Curr. Opin. Biotechnol. 24, 1044-1053.

36. Durante-Rodríguez, G., Gómez-Álvarez, H., Nogales, J., Carmona, M., and Díaz, E. (2016) One-Component Systems that Regulate the Expression of Degradation Pathways for Aromatic Compounds, In Cellular Ecophysiology of Microbe (Krell, T., Ed.), pp 1-39, Springer International Publishing.

37. Szokol, J., Rucka, L., Simcikova, M., Halada, P., Nesvera, J., and Patek, M. (2014) Induction and carbon catabolite repression of phenol degradation genes in Rhodococcus erythropolis and Rhodococcus jostii, Appl. Microbiol. Biotechnol. 98, 8267-8279.

38. Kosa, M., and Ragauskas, A. J. (2013) Lignin to lipid bioconversion by oleaginous Rhodococci, Green Chem 15, 2070-2074.

39. Patrauchan, M. A., Florizone, C., Dosanjh, M., Mohn, W. W., Davies, J., and Eltis, L. D. (2005) Catabolism of benzoate and phthalate in Rhodococcus sp. strain RHA1: redundancies and convergence, Journal of bacteriology 187, 4050-4063.

40. Johnson, C. W., Salvachúa, D., Khanna, P., Smith, H., Peterson, D. J., and Beckham, G. T. (2016) Enhancing muconic acid production from glucose and lignin-derived aromatic compounds via increased protocatechuate decarboxylase activity, Metabolic Engineering Communications 3, 111-119.

41. Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., 3rd, Burland, V., Mau, B., Petrosino, J. F., Qin, X., Muzny, D. M., Ayele, M., Gibbs, R. A., Csorgo, B., Posfai, G., Weinstock, G. M., and Blattner, F. R. (2008) The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse, Journal of bacteriology 190, 2597-2606.

42. Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability, PloS one 3, e3647.

43. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases, Nature methods 6, 343-345.

44. Kalscheuer, R., Arenskotter, M., and Steinbuchel, A. (1999) Establishment of a gene transfer system for Rhodococcus opacus PD630 based on electroporation and its application for recombinant biosynthesis of poly(3-hydroxyalkanoic acids), Applied microbiology and biotechnology 52, 508-515.

45. Georgescu, R., Bandara, G., and Sun, L. (2003) Saturation Mutagenesis, In Directed Evolution Library Creation (Arnold, F., and Georgiou, G., Eds.), pp 75-83, Humana Press.

Chapter 3: Construction of genetic logic gates based on the T7 RNA polymerase expression system in *Rhodococcus opacus* PD630

Reprinted with permission from DeLorenzo, D.M. and Moon, T.S. Construction of genetic logic gates based on the T7 RNA polymerase expression system in *Rhodococcus opacus* PD630. *ACS Synthetic Biology* (2019). 8(8), 1921-1930.

The previous chapter detailed the development of a number of tools for tunable gene expression in *Rhodococcus opacus* PD630. The following chapter will discuss the implementation of several of these parts, along with newly developed tools for high levels of gene expression (e.g., T7 RNAP expression platform, IPTG-inducible $pLac_{RO}$ promoters, selective aromatic sensors), into programmable genetic logic gates. Genetic logic allows for more controllable and sophisticated genetic constructs that can lead to improvements in cellular functions and enhance productivity. I performed all experiments and wrote the manuscript.

3.1 Abstract

Rhodococcus opacus PD630 (*R. opacus*) is a non-model, gram-positive bacterium which holds promise as a biological catalyst for the conversion of lignocellulosic biomass into value-added products. In particular, it demonstrates both a high tolerance for and an ability to consume inhibitory lignin-derived aromatics, generates large quantities of lipids, exhibits a relatively rapid growth rate, and has a growing genetic toolbox for engineering. However, the availability of genetic parts for tunable, high-activity gene expression is still limited in *R. opacus*. Furthermore, genetic logic circuits for sophisticated gene regulation have never been demonstrated in *Rhodococcus* spp. To address these shortcomings, two inducible T7 RNA polymerase-based

expression systems were implemented for the first time in *R. opacus* and applied to constructing AND and NAND genetic logic gates. Additionally, three IPTG-inducible promoters were created by inserting LacI binding sites into newly-characterized constitutive promoters. Furthermore, four novel aromatic sensors for 4-hydroxybenzoic acid, vanillic acid, sodium benzoate, and guaiacol were developed, expanding the gene expression toolbox. Finally, the T7 RNA polymerase platform was combined with a synthetic IPTG-inducible promoter to create an IMPLY logic gate. Overall, this work represents the first demonstration of a heterologous RNA polymerase system and synthetic genetic logic in *R. opacus*, enabling complex and tunable gene regulation in this promising non-model host for bioproduction.

3.2 Introduction

Living cells naturally use genetic circuits to sense, process, and respond to their environment. Bacteria possess thousands of genes, and it would be deleterious if all genes were expressed at a constitutive level. In fact, cells implement complex regulatory networks that determine when and to what extent each of those genes should be expressed. Initial engineering efforts, however, typically begin with high-level, constitutive expression of heterologous genes, which can be metabolically taxing to the organism.[1] Dynamic pathway regulation, which adds a layer of control over gene expression by enabling autonomous feedback in the system, has demonstrably improved productivity and cellular fitness.[2-4] Furthermore, Boolean logic (e.g., AND and NAND) has been implemented in microbial hosts, where the output of a genetic circuit is reduced to a binary "true" or "false" based on an array of different inputs, allowing the organism to alter gene expression based on a specific set of conditions.[5-7] Researchers have also expanded these engineered regulatory motifs to construct large synthetic genetic circuits that respond to a variety of stimuli and replicate native regulatory networks.[8]

In order to implement these kinds of gene control architecture, genetic parts—particularly regulatory elements such as promoters and their cognate transcription factors—must first be developed and characterized. Unfortunately, outside of well-studied model organisms (e.g., *Escherichia coli*), genetic toolboxes are often underdeveloped and thus insufficient for constructing genetic circuits. Implementing genetic circuits in non-model organisms is further complicated by the fact that genetic part performance may differ between organisms and thus require additional optimization.[9-10] More recently, however, non-model organisms have received heightened research attention due to their unusual evolutionary adaptations, diverse metabolic pathways, and a variety of novel enzymes and metabolites.[11-12]

Rhodococcus opacus PD630 (hereafter *R. opacus*) is a non-model, gram-positive bacterium that holds great promise as a microbial chassis for the conversion of waste lignocellulosic biomass into value-added products (e.g., lipids).[13-15] Due to its extensive catabolic pathways, notably for lignin-derived aromatics, and its tolerance towards inhibitory lignocellulose-derived compounds (e.g., furans, phenolics, etc.), *R. opacus* is ideally suited for the valorization of biomass breakdown products.[15-19] While its genetic toolbox has steadily grown to comprise an array of parts and techniques, including plasmid backbones, selection markers, a methodology for genome engineering, CRISPR interference for targeted gene repression, reference genes to quantify gene expression, metabolite sensors to detect relevant aromatic feedstocks, and a library of constitutive promoters, these parts have yet to be combined into genetic circuits to perform logic.[10, 20-21] Moreover, the ability to highly express and reliably tune gene expression in *R. opacus* is still limited to a handful of promoters.[10]

The most readily employed type of promoter for engineering in *R. opacus* is a constitutive promoter, which may reduce productivity due to metabolic burden.[20, 22-24] An alternative to this

approach is the conversion of a constitutive promoter to an inducible promoter. Native regulatory systems can provide inspiration for ways to enact such an alteration. One of the inducible promoters, pLac, was derived from the lactose degradation operon in *E. coli*, which is regulated by the LacI repressor.[25] In the absence of lactose or a synthetic substitute (e.g., isopropyl β-D-1-thiogalactopyranoside [IPTG]), LacI binds to a short nucleotide sequence referred to as the *lac* operator site (lacO), preventing RNA polymerase from binding to the promoter.[26] When its ligand is present, LacI dissociates from the operator, allowing for transcription. Synthetic IPTG-dependent promoters have previously been constructed for other non-model organisms through the insertion of a lacO site into or downstream of a constitutive promoter.[27-28] The development of tunable promoters is critical for both the optimization of gene expression in *R. opacus* and the construction of genetic logic gates.

An additional system that can simultaneously advance both promoter development and genetic circuit construction in *R. opacus* is the T7 RNA polymerase (T7 RNAP) expression platform. T7 RNAP was originally isolated from the T7 bacteriophage and was found to recognize only its cognate promoter (pT7); furthermore, pT7 is not recognized by native microbial RNA polymerases, thus allowing for an orthogonal expression system.[29-30] Due to T7 RNAP's high activity, processivity, and selectivity, substantial levels of heterologous gene expression can be achieved both *in vivo* and *in vitro* compared to traditional expression systems.[30-31] The integration of an inducible T7 RNAP cassette into the BL21 strain of *E. coli* [BL21(DE3)] facilitated extremely high levels of target protein production (up to 50% of the cellular proteome) and paved the way for BL21(DE3) to become a pivotal industrial strain for the production of a range of products (e.g., therapeutic proteins).[30, 32-33] Furthermore, pT7 can be mutated to change its rate of transcription initiation while still retaining specificity to the T7 RNAP, and the T7 RNAP and pT7

system can be co-evolved to form novel orthogonal pairs that can drive the expression of multiple gene cassettes independently.[34-35] Additionally, several forms of genetic logic gates have been constructed in *E. coli* using the T7 RNAP, including AND and IMPLY logic gates.[36-37]

The T7 RNAP expression platform has several advantages over a traditional inducible promoter (e.g., pBAD, pTet, and pLac) for both the enactment of genetic circuits and the tunable expression of genes. Firstly, the reliance on a native RNAP can hinder the implementation of predictable, independently-functioning circuits due to multiple factors, including its fluctuating total level and the changing demand for endogenous RNAP across the genome.[37] Secondly, a host's native RNAP is always expressed at some level, which can lead to leaky basal expression from a standard sigma70 promoter.[38] In contrast to the native RNAP, the T7 RNAP is typically under the control of an inducible promoter, allowing finer tuning over its expression level. This system can be further regulated by a secondary mechanism, such as the placement of a lacO site downstream of pT7, which can prevent transcription by any basally expressed T7 RNAP.[39]

In this work, two different inducible T7 RNAP systems are demonstrated in *R. opacus*, representing the first time that a heterologous RNA polymerase has been implemented for gene expression in any *Rhodococcus* spp. Additionally, these T7 RNAP-based expression platforms were used as the basis for AND and NAND genetic logic gates in *R. opacus*. To further expand the number of genetic parts, in particular promoter/regulator pairs that can be used in additional genetic circuits, three synthetic IPTG-dependent promoters were developed from newly characterized constitutive promoters in *R. opacus*. These represent the first IPTG-dependent promoters demonstrated in *Rhodococcus* spp. Furthermore, four novel aromatic sensors for 4-hydroxybenzoic acid, vanillic acid, sodium benzoate, and guaiacol were developed, providing potentially useful gene expression systems for future lignin valorization. Finally, the utility of both

the T7 RNAP system and one of the new IPTG-dependent promoters was demonstrated by linking them together to create an IMPLY genetic circuit. Overall, this work expands the ability to highly express heterologous genes in a tunable manner in *R. opacus* and demonstrates the capacity for sophisticated gene regulation in this promising non-model host for bioproduction.

3.3 Results and discussion

3.3.1 T7 RNAP expression platform.

Similar to the *E. coli* protein production strain BL21(DE3), the T7 RNAP gene was integrated into a previously described *R. opacus* chromosomal neutral site (ROCI3) under the control of an inducible promoter.[20] Two different *R. opacus* strains were created using either a previously reported phenol-inducible promoter (pLPD06575; hereafter referred to as pPhenol) or an *R. opacus*-optimized anhydrotetracycline (aTc)-inducible pTet promoter to control T7 RNAP expression.[10, 20] A hygromycin B antibiotic resistance cassette was co-integrated with the T7 RNAP expression cassette for selection.[20] The consensus T7 promoter (pT7) was placed upstream of an enhanced GFP (*eGFP*) gene on the pAL5000(S) shuttle plasmid.

To examine the inducibility of the two integrated T7 RNAP systems, the normalized fluorescent output (see Methods for details) was quantified in response to a range of concentrations of either phenol or aTc (Figure 3.1). For pPhenol-*T7 RNAP*, a fold change of 55 was observed in the normalized fluorescent output between 0 and 0.5 g/L phenol (Figure 3.1B). For pTet-*T7 RNAP*, a fold change of 5.3 was observed in the normalized fluorescent output between 0 and 1 ng/mL aTc (Figure 3.1C). The pPhenol-*T7 RNAP* strain generated 2.8-fold lower minimum (i.e. tighter OFF) and 3.7-fold higher maximum (i.e. higher maximum ON) fluorescent outputs than the pTet-*T7 RNAP* strain.

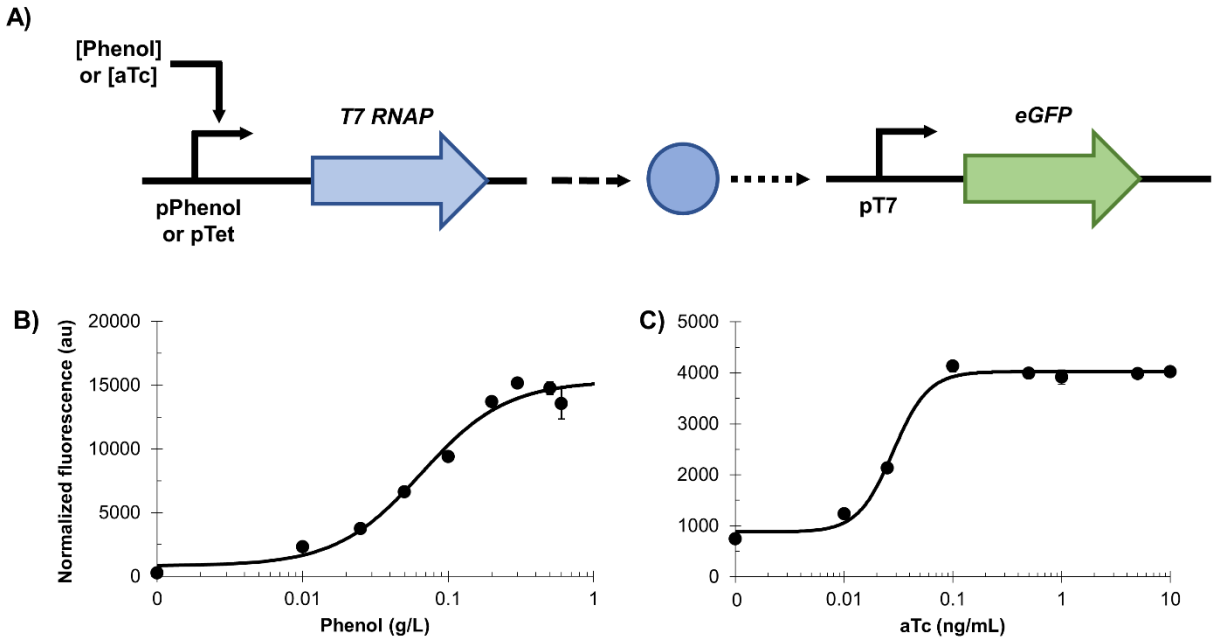


Figure 3.1. Inducible T7 RNAP system in *R. opacus*. **A)** Schematic of the phenol- or aTc-inducible T7 RNA polymerase (T7 RNAP) expression platform. A phenol-inducible promoter (pLPD06575; referred to as pPhenol) or an optimized pTet promoter was placed upstream of the *T7 RNAP* gene and integrated into the *R. opacus* genome at a previously determined neutral site (ROCI3).[20] The T7 promoter (pT7) was placed upstream of *eGFP* on the pAL5000(S) plasmid backbone. **B)** Normalized fluorescence of a strain containing pPhenol-*T7 RNAP* in response to 0, 0.01, 0.025, 0.05, 0.1, 0.2, 0.3, 0.5, or 0.6 g/L phenol. The increase in fluorescent output from 0 to 0.5 g/L phenol was 55-fold. **C)** Normalized fluorescence of a strain containing pTet-*T7 RNAP* in response to 0, 0.01, 0.025, 0.1, 0.5, 1, 5, or 10 ng/mL aTc. The increase in fluorescent output from 0 to 1 ng/mL aTc was 5.3-fold. Values are averages of three replicates, and error bars represent one standard deviation. The solid black line in each case represents a fitted curve (see Supplementary Methods B and Supplementary Table B.4).

3.3.2 AND genetic logic gate

The pPhenol- and pTet-*T7 RNAP* systems were next demonstrated to act as AND logic gates. This was accomplished by inserting a *lac* operator site (lacO) downstream of pT7 and adding a constitutive *lacI* expression cassette to the reporter plasmid. An AND logic gate is a circuit that is "true" only when all inputs are present, which helps to reduce leaky output expression as multiple conditions must first be met before the circuit turns on. As two inducers (IPTG and either

phenol or aTc) are required for complete induction of *eGFP* (Figure 3.2A), these circuits are classified as two-input AND gates (Figure 3.2B).[40] Both the pPhenol-*T7 RNAP* and pTet-*T7 RNAP* systems were highly repressed in response to the uninduced [0 0] input, showing no statistically significant difference between their normalized fluorescence and the empty vector control strain's normalized fluorescence (Figures 3.2C and 3.2D). When only one inducer was provided, either phenol/aTc [1 0] or IPTG [0 1], some fluorescence was observed; however, it was at least an order of magnitude lower than that of the two-input [1 1] condition. The increase in normalized fluorescence between either the [1 0] or [0 1] condition and the [1 1] condition for the pPhenol-*T7 RNAP* construct was 18- and 27-fold, respectively. The increase in normalized fluorescence between either the [1 0] or [0 1] condition and the [1 1] condition for the pTet-*T7 RNAP* construct was 39- and 16-fold, respectively.

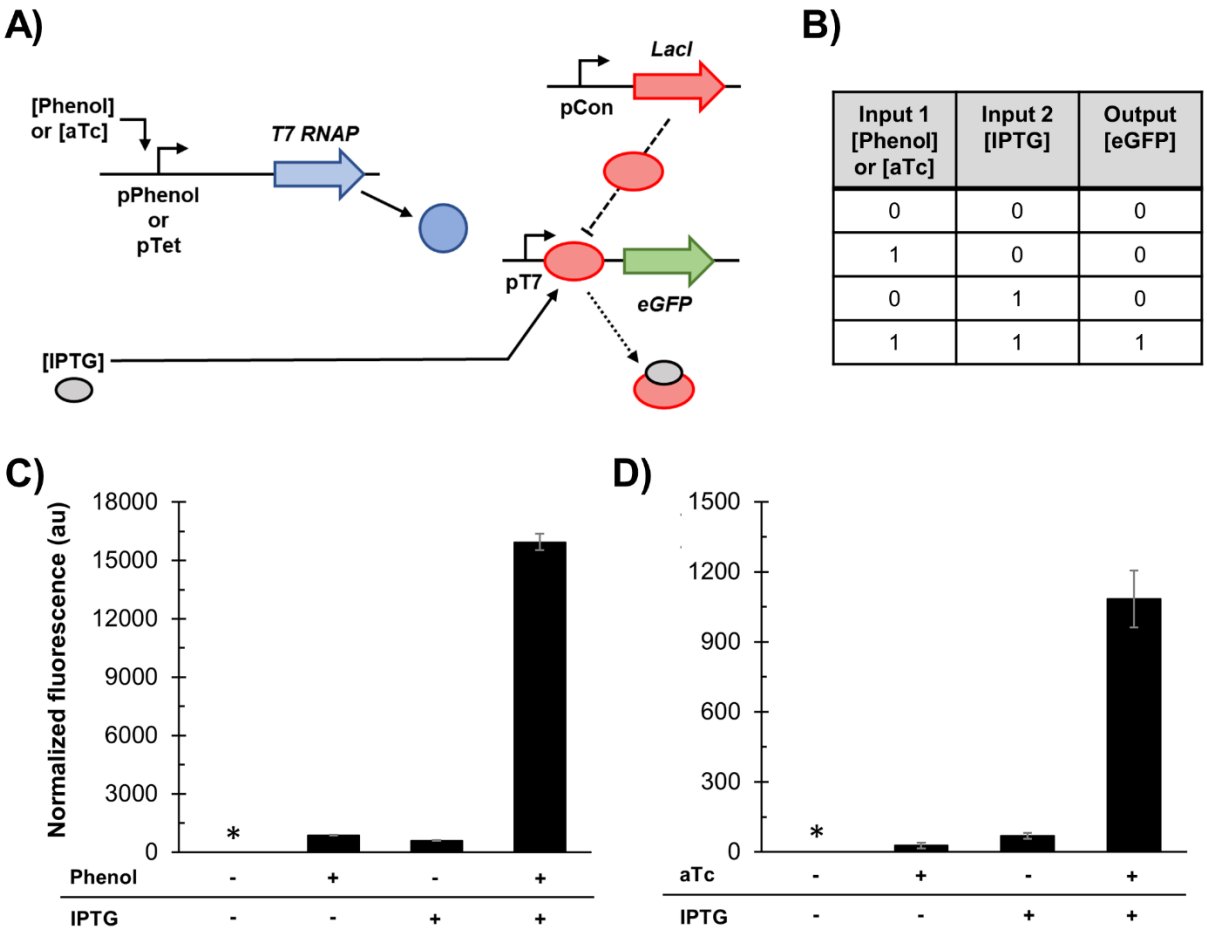
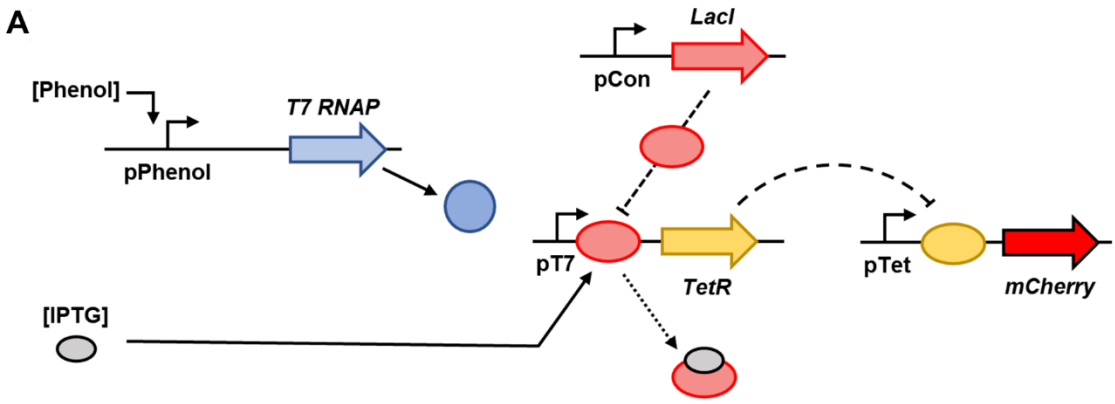


Figure 3.2. AND logic gate in *R. opacus*. **A)** Schematic of the genetic AND gate. **B)** AND truth table. **C)** Normalized fluorescence of the pPhenol-*T7 RNAP* circuit in response to different combinations of inputs. **D)** Normalized fluorescence of the pTet-*T7 RNAP* circuit in response to different combinations of inputs. IPTG, phenol, and aTc were provided as appropriate at concentrations of 1 mM, 0.5 g/L, and 1 ng/mL, respectively. Due to the [0 0] condition having no discernable fluorescence, a fold change between [0 0] and [1 1] cannot be calculated. The increase in normalized fluorescence between either the [1 0] or [0 1] conditions and the [1 1] condition for the pPhenol-*T7 RNAP* construct was 18- and 27-fold, respectively. The increase in normalized fluorescence between either the [1 0] or [0 1] conditions and the [1 1] condition for the pTet-*T7 RNAP* construct was 39- and 16-fold, respectively. Values are averages of three replicates, and error bars represent one standard deviation. An asterisk (*) represents that there was no statistical difference in normalized fluorescence between the experimental strain and the control strain containing an empty control plasmid.

3.3.3 NAND genetic logic gate.

The pPhenol-*T7 RNAP* expression system was further applied to build a NAND gate (Figure 3.3A), a more complex circuit than the previously demonstrated AND gates. A NAND (NOT-AND) logic gate is only "false" when all inputs are present (Figure 3.3B). NAND gates are particularly useful as they are functionally complete, meaning that any form of Boolean logic can be implemented by layering multiple NAND gates together. A NAND logic gate was constructed in the pPhenol-*T7 RNAP* strain by placing pT7 and a lacO site upstream of *tetR* and placing the TetR-repressible pTet promoter upstream of *mCherry* on the pAL5000(S) shuttle vector containing constitutively expressed *lacI*. When both phenol and IPTG are provided, the T7 RNAP is expressed, the lacO site is not occupied by LacI, and thus TetR is produced, leading to the repression of pTet-*mCherry*. The decrease in normalized fluorescence between the uninduced [0 0] and the two-input [1 1] conditions was 27-fold (Figure 3.3C). When only one inducer was provided, either phenol [1 0] or IPTG [0 1], there was a decrease in the normalized fluorescence relative to the [0 0] condition; however, fluorescence was still at least an order of magnitude higher than that of the two-input [1 1] condition, with observed decreases of 15- and 20-fold, respectively.



B)

Input 1 [Phenol]	Input 2 [IPTG]	Output [mCherry]
0	0	1
1	0	1
0	1	1
1	1	0

C)

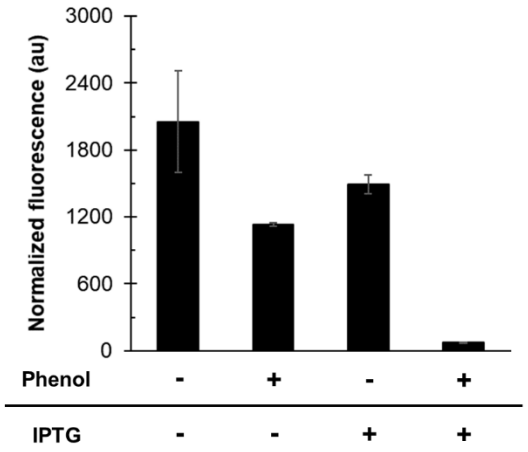


Figure 3.3. NAND logic gate in *R. opacus*. **A)** Schematic of the genetic NAND gate. **B)** NAND truth table. **C)** Normalized fluorescence of the NAND genetic circuit in response to different combinations of inputs. Phenol and IPTG were provided as appropriate at concentrations of 0.2 g/L and 1 mM, respectively. aTc was not provided during the main experimental culture (see Methods for details). An initial OD_{600} of 0.1 was used, and cultures were grown in minimal media B with 10 g/L glucose as a carbon source. The decrease in normalized fluorescence between uninduced [0 0] (the highest ON) and phenol- and IPTG-induced [1 1] (OFF) was 27-fold. The decrease in normalized fluorescence between when either phenol [1 0] or IPTG [0 1] was individually provided and when both were provided [1 1] was 15- and 20-fold, respectively. The response of the NAND logic gate when 0.5 g/L phenol was used is shown in Supplementary Figure B.3. Values are averages of three replicates, and error bars represent one standard deviation.

3.3.4 Synthetic IPTG inducible promoters.

Two-input AND and NAND genetic circuits are just two demonstrations of the many kinds of Boolean logic gates that can be constructed in a microbial host. In an extreme example, Nielsen et al. constructed numerous multi-tiered genetic circuits consisting of up to ten orthogonal regulators to create synthetic signaling networks in *E. coli*.[8] In order to expand the types of genetic circuits that can be constructed and to add additional layers of complexity in *R. opacus*, more circuit components (i.e., regulators and promoters) must be developed. To date, two heterologous, one-component regulators have been characterized in *R. opacus* (TetR [repressor] and AraC [activator]), in addition to several *Rhodococcus* regulators that have been identified and characterized to varying degrees (e.g., TipA [activator], NlpR [repressor], and NpdR [repressor]).[10, 41-43]

A common heterologous regulator/promoter system that has not yet been demonstrated in *Rhodococcus* spp. is the IPTG-inducible pLac/LacI system. Constitutive variants of pLac and its derivatives (e.g., pTac) have been employed in *Rhodococcus* spp., but none have been utilized as IPTG-dependent promoters.[23-24, 44] The previously discussed AND and NAND gates utilize a T7 promoter that includes a downstream lacO site, as well as a *lacI* expression cassette. The absence or presence of IPTG was sufficient to repress or de-repress, respectively, the expression of the fluorescent reporter protein under control of pT7(lacO). The next step was to create novel synthetic IPTG-inducible promoters that could be transcribed by the native *R. opacus* RNAP. To ensure that these new promoters would have suitably high levels of gene expression for engineering purposes, it was necessary to identify constitutive promoters with strong activity in *R. opacus*.

R. opacus native promoters, as well as highly active promoters from related Actinomycetales species, were compared to a baseline promoter (pConstitutive) that was recently reported to generate high levels of GFP+ fluorescence in *R. opacus* (Supplementary Figure B.1).[20] To choose *R. opacus* native promoters, two criteria were implemented. First, the promoter should have high activity based on previously published transcriptomic studies in *R. opacus*.[16-17] Second, the promoter should be for ribosomal RNAs (rRNAs) or transfer RNAs (tRNAs). The primary reason for the second criterion was that these RNAs, particularly rRNAs, have a substantial demand within the cell and are thus transcribed at high levels.[45] The secondary reason was that the transcription start sites for these promoters are readily identified as the rRNA and tRNA sequences are evolutionarily conserved and clearly annotated. The identified endogenous promoters were those for rRNA (pRS23365), tryptophan tRNA (pRS28745), threonine tRNA (pRS28770), isoleucine tRNA (pRS00085), glutamate tRNA (pRS06620), and serine tRNA (pRS02495). These six promoters, in addition to pConstitutive, were placed upstream of RBS-1 and *GFP+* (Supplementary Figure B.1A). pRS00085 was the only promoter demonstrating a statistically higher normalized fluorescence ($p < 0.005$ based on two-sample t-test) than the normalized fluorescence generated by pConstitutive, with an increase of 1.4-fold.

Promoters demonstrated to be highly active in other related gram-positive microbes, in particular *Rhodococcus ruber* TH and *Corynebacterium glutamicum*, were also examined. Three strong promoters derived from the *R. ruber* TH genome (pamiM, pnhM, and pcs) were selected and placed upstream of RBS-2 and *GFP+* (Supplementary Figure B.1B; Supplementary Table B.1).[46] Plasmids containing RBS-1 occasionally demonstrated instability in *R. opacus*, with cells exhibiting loss of fluorescence. This led to the use of the slightly weaker RBS-2 (Supplementary Figure B.2) as no such instability was observed. The pamiM promoter-containing construct failed

to generate any *R. opacus* transformant. Additionally, two strong synthetic promoters designed for *C. glutamicum* (pCory1 and pCory2), as well as pConstitutive, were placed upstream of RBS-2 and *GFP+* (Supplementary Figure B.1B; Supplementary Table B.1).[47] Two promoters demonstrated normalized fluorescence that was statistically higher than pConstitutive based on a two-sample t-test, namely pnhM ($p < 0.0001$) and pCory1 ($p < 0.005$). Of these two, however, only pnhM demonstrated a meaningful improvement over pConstitutive, with a 2.4-fold increase in normalized fluorescence. To enable the use of the pamiM -10 and -35 sites, which were predicted to be strong recruiters of RNA polymerase in *R. ruber*, these nucleotide sequences were placed into a new genetic context. Specifically, by replacing the pcs promoter's -10 and -35 sites with those of the pamiM, a hybrid synthetic promoter was built. The hybrid promoter (referred to as pcs/pamiM) had a normalized fluorescence statistically higher than pConstitutive ($p < 0.005$) and also had an 8-fold higher fluorescence level than pcs.

Having identified a set of functional constitutive promoters, the next step to creating an IPTG-dependent promoter was to insert lacO into the expression cassette of those promoters with the strongest expression profiles. To minimize disturbances to surrounding part sequences, the initial construct ($pLac_{RO1}$) utilized an insulated 33-nucleotide (nt) lacO sequence which was inserted between pnhM and RBS-2 (Figure 3.4A; Supplementary Table B.1; Supplementary Figure B.4A). The normalized fluorescence of $pLac_{RO1}$, which was measured in response to a range of IPTG concentrations, exhibited a 26-fold increase from 0 to 1 mM IPTG. However, the maximum normalized fluorescence output of $pLac_{RO1}$ only reached ~12% of the output of pnhM. To improve the maximum output, a minimal 17-nt version of lacO ($lacO_{min}$) was overlaid with RBS-2 to create $pLac_{RO2}$ via several nucleotide changes (Supplementary Table B.1; Supplementary Figure B.4B). $pLac_{RO2}$ exhibited a 15-fold increase in normalized fluorescence from 0 to 1 mM IPTG (Figure

3.4B). Furthermore, the maximum normalized fluorescence of $pLac_{RO2}$ was nearly double that of $pLac_{RO1}$, reaching ~21% of the output of pnhM.

For both $pLac_{RO1}$ and $pLac_{RO2}$, the lacO site was placed downstream of the transcription start site (TSS), which means that this untranslated region (UTR) is appended to the transcript. In some scenarios, such as the expression of small non-coding RNAs (e.g., small guide RNA), a self-contained IPTG-dependent promoter would be desirable. To enable such leaderless transcription, a third IPTG-dependent promoter where $lacO_{min}$ was placed between the -35 and -10 sites was developed. The pcs/pamiM promoter was selected because it has nine identical nucleotides when aligned with the 17-nt $lacO_{min}$, requiring only 8-nt changes to create $pLac_{RO3}$ (Supplementary Table B.1; Supplementary Figure B.4C). The normalized fluorescence of $pLac_{RO3}$ was measured in response to a range of IPTG concentrations and exhibited a 13-fold increase from 0 to 1 mM IPTG (Figure 3.4C). Furthermore, the maximum normalized fluorescence of $pLac_{RO3}$ reached ~20% of the maximum output of pcs/pamiM.

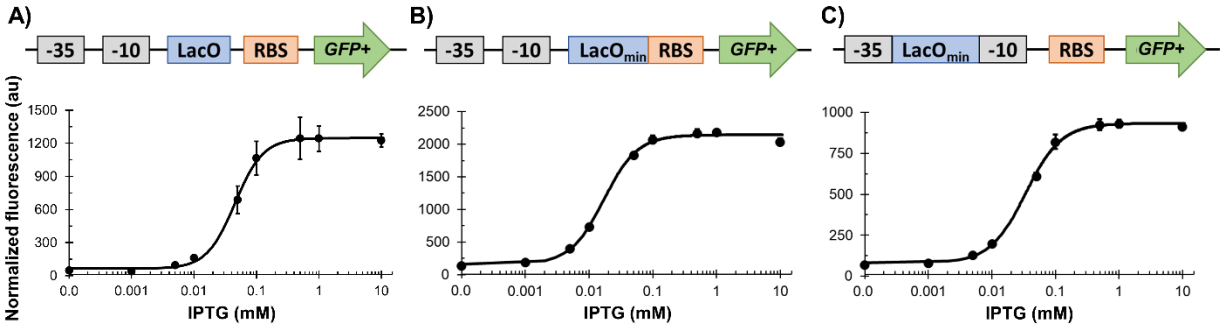
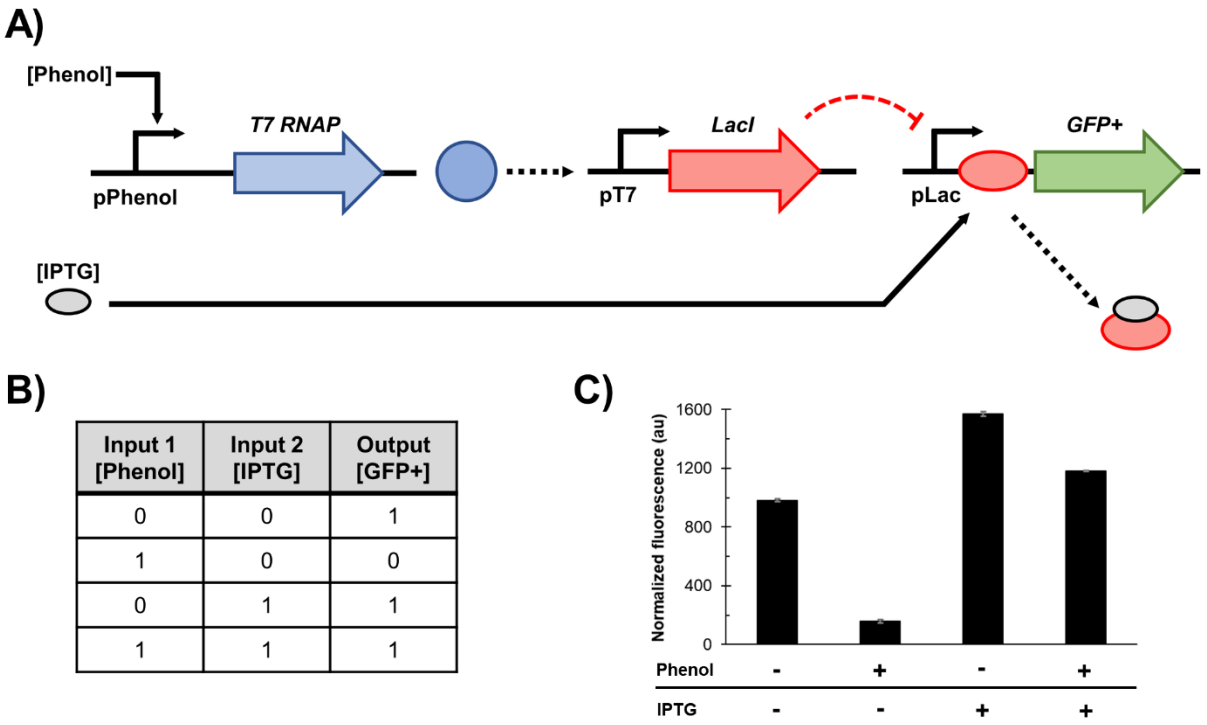


Figure 3.4. Synthetic IPTG-inducible promoters in *R. opacus*. The lac operator (lacO) was inserted into different regions of two constitutive promoters (pnhM and pcs/pamiM) to create three IPTG-inducible promoters. The plots report the normalized fluorescence in response to 0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, and 10 mM IPTG for each respective promoter. **A)** A 33 nucleotide (nt) lacO site was inserted downstream of the pnhM promoter and upstream of RBS-2 and *GFP+* to create $pLac_{RO1}$. An increase in normalized fluorescence of 26-fold was observed between 0 and 10 mM IPTG. **B)** A minimal 17 nt version of lacO ($LacO_{min}$) was overlaid with the RBS-2

(downstream of pnhM) to create $pLac_{RO2}$. An increase in normalized fluorescence of 15-fold was observed between 0 and 10 mM IPTG. **C)** $LacO_{min}$ was inserted between the -10 and -35 sites of the pcs/pamiM hybrid constitutive promoter to create $pLac_{RO3}$. An increase in normalized fluorescence of 13-fold was observed between 0 and 10 mM IPTG. Values are averages of three replicates, and error bars represent one standard deviation. The solid black line in each case represents a fitted curve (see Supplementary Methods B and Supplementary Table B.4). Promoter sequences are provided in Supplementary Table B.1 and Supplementary Figure B.4.

3.3.5 IMPLY genetic logic gate.

Combining the T7 RNAP expression system with one of the newly developed IPTG-dependent promoters allowed for the construction of an IMPLY genetic logic gate (Figure 3.5A). An IMPLY gate generates a "false" output only when the [1 0] input is given (Figure 3.5B). An IMPLY gate was constructed in the pPhenol-*T7 RNAP* strain by placing pT7 upstream of *lacI* and placing $pLac_{RO1}$ upstream of *GFP+* (Figure 3.5A). The [1 0] input (phenol only) showed 6-fold, 10-fold, and 7-fold decreases in normalized fluorescence compared to the [0 0] (no inducer), [0 1] (IPTG only), and [1 1] (phenol and IPTG) conditions, respectively (Figure 3.5C). Together with the successful construction of AND and NAND gates in *R. opacus*, this result demonstrates that large layered genetic circuits can be built in this non-model organism by assembling engineered genetic parts, all of which were encoded in 12 (e.g., IMPLY) or 13 kilobases (e.g., NAND, the largest logic gate built so far in any *Rhodococcus* spp.).



Input 1 [Phenol]	Input 2 [IPTG]	Output [GFP+]
0	0	1
1	0	0
0	1	1
1	1	1

Figure 3.5. IMPLY logic gate in *R. opacus*. **A)** Schematic of the genetic IMPLY gate. **B)** IMPLY truth table. **C)** Normalized fluorescence in response to the appropriate combinations of 0.4 g/L phenol and 1 mM IPTG. The decrease in normalized fluorescence between the uninduced [0 0] (the lowest ON) and phenol-induced [1 0] (OFF) conditions was 6-fold. The normalized fluorescent response of this circuit was additionally screened with a range of phenol concentrations (Supplementary Figure B.5). Values are averages of three replicates, and error bars represent one standard deviation.

3.3.6 Aromatic sensors

A phenol responsive sensor was utilized in the previously described experiments to control T7 RNAP expression, but lignin breakdown products contain a diversity of aromatic compounds and additional orthogonal aromatic sensors may be beneficial for future lignin valorization. Our previously published transcriptomic data identified catabolic funneling pathways that were selectively upregulated in response to one of four aromatic compounds: 4-hydroxybenzoic acid (HBA), vanillic acid (VAN), sodium benzoate (BEN), and guaiacol (GUA).[17] The region containing the promoter and RBS of each of these identified degradation pathways was placed

upstream of *GFP+* to develop four novel aromatic sensors, referred to as pLPD06764 (pHBA), pLPD00563 (pVAN), pLPD06580 (pBEN), and pLPD06578 (pGUA). The fluorescent output of each of these constructs was measured in response to a range of relevant aromatic compound concentrations (Figures 3.6A-D). Concentrations were chosen such that they would not cause substantial reductions in growth, based on previous toxicity data collected when these compounds were used as sole carbon sources, and thus sensor output saturation was not reached for all sensors.[17] pHBA demonstrated a fold change of 6 in normalized fluorescence from 0 to 1 g/L HBA (Figure 3.6A). The fluorescence value of pVAN at 1 g/L glucose was indistinguishable from the background level, and a fold change could not be calculated (Figure 3.6B). pBEN demonstrated a fold change of 35 in normalized fluorescence from 0 to 5 g/L BEN (Figure 3.6C). pGUA demonstrated a fold change of 137 in normalized fluorescence from 0 to 1.5 g/L GUA (Figure 3.6D). To investigate the selectivity of each of these sensors, each strain was grown on glucose or glucose plus protocatechuic acid, phenol, 4-hydroxybenzoic acid, vanillic acid, sodium benzoate, or guaiacol (Figures 3.6E-H). Each of these constructs was confirmed to be selectively upregulated in the presence of its cognate compound, with different degrees of response to noncognate compounds. In addition to the IPTG-inducible promoters developed in this work, these novel aromatic sensors expand the gene expression toolbox by providing inducible sensors with different output ranges (Figures 3.6A-D and Supplementary Figure B.6).

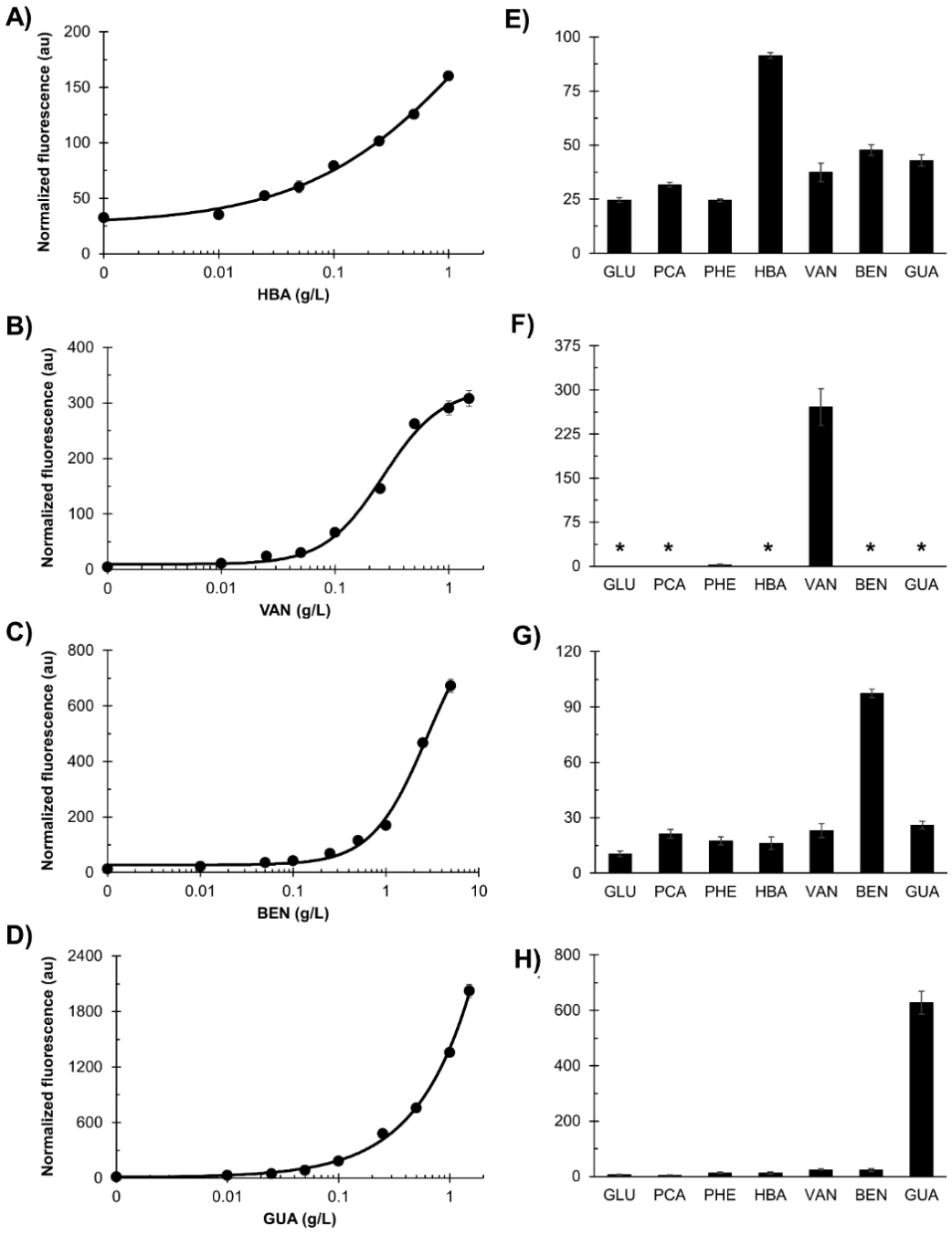


Figure 3.6. Novel aromatic sensors in *R. opacus*. pLPD06764 (pHBA), pLPD00563 (pVAN), pLPD06580 (pBEN), and pLPD06578 (pGUA) were placed upstream of GFP+. **A)** Normalized fluorescence of the strain containing pHBA-GFP+ in response to 0, 0.001, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, and 1 g/L 4-hydroxybenzoic acid (HBA). An increase in normalized fluorescence of 6-fold was observed between 0 and 1 g/L HBA. **B)** Normalized fluorescence of the strain containing pVAN-GFP+ in response to 0, 0.001, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 1, and 1.5 g/L vanillic acid (VAN). The fluorescence value of pVAN at 1 g/L glucose was indistinguishable from the background level (i.e. within one standard deviation) and a fold change could not be calculated. **C)** Normalized fluorescence of the strain containing pBEN-GFP+ in response to 0, 0.001, 0.01,

0.05, 0.1, 0.25, 0.5, 1, 2.5, and 5 g/L sodium benzoate (BEN). An increase in normalized fluorescence of 35-fold was observed between 0 and 5 g/L BEN. **D)** Normalized fluorescence of the strain containing pGUA-GFP+ in response to 0, 0.001, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 1, and 1.5 g/L guaiacol (GUA). An increase in normalized fluorescence of 137-fold was observed from 0 to 1.5 g/L GUA. The solid black line in each case (A-D) represents a fitted curve (see Supplementary Methods B and Supplementary Table B.4). Cultures in A-D were grown in 1 g/L glucose. **E-H)** Selective response of aromatic sensors. The fluorescent response of each construct (E: pHBA; F: pVAN; G: pBEN; and H: pGUA) was tested in the presence of 1 g/L glucose (GLU); or the combination of 1 g/L GLU and either 0.3 g/L phenol (PHE) or equimolar amount of protocatechuic acid (PCA), HBA, VAN, BEN, or GUA. Values are averages of three replicates, and error bars represent one standard deviation. An asterisk (*) represents that there was no statistical difference in normalized fluorescence between the experimental strain and the control strain containing an empty control plasmid.

3.4 Conclusion

The availability of well-characterized genetic parts in non-model microbes, such as *R. opacus*, is a common hindrance to the successful engineering of these organisms. In this work, multiple genetic logic circuits (AND, NAND, and IMPLY) were constructed in *R. opacus* using newly developed and previously reported parts. To enable predictable and high-activity gene transcription, a T7 RNAP-based expression system, under the control of either a phenol- or an aTc-inducible promoter, was integrated into the chromosome. This is the first demonstration of the T7 RNAP system in *Rhodococcus* spp. and is an important step toward developing *R. opacus* as a robust industrial strain. Furthermore, the ability to use the T7 RNAP platform for gene expression facilitates potential future developments, including the implementation of a high-throughput, cell-free platform for prototyping pathways prior to their insertion into *R. opacus*, such as that developed for *Pseudomonas putida*.[31, 48-49] Additionally, three synthetic IPTG-dependent promoters, including a version capable of leaderless transcription, were developed. These are the first IPTG-dependent promoters described in *Rhodococcus* spp. and present a significant addition to the promoter toolbox. Finally, four novel aromatic sensors for 4-hydroxybenzoic acid, vanillic acid, sodium benzoate, and guaiacol were developed for use in future genetic circuits and for future

lignin valorization. As the number of genetic parts for *R. opacus* increases, building complex genetic circuits in this host becomes easier, enabling sophisticated gene regulation for complicated cellular functions required for future applications.

3.5 Materials and methods

3.5.1 Strains, plasmids and growth conditions

All plasmids were assembled in *E. coli* DH10B using the Gibson assembly method.[50] Kanamycin (20 μg/mL), chloramphenicol (34 μg/mL), or hygromycin B (200 μg/mL) was added as appropriate to *E. coli* cultures. Kanamycin (50 μg/mL), chloramphenicol (15 μg/mL), and/or hygromycin B (50 μg/mL) were added as appropriate to *R. opacus* cultures. Electrocompetent *E. coli* DH10B and *R. opacus* cells were transformed as previously described.[51-52] Plasmid DNA was isolated using Zyppy™ Plasmid Miniprep Kit (Zymo), and PCR products were extracted from electrophoresis gels using Zymoclean™ Gel DNA Recovery Kit (Zymo). DNA sequencing was performed by Genewiz. Chromosomal integrations in *R. opacus* were performed as previously described using a helper plasmid expressing two recombinases to facilitate homologous recombination.[20] All genetic part sequences, plasmids, and strains used in this study are listed in Supplementary Tables B.1, B.2, and B.3, respectively. pTara:500 and pET28:GFP were gifts from Prof. Matthew Bennett (Addgene plasmids #60717 and #60733). Enzymes were purchased from New England Biolabs. Chemicals were purchased from Sigma-Aldrich.

R. opacus was cultured in tryptic soy broth (TSB) or minimal media B as previously described at 30 °C and 250 rpm using 50 mL glass culture tubes.[10, 53] Unless otherwise noted, minimal media B was supplemented with 4 g/L glucose as the carbon source and 1 g/L ammonium sulfate as the nitrogen source. For all cultures, a single colony was re-streaked from a TSB agar plate onto a new TSB agar plate containing appropriate antibiotics and incubated for 48 hours at

30 ºC. A loopful of cells was used to start a 3 mL seed culture in minimal media B, except for strain DMD377 [NAND circuit] which was grown in TSB with 10 ng/mL aTc, and incubated until cells reached early stationary phase. Main experimental cultures were performed in triplicate at a 5 mL scale, with cultures diluted to an initial optical density at 600 nm (OD_{600}) of ~0.2, unless otherwise noted. The DMD377 seed culture (NAND circuit) was washed with minimal media prior to use in the main experimental culture to remove residual aTc. Inducers (phenol, anhydrotetracycline [aTc], isopropyl β-D-1-thiogalactopyranoside [IPTG], protocatechuic acid [PCA], 4-hydroxybenzoic acid [HBA], sodium benzoate [BEN], or guaiacol [GUA]) were added at the beginning of the main experimental cultures.

3.5.2 Growth and fluorescence measurements

The optical density (OD_{600}) of cultures was measured in VWR semi-micro polystyrene cuvettes using a Tecan Infinite M200Pro plate reader. Cell fluorescence and absorbance at 600 nm were measured at late exponential phase in black-walled 96-well plates (Greiner Bio-One flat bottom, chimney well, μclear) using a Tecan Infinite M200Pro plate reader. The excitation and emission wavelengths for GFP+, eGFP, and mCherry were 502/536 nm, 490/525 nm, and 570/630 nm, respectively. Fluorescence measurements were normalized using Equation 1, where $Fluo_{norm}$ is the normalized fluorescence, $Abs600_{sample}$ is the test strain absorbance at 600 nm, $Fluo_{sample}$ is the test strain fluorescence, $Fluo_{control}$ is the empty vector control strain fluorescence, and $Abs600_{control}$ is the empty vector control strain absorbance at 600 nm.[10]

Eq. 1)
$$Fluo_{norm} = \frac{Fluo_{sample}}{Abs600_{sample}} - \frac{Fluo_{control}}{Abs600_{control}}$$

3.6 Supplementary materials

Supplementary Tables B.1-4, Supplementary Figures B.1-6, Supplementary Methods B.

3.7 References

1. Wu, G.; Yan, Q.; Jones, J. A.; Tang, Y. J.; Fong, S. S.; Koffas, M. A. G., Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol.* **2016,** *34* (8), 652-664.

2. Xu, P.; Li, L.; Zhang, F.; Stephanopoulos, G.; Koffas, M., Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proc Natl Acad Sci U S A* **2014,** *111* (31), 11299-304.

3. Cress, B. F.; Trantas, E. A.; Ververidis, F.; Linhardt, R. J.; Koffas, M. A., Sensitive cells: enabling tools for static and dynamic control of microbial metabolic pathways. *Curr. Opin. Biotechnol.* **2015,** *36*, 205-14.

4. Hoynes-O'Connor, A.; Moon, T. S., Programmable genetic circuits for pathway engineering. *Curr. Opin. Biotechnol.* **2015,** *36*, 115-121.

5. Miyamoto, T.; Razavi, S.; DeRose, R.; Inoue, T., Synthesizing biomolecule-based Boolean logic gates. *ACS synthetic biology* **2013,** *2* (2), 72-82.

6. Hoynes-O'Connor, A.; Shopera, T.; Hinman, K.; Creamer, J. P.; Moon, T. S., Enabling complex genetic circuits to respond to extrinsic environmental signals. *Biotechnology and bioengineering* **2017,** *114*, 1626-1631.

7. Immethun, C. M.; Ng, K. M.; DeLorenzo, D. M.; Waldron-Feinstein, B.; Lee, Y. C.; Moon, T. S., Oxygen-responsive genetic circuits constructed in synechocystis sp. PCC 6803. *Biotechnology and bioengineering* **2016,** *113*, 433-442.

8. Nielsen, A. A.; Der, B. S.; Shin, J.; Vaidyanathan, P.; Paralanov, V.; Strychalski, E. A.; Ross, D.; Densmore, D.; Voigt, C. A., Genetic circuit design automation. *Science* **2016,** *352* (6281), aac7341.

9. Jin, H.; Lindblad, P.; Bhaya, D., Building an Inducible T7 RNA Polymerase/T7 Promoter Circuit in Synechocystis sp. PCC6803. *ACS synthetic biology* **2019,** *8* (4), 655-660.

10. DeLorenzo, D. M.; Henson, W. R.; Moon, T. S., Development of Chemical and Metabolite Sensors for Rhodococcus opacus PD630. *ACS synthetic biology* **2017,** *6* (10), 1973-1978.

11. Russell, J. J.; Theriot, J. A.; Sood, P.; Marshall, W. F.; Landweber, L. F.; Fritz-Laylin, L.; Polka, J. K.; Oliferenko, S.; Gerbich, T.; Gladfelter, A.; Umen, J.; Bezanilla, M.; Lancaster, M. A.; He, S.; Gibson, M. C.; Goldstein, B.; Tanaka, E. M.; Hu, C.-K.; Brunet, A., Non-model model organisms. *BMC Biol.* **2017,** *15* (55).

12. Yan, Q.; Fong, S. S., Challenges and Advances for Genetic Engineering of Non-model Bacteria and Uses in Consolidated Bioprocessing. *Front Microbiol* **2017,** *8*, 2060.

13. Waltermann, M.; Luftmann, H.; Baumeister, D.; Kalscheuer, R.; Steinbuchel, A., Rhodococcus opacus strain PD630 as a new source of high-value single-cell oil? Isolation and characterization of triacylglycerols and other storage lipids. *Microbiology* **2000,** *146 (Pt 5)*, 1143-9.

14. Alvarez, H. M.; Mayer, F.; Fabritius, D.; Steinbuchel, A., Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630. *Arch Microbiol* **1996,** *165* (6), 377-86.

15. Kurosawa, K.; Wewetzer, S. J.; Sinskey, A. J., Triacylglycerol Production from Corn Stover Using a Xylose-Fermenting Rhodococcus opacus Strain for Lignocellulosic Biofuels. *Journal of Microbial & Biochemical Technology* **2014,** *6* (5), 254-259.

16. Yoneda, A.; Henson, W. R.; Goldner, N. K.; Park, K. J.; Forsberg, K. J.; Kim, S. J.; Pesesky, M. W.; Foston, M.; Dantas, G.; Moon, T. S., Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630. *Nucleic Acids Res* **2016,** *44* (5), 2240-54.

17. Henson, W. R.; Campbell, T.; DeLorenzo, D. M.; Gao, Y.; Berla, B.; Kim, S. J.; Foston, M.; Moon, T. S.; Dantas, G., Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus. *Metab. Eng.* **2018,** *49*, 69-83.

18. Kurosawa, K.; Laser, J.; Sinskey, A. J., Tolerance and adaptive evolution of triacylglycerol-producing Rhodococcus opacus to lignocellulose-derived inhibitors. *Biotechnology for Biofuels* **2015,** *8* (76).

19. He, Y.; Li, X.; Ben, H.; Xue, X.; Yang, B., Lipid Production from Dilute Alkali Corn Stover Lignin by Rhodococcus Strains. *ACS Sustainable Chemistry & Engineering* **2017,** *5* (3), 2302-2311.

20. DeLorenzo, D. M.; Rottinghaus, A. G.; Henson, W. R.; Moon, T. S., Molecular Toolkit for Gene Expression Control and Genome Modification in Rhodococcus opacus PD630. *ACS synthetic biology* **2018,** *7* (2), 727-738.

21. DeLorenzo, D. M.; Moon, T. S., Selection of stable reference genes for RT-qPCR in Rhodococcus opacus PD630. *Sci Rep* **2018,** *8* (1), 6019.

22. Bienick, M. S.; Young, K. W.; Klesmith, J. R.; Detwiler, E. E.; Tomek, K. J.; Whitehead, T. A., The interrelationship between promoter strength, gene expression, and growth rate. *PLoS ONE* **2014,** *9* (10), e109105.

23. Hetzler, S.; Broker, D.; Steinbuchel, A., Saccharification of Cellulose by Recombinant Rhodococcus opacus PD630 Strains. *Appl Environ Microbiol* **2013,** *79* (17), 5159-66.

24. Xiong, X.; Wang, X.; Chen, S., Engineering of a xylose metabolic pathway in Rhodococcus strains. *Appl Environ Microbiol* **2012,** *78* (16), 5483-91.

25. Marbach, A.; Bettenbrock, K., lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *J Biotechnol* **2012,** *157* (1), 82-8.

26. Bahl, C. P.; Wu, R.; Stawinsky, J.; Narang, S. A., Minimal length of the lactose operator sequence for the specific recognition by the lactose repressor. *Proc Natl Acad Sci U S A* **1977,** *74* (3), 966-70.

27. Brosius, J.; Holy, A., Regulation of ribosomal RNA promoters with a synthetic lac operator. *Proc Natl Acad Sci U S A* **1984,** *81* (22), 6929-33.

28. Yansura, D. G.; Henner, D. J., Use of the Escherichia coli lac repressor and operator to control gene expression in Bacillus subtilis. *Proc Natl Acad Sci U S A* **1984,** *81* (2), 439-43.

29. Shultzaberger, R. K.; Chen, Z.; Lewis, K. A.; Schneider, T. D., Anatomy of Escherichia coli sigma70 promoters. *Nucleic Acids Res* **2007,** *35* (3), 771-88.

30. Studier, F. W.; Moffatt, B. A., Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **1986,** *189* (1), 113-30.

31. Kim, D. M.; Kigawa, T.; Choi, C. Y.; Yokoyama, S., A highly efficient cell-free protein synthesis system from Escherichia coli. *Eur J Biochem* **1996,** *239* (3), 881-6.

32. Baneyx, F., Recombinant protein expression in Escherichia coli. *Curr. Opin. Biotechnol.* **1999,** *10* (5), 411-21.

33. Graumann, K.; Premstaller, A., Manufacturing of recombinant therapeutic proteins in microbial systems. *Biotechnology journal* **2006,** *1* (2), 164-86.

34. Meyer, A. J.; Ellefson, J. W.; Ellington, A. D., Directed Evolution of a Panel of Orthogonal T7 RNA Polymerase Variants for in Vivo or in Vitro Synthetic Circuitry. *ACS synthetic biology* **2015,** *4* (10), 1070-6.

35. Bandwar, R. P.; Jia, Y.; Stano, N. M.; Patel, S. S., Kinetic and thermodynamic basis of promoter strength: multiple steps of transcription initiation by T7 RNA polymerase are modulated by the promoter sequence. *Biochemistry* **2002,** *41* (11), 3586-95.

36. Shis, D. L.; Bennett, M. R., Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proc Natl Acad Sci U S A* **2013,** *110* (13), 5028-33.

37. Iyer, S.; Karig, D. K.; Norred, S. E.; Simpson, M. L.; Doktycz, M. J., Multi-input regulation and logic with T7 promoters in cells and cell-free systems. *PLoS ONE* **2013,** *8* (10), e78442.

38. Ghosh, T.; Bose, D.; Zhang, X., Mechanisms for activating bacterial RNA polymerase. *FEMS Microbiol. Rev.* **2010,** *34* (5), 611-27.

39. Dubendorff, J. W.; Studier, F. W., Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J Mol Biol* **1991,** *219* (1), 45-59.

40. Moon, T. S.; Lou, C.; Tamsir, A.; Stanton, B. C.; Voigt, C. A., Genetic programs constructed from layered logic gates in single cells. *Nature* **2012,** *491* (7423), 249-53.

41. Dong, L.; Nakashima, N.; Tamura, N.; Tamura, T., Isolation and characterization of the Rhodococcus opacus thiostrepton-inducible genes tipAL and tipAS: application for recombinant protein expression in Rhodococcus. *FEMS Microbiol. Lett.* **2004,** *237* (1), 35-40.

42. Hernandez, M. A.; Lara, J.; Gago, G.; Gramajo, H.; Alvarez, H. M., The pleiotropic transcriptional regulator NlpR contributes to the modulation of nitrogen metabolism, lipogenesis and triacylglycerol accumulation in oleaginous rhodococci. *Mol Microbiol* **2017,** *103* (2), 366-385.

43. Nga, D. P.; Altenbuchner, J.; Heiss, G. S., NpdR, a repressor involved in 2,4,6-trinitrophenol degradation in Rhodococcus opacus HL PM-1. *J Bacteriol* **2004,** *186* (1), 98-103.

44. Ellinger, J.; Schmidt-Dannert, C., Construction of a BioBrick compatible vector system for Rhodococcus. *Plasmid* **2017,** *90*, 1-4.

45. He, S.; Wurtzel, O.; Singh, K.; Froula, J. L.; Yilmaz, S.; Tringe, S. G.; Wang, Z.; Chen, F.; Lindquist, E. A.; Sorek, R.; Hugenholtz, P., Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nat. Methods* **2010,** *7* (10), 807-12.

46. Jiao, S.; Yu, H.; Shen, Z., Core element characterization of Rhodococcus promoters and development of a promoter-RBS mini-pool with different activity levels for efficient gene expression. *N Biotechnol* **2018,** *44*, 41-49.

47. Rytter, J. V.; Helmark, S.; Chen, J.; Lezyk, M. J.; Solem, C.; Jensen, P. R., Synthetic promoter libraries for Corynebacterium glutamicum. *Appl. Microbiol. Biotechnol.* **2014,** *98* (6), 2617-23.

48. Karim, A. S.; Jewett, M. C., Cell-Free Synthetic Biology for Pathway Prototyping. *Methods Enzymol* **2018,** *608*, 31-57.

49. Wang, H.; Jewett, M. C.; Li, J., Development of a Pseudomonas putida cell-free protein synthesis platform for rapid screening of gene regulatory elements. *Synthetic Biology* **2018,** *3* (1).

50. Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A., 3rd; Smith, H. O., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **2009,** *6* (5), 343-5.

51. Durfee, T.; Nelson, R.; Baldwin, S.; Plunkett, G., 3rd; Burland, V.; Mau, B.; Petrosino, J. F.; Qin, X.; Muzny, D. M.; Ayele, M.; Gibbs, R. A.; Csorgo, B.; Posfai, G.; Weinstock, G. M.; Blattner, F. R., The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. *J Bacteriol* **2008,** *190* (7), 2597-606.

52. Kalscheuer, R.; Arenskotter, M.; Steinbuchel, A., Establishment of a gene transfer system for Rhodococcus opacus PD630 based on electroporation and its application for recombinant biosynthesis of poly(3-hydroxyalkanoic acids). *Appl. Microbiol. Biotechnol.* **1999,** *52* (4), 508-15.

53. Hollinshead, W. D.; Henson, W. R.; Abernathy, M.; Moon, T. S.; Tang, Y. J., Rapid metabolic analysis of Rhodococcus opacus PD630 via parallel 13C-metabolite fingerprinting. *Biotechnol. Bioeng.* **2016,** *113* (1), 91-100.

Chapter 4: Molecular toolkit for gene expression control and genome modification in *Rhodococcus opacus* PD630

Reprinted with permission from DeLorenzo, D.M., Rottinghaus, A.G., Henson, W.R., Moon, T.S. Molecular toolkit for gene expression control and genome modification in *Rhodococcus opacus* PD630. ACS Synthetic Biology (2018). 7, 727−738.

The previous two chapters have detailed the development of genetic parts for tunable gene expression. More genetic elements, however, are required for heterologous gene expression than just promoters. The following chapter will discuss the characterization of tools that facilitate the introduction of heterologous DNA into *R. opacus*, including plasmid backbones, antibiotic resistance markers, and neutral integration sites in the genome. Furthermore, a methodology for genome recombineering was developed for *R. opacus* to enable stable gene integrations or gene knockouts. Finally, while the preceding chapters have dealt with gene expression, this chapter will discuss a newly developed tool for targeted gene repression (CRISPRi). I performed or directed all experiments and wrote the manuscript.

4.1 Abstract

Rhodococcus opacus PD630 is a non-model, gram-positive bacterium that possesses desirable traits for lignocellulosic biomass conversion. In particular, it has a relatively rapid growth rate, exhibits genetic tractability, produces high quantities of lipids, and can tolerate and consume toxic, lignin-derived aromatic compounds. Despite these unique, industrially relevant characteristics, *R. opacus* has been underutilized due to a lack of reliable genetic parts and engineering tools. In this work, we developed a molecular toolbox for reliable gene expression control and genome modification in *R. opacus*. To facilitate predictable gene expression, a

constitutive promoter library spanning ~45-fold in output was constructed. To improve the characterization of available plasmids, the copy numbers of four heterologous and nine endogenous plasmids were determined using quantitative PCR. The molecular toolbox was further expanded by screening a previously unreported antibiotic resistance marker (HygR) and constructing a curable plasmid backbone for temporary gene expression (pB264). Furthermore, a system for genome modification was devised, and three neutral integration sites were identified using a novel combination of transcriptomic data, genomic architecture, and growth rate analysis. Finally, the first reported system for targeted, tunable gene repression in *Rhodococcus* was developed by utilizing CRISPR interference (CRISPRi). Overall, this work greatly expands the ability to manipulate and engineer *R. opacus*, making it a viable new chassis for bioproduction from renewable feedstocks.

4.2 Introduction

Non-model organisms are gaining traction as new chassis for addressing questions that traditional model organisms are not well suited to answer.[1, 2] Recent advances in "omics" analysis (whole genome sequencing, transcriptomics, etc.) and developments in molecular biology techniques have made previously inaccessible organisms with unique properties available to investigate and engineer.[1, 3] Non-model organisms possess a vast array of novel enzymes, adaptation abilities, and metabolic networks that can help address challenges in bioproduction from renewable feedstocks.[1, 2, 4] A field of study that may benefit from the use of new microbial chassis is the bioconversion of lignocellulosic biomass into higher value bioproducts. The pretreatment and depolymerization of lignocellulose, particularly lignin, generates an array of toxic, aromatic compounds that traditional model organisms are not able to effectively tolerate or consume.[5-7]

Rhodococcus opacus PD630 (hereafter *R. opacus*) is a gram-positive, oleaginous Actinobacterium that possesses a native and adaptable tolerance towards biomass breakdown products (e.g. furans, organic acids, halogenated compounds, and phenolics), and diverse enzymatic pathways for aromatic compound consumption.[8-14] Furthermore, *R. opacus* can dedicate a large portion of its cellular resources to lipid production (up to ~78% triacylglycerol [TAG] of cell dry weight), has high flux β-ketoadipate and Entner Doudoroff pathways, and can consume multiple carbon sources simultaneously.[10, 15] This unique metabolic topology can be utilized in conjunction with catalytic depolymerization of lignocellulose for the production of lipid-based fuels and chemicals, or the production of petrochemical replacements (e.g. muconic acid).[16-21]

R. opacus has been previously engineered to consume different fractions of biomass (e.g. xylose, arabinose, and cellobiose) or to produce simple bio-based products (e.g. TAG and wax esters), but the modifications were achieved using borrowed and generally uncharacterized genetic parts from related Actinobacteria.[22-26] Recent work has sought to develop and characterize a number of inducible promoters in *R. opacus* for static and dynamic control of gene expression, including those responsive to aromatics, ammonium, acetamide, anhydrotetracycline (aTc), and arabinose.[13, 27] A summary of genetic parts available for *R. opacus*, including plasmid backbones, promoters, antibiotic resistance markers, and reporters can be found in Table 4.1.

To further expand the available genetic toolbox in *R. opacus*, this work seeks to develop and characterize parts and methodologies related to gene expression control and genome modification. A constitutive promoter library spanning a ~45-fold range in strength was developed to facilitate predicable expression of heterologous genes. As described in Table 4.1, several heterologous plasmids are known to be stably maintained in *R. opacus*, but the number of copies maintained per cell was unknown. The numbers of plasmid copies for two common backbones

(pNG2 and pAL5000), the newly characterized pB264, and the nine endogenous plasmids maintained in *R. opacus*, were measured using quantitative PCR (qPCR). Two antibiotic resistance markers (chloramphenicol and hygromycin B) were either optimized for *R. opacus* or screened in this host to broaden the number of selection markers. To further expand the ability to express heterologous genes or to modify the genome, a recombination-based system for genomic modification was developed by repurposing two bacteriophage recombinases (Che9c60 and Che9c61). By using both a previously published genome assembly and transcriptomic data, in addition to experimental measurements, three neutral sites where expression cassettes can be integrated without growth defect were identified. Finally, a system for tunable, targeted gene repression was developed for the first time in *Rhodococcus* by utilizing CRISPR interference (CRISPRi). Together, these new genetic tools provide the groundwork for the further development of *R. opacus* as an improved chassis for renewable bioproduction.

Table 4.1. Published genetic elements previously used in *R. opacus*. Common naming conventions of plasmid backbones containing the same or similar origins of replication are given due to various nomenclature changes over decades of publication. The two pAL5000 variants (S and L) are derived from the same *Mycobacterium fortuitum* plasmid and are not compatible.[85] The ancestrally related pNG2 and pGA1 rolling circle origins of replication are derived from cryptic *Corynebacterium diphtheriae* and *Corynebacterium glutamicum* plasmids, respectively, and share ~50% replication protein identity.[48, 86-88] The pAL5000 and pNG2 origins of replication are compatible and maintained concurrently (data not shown).

Genetic Part	**Note**
Plasmid Backbones	pXYLA/pNV18 (short variant of pAL5000; S)[85, 89-93]; pJAM2/pJEM (long variant of pAL5000; L)[29, 85, 92]; pNG2/pAL358/pEP2[22, 86]; pGA1[25, 29, 87, 94]
Promoters	pAcet[13] (acetamide-inducible); pTipA[95] (thiostrepton-inducible); pLac[29] (constitutive); pTac[89] (constitutive); pBAD[13] (arabinose-inducible); pTet[13] (aTc-inducible); a suite of aromatic inducible promoters[13]
Selection markers	kanamycin (50 μg/mL)[64, 96]; gentamicin (10 μg/mL)[22]; spectinomycin (100 μg/mL)[22]; thiostrepton[64]; chloramphenicol (34 μg/mL)[26]
Reporters	eGFP[97]; lacZ[97]; RFP[13]; sfGFP[13]; GFP+[13]; mCherry[13]; CFP[13]; EYFP[13]

4.3 Results and discussion

4.3.1 Constitutive promoter library

Characterized and predictable genetic parts are required when engineering and balancing cellular pathways.[28] One common component needed for synthetic biology is the promoter, which drives gene transcription. Previous work in *R. opacus* has used borrowed and uncharacterized constitutive promoters from related Actinomycetales, such as *Mycobacterium* spp. and *Streptomyces* spp., or gram-negative bacteria, such as *E. coli*, for heterologous gene expression.[22, 25, 29] Recently, several inducible promoters (e.g. pBAD, pTet, pAcet, etc.) were characterized in *R. opacus* (Table 4.1), but these require the use of a chemical inducer to modulate gene expression, which can be expensive at an industrial scale and lead to heterogeneous expression levels.[13, 28] An alternative is to have a well characterized and predictable constitutive promoter library that spans a range of expression levels. The benefit of a well-defined promoter library was recently demonstrated when a metabolic pathway was rapidly optimized by combinatorial assembly of each enzyme with a number of constitutive promoters of varying strengths.[30]

To engineer such a library, a strong synthetic promoter from *Streptomyces lividans* TK24 (pConstitutive)[31] was cloned in front of the gene encoding GFP+ (a modified green fluorescent protein with improved folding efficiency and fluorescence yield[13, 32, 33]), and saturation mutagenesis was performed on the -35 and -10 sites, either individually or concurrently. A subset of 25 new promoters was selected that spans a fluorescent output range of ~45-fold from the weakest to strongest promoter (Figure 4.1). All promoter sequences are listed in Supplementary Table B.1. Interestingly, none of the tested promoters exhibited significantly higher fluorescent output than the original promoter (pConstitutive). To develop constitutive promoters with very high levels of expression in *Streptomyces* strains, researchers have screened endogenous promoters

mined from the genome, or randomized sequences surrounding the -35 and -10 sites to better suit the organism's sigma factor.[34-36] Future work in *R. opacus* could pursue a similar path to develop even stronger promoters than those reported here and further expand the ability to reliably express heterologous genes or non-coding elements.

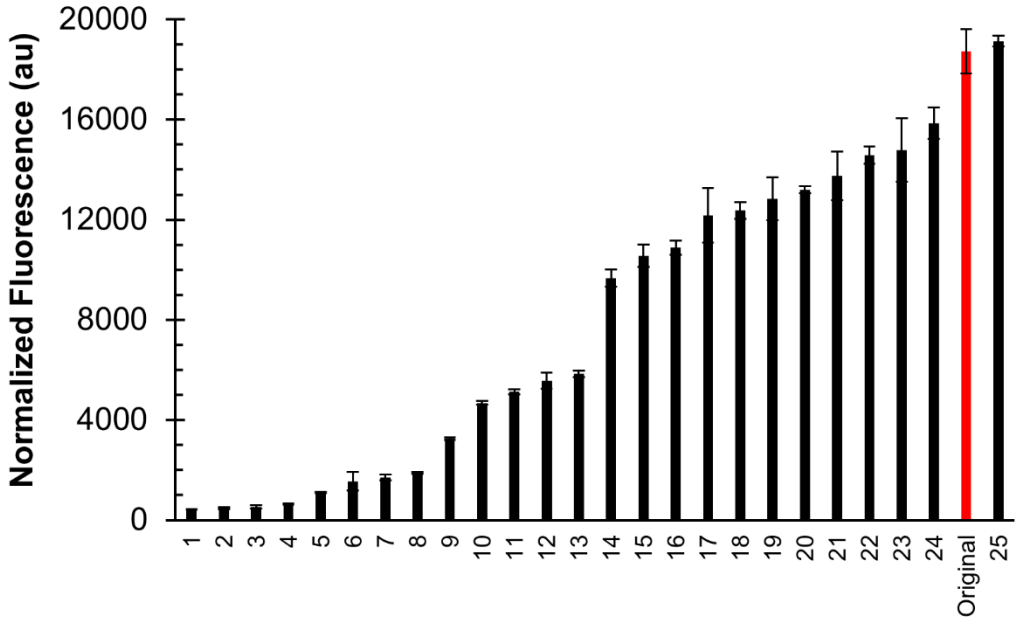


Figure 4.1. Constitutive promoter library in *R. opacus*. The -10 and -35 regions of a strong constitutive promoter (Original or pConstitutive) transcribing GFP+ underwent saturation mutagenesis to create a library of constitutive promoters (see 4.5 Materials and Methods).[31, 32] Normalized fluorescence (see 4.5 Materials and Methods) spans ~45-fold from the weakest to strongest promoter. All promoter sequences are listed in Supplementary Table B.1. Values represent the average of three replicates grown in minimal media A, and error bars represent one standard deviation.

4.3.2 Heterologous and endogenous plasmid copy numbers

The overexpression of genes is typically performed using self-replicating plasmids that can be maintained at multiple copies per cell. The number of copies of a plasmid per cell, typically measured as copies per chromosome, is determined predominately by the origin of replication.[37]

In addition to the heterologous origins described in Table 4.1, *R. opacus* also possesses nine endogenous plasmids that range from 37 to 172 kilo base pairs (kbp) in size.[38] The role of these large plasmids is generally unknown, as 54.5% of all their annotated genes (733) are listed as hypothetical proteins (NCBI reference sequence NZ_CP003949.1). However, there are annotated pathways for hetero- and polycyclic aromatics (e.g. biphenyl and chlorophenol) in addition to many mono- and dioxygenases. Other *Rhodococcus* spp. host a number of catabolic pathways on their plasmids for compounds such as alkanes, aromatics, herbicides, and halogenated compounds.[39, 40] Interestingly, seven of the native plasmids are linear (endogenous plasmid 3-9).[38, 41] Although plasmids are commonly believed to be exclusively circular, linear plasmids have also been observed in related Actinomycetales, including *Streptomyces* spp. and *Mycobacterium* spp.[39, 42] The frequent observation of linear plasmids and occurrence of illegitimate recombination in Actinobacteria has been hypothesized to be related to a hyper-recombinational gene storage strategy.[39, 41, 43, 44] In other words, Actinobacteria such as *Rhodococcus* spp. store a large number of catabolic genes for compounds, which they may encounter in their native soil environment, as future recombination sources upon adaptation.

For the purpose of engineering, all plasmids, both heterologous and endogenous, can be used for gene overexpression, making it critical to quantify the number of plasmid copies relative to the chromosome number. qPCR was used to quantify the absolute number of chromosomal DNA and plasmid DNA copies (Figure 4.2).[45] *R. opacus* is believed to have a single copy chromosome. The endogenous plasmids had 1 to 2 plasmid copies per chromosome, except plasmid 8 which had ~5 copies per chromosome. This observation is consistent with the hypothesis that these plasmids are being maintained as gene storage for future recombination with the chromosome, as a low copy number would reduce replication costs while increasing genetic

potential. The two variants of pAL5000, originally isolated from *Mycobacterium* sp., had significantly different numbers: ~11 and ~3 copies per chromosome for the short (S) and long versions (L), respectively.[22, 32] This result is generally consistent with the low copy number observed in *Mycobacterium smegmatus*, where pAL5000 variants have ~5 copies per cell.[46] pNG2, originally isolated from *Corynebacterium diphtheriae*, demonstrated ~10 copies per chromosome, which is consistent with findings of low to medium copy numbers in *Corynebacterium glutamicum.*[47, 48] pB264, derived from a plasmid originally isolated from *Rhodococcus* sp. B264, was found to have ~8 copies per chromosome in *R. opacus*.[49] The origin of replication on pB264 shares homology with the pAL5000 family of origins, although from a divergent evolutionary lineage.[49] This is the first reported use of a minimal B264 replicon in *R. opacus*, expanding available origins of replication for *R. opacus* (see the Genome recombineering section for the utility of B264). While all the characterized plasmids herein demonstrate low copy numbers in *R. opacus* (1 - 11 copies per cell), a pAL5000 vector demonstrating high copy numbers in *Mycobacteria* (32 - 64 copies per cell) was created using a method that involved enrichment of high copy number plasmids by repeated isolation and transformation of mutated plasmids.[46] Future work in *R. opacus* could pursue a similar method to obtain higher copy number plasmids. In summary, the copy number of 13 uncharacterized plasmids in *R. opacus* was quantified for predictable heterologous gene expression, and a subset of these results may support a hypothesis regarding the role of large endogenous plasmids in *Rhodococcus* spp.

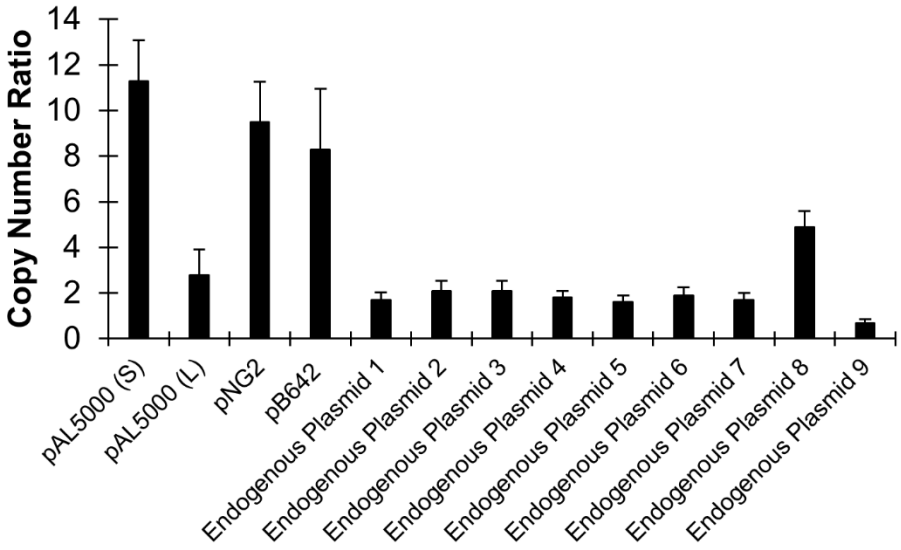


Figure 4.2. Copy number ratios of heterologous and endogenous plasmids in *R. opacus*. The numbers of copies of the four heterologous and nine endogenous plasmids relative to the single copy chromosome were determined by qPCR as described previously.[45] Values represent the average of three replicates grown in minimal media A, and error bars represent the propagated standard deviation (see B.3 Supplementary Methods). All qPCR primers and PCR amplification efficiencies are reported in Supplementary Table B.4.

4.3.3 Antibiotic resistance markers

Several antibiotic resistance cassettes have been previously described in *R. opacus* (Table 4.1). However, it is beneficial to have additional options for selection as the number of plasmid backbones and potential genomic integration sites expands. In this work, an optimized chloramphenicol resistance cassette was constructed, in addition to characterizing a previously unreported hygromycin B resistance marker in *R. opacus*. The first attempt to characterize chloramphenicol resistance in *R. opacus* failed to bestow antibiotic tolerance. As the marker was taken from an *E. coli* vector, there could have been differences in the gene GC content, codon usage, or promoter/RBS activity between the two organisms. To address GC content and codon usage variations, the chloramphenicol resistance gene was codon optimized for *R. opacus* that has a high GC content (Integrated DNA Technologies). To improve expression, a strong constitutive

promoter (pConstitutive; Figure 4.1) and a synthetic RBS was placed upstream of the optimized marker, while a synthetic terminator was placed downstream. The new cassette was placed on pNG2 and successfully restored growth (to wild type levels measured without antibiotics) at up to 30 μg/mL chloramphenicol (Figure 4.3A). In 2017, Lanfranconi and Alvarez reported the use of an alternate chloramphenicol resistance marker in *R. opacus*.[26] The second antibiotic marker tested in *R. opacus* was a hygromycin B resistance gene, which had not been previously reported in this organism. The hygromycin B resistance cassette on pAL5000 (L) restored growth (to wild type levels measured without antibiotics) at up to 100 μg/mL hygromycin B (Figure 4.3B). Part sequences are listed in Supplementary Table B.4.

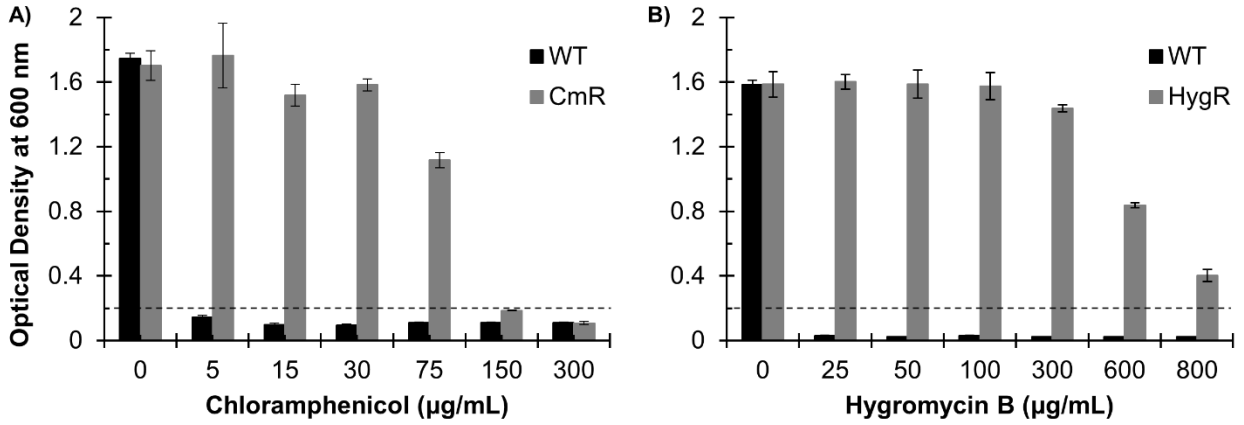


Figure 4.3. Antibiotic resistance cassettes for *R. opacus*. Dashed line represents the starting optical density ($OD_{600,\ initial}$ = 0.2). **A)** A chloramphenicol resistance marker (CmR) was codon optimized (Integrated DNA Technologies) for *R. opacus*. A strong promoter from the constitutive promoter library and a synthetic RBS was placed upstream, while a synthetic terminator was inserted downstream of the gene (see Supplementary Table B.2 for the cassette sequence). The optimized CmR cassette was placed on pNG2 and screened with 0 to 300 μg/mL chloramphenicol. The tolerance of wild type (WT) is also shown. **B)** A mycobacterial hygromycin B resistance (HygR) cassette on pAL5000 (L) was screened with 0 to 800 μg/mL hygromycin B.[32] Values represent the average of three replicates grown in minimal media A, and error bars represent one standard deviation.

4.3.4 Genome recombineering

Plasmids are not the only option for the expression of heterologous genes. The genome is also a candidate for gene integration, having the added benefit of enhanced genetic stability and no requirement for long term selection (e.g. addition of antibiotics). Additionally, the ability to modify the genome of an organism (e.g. gene deletion or replacement) in a site-specific manner is central to gene function analysis and strain engineering. Typically, double homologous recombination can be used in conjunction with a selection marker for site directed insertion or deletion of DNA (Figure 4.4).[50] The DNA to be introduced (referred to as the integration cassette) is designed to have flanking DNA homologous to the target site (homologous arm in Figure 4.4B).[44, 51] Many organisms, such as cyanobacteria, can readily facilitate recombination at the correct location with both circular and linear integration cassettes when a selection marker is utilized.[52-54] However, several Actinobacteria, including *Mycobacterium tuberculosis* and *Rhodococcus fascians*, demonstrate recombination of the integration cassette in the incorrect location (illegitimate integration).[43, 44] Our initial recombination experiments in *R. opacus* also showed that the cassettes recombined with the genome in the incorrect location (Supplementary Figure 4.6). Another method for site-specific recombination demonstrated in *E. coli* and *Streptomyces* spp. requires the use of a bacteriophage recombinase that recognizes an *attB* site in the genome and an *attP* site in the exogenous plasmid DNA and facilitates integration into the genome.[55-57] However, no such system has been demonstrated in *Rhodococcus* spp., and this system does not allow flexibility in the location of DNA integration.

To better facilitate recombination at the desired site, several techniques were applied to *R. opacus* for the first time. Previous work in *M. tuberculosis* found that a linearized integration cassette, in which the *E. coli* origin of replication (used for plasmid assembly) was removed, helped

to minimize illegitimate recombination (Supplementary Figure B.2).[44] Additionally, a pair of bacteriophage (Che9c) recombinases, Che9c60 and Che9c61 (GC rich homologs of RecE and RecT), were constitutively expressed to facilitate site specific recombination.[44] RecE is an exonuclease that produces a single stranded DNA overhang on the integration cassette, while RecT is a DNA binding protein that assists in strand invasion of the single stranded integration cassette overhang into the targeted DNA.[58] The two-gene recombinase operon (Che9c60 and Che9c61) was initially acquired on the plasmid pJV53 and was originally expressed using the acetamide promoter (pAcet) on an uncharacterized variant of pAL5000.[44] While these recombinases consistently facilitated site specific recombination, the stably maintained plasmid could not be cured post recombination even after the exclusion of selection pressure and multiple rounds of serial cultures (data not shown).

A plasmid that can be selectively cured out after successful recombination is required to remove the recombinase-containing plasmid. Previously, a cryptic plasmid from *Rhodococcus* sp. B264 was demonstrated to exhibit temperature sensitivity and could be cured out when the strain was grown at or above 37 ^{o}C.[49] Lessard *et al*. determined the minimal B264 replicon required for stable plasmid replication through a rigorous deletion assay.[49] This minimal replicon of 1500 bp was synthesized (Integrated DNA Technologies) and assembled with the *E. coli* origin pBR322 and a kanamycin resistance marker to create pB264. Interestingly, this plasmid could be cured just by removing selection pressure (kanamycin) and performing a round of colony purification; a 37 ^{o}C temperature passage was not required (Supplementary Figure B.4). The Che9c recombinases under pAcet were added to this new backbone. However, site specific recombination was no longer successful, presumably due to a change in the expression level of Che9c60 and Che9c61. The acetamide promoter was replaced with the strong pConstitutive promoter (Supplementary Figure

B.3), which restored the ability of the recombinases to facilitate site specific recombination. The multiple rounds of optimization required to implement a reliable genome recombineering system in *R. opacus* reiterate the inherent challenges in developing engineering tools for a non-model organism.

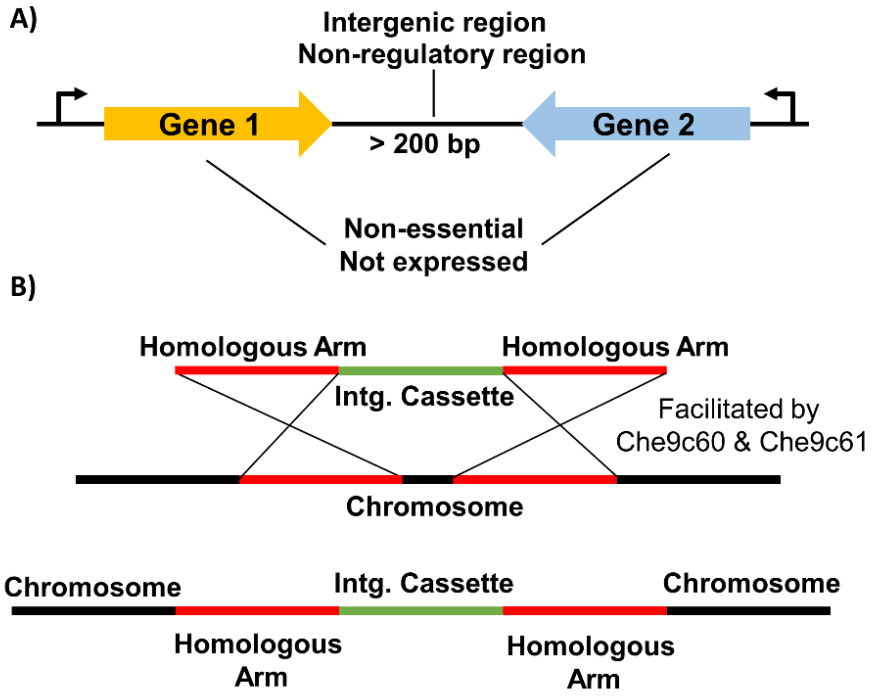


Figure 4.4. Neutral integration site identification and recombination methodology. **A)** A schematic detailing the criteria developed for identification of candidate neutral integration sites in the genome. A previously published *R. opacus* genome sequence from the NCBI database (Refseq, NZ_CP003949.1) was used to identify inward facing genes (sequential 5' to 3' and 3' to 5') with non-critical (hypothetical or putative) gene annotation and an intergenic region greater than 200 bp.[38] The intergenic region between two inward facing genes is anticipated to contain minimal regulatory components, besides transcriptional terminator sequences near 3' end of ORFs, and thus minimizes the probability of disturbing critical non-coding regions. Previously published transcriptomic data was additionally used to screen the candidate gene pairs and refine the list to only include sites with low or zero expression levels.[8] **B)** A schematic demonstrating double homologous recombination of a linearized double stranded integration cassette (Intg. Cassette) facilitated by the two heterologous recombinases Che9c60 and Che9c61.[44] The recombinases were constitutively expressed on the curable pB264 backbone, which is lost when antibiotic selection pressure is removed (see Supplementary Figure B.4).

4.3.5 Neutral integration sites

With a platform for site specific recombination available, neutral integration sites whose disruption does not have a negative impact on the host are necessary if heterologous genes are to be inserted into the genome. The definition of a neutral site is variable, but in this work and in other reports, it is defined as a site where gene insertion does not cause a reduction in growth rate relative to that of the wild type strain.[59-61] One approach to neutral site identification is to search the genome annotation for redundant, non-essential, or non-functional genes (e.g., pseudogenes).[59-61] This method may require additional trial and error due to the necessity of certain genes not always being clear based solely on the gene annotation. For example, pseudogenes may actually play a role as trans-acting antisense RNA and may be important for cellular regulation.[62] An alternative option is to use transcriptomic analysis or RT-qPCR to identify regions or loci of minimal transcription, with the assumption that disruption of these non-transcribed regions will have a minimal impact.[60, 61] This method alone does not consider that non-transcribed DNA may still contain critical non-coding elements (e.g. promoters, transcription factor binding sites, etc.) whose deletion may lead to unintended consequences.[63]

To maximize the likelihood of identifying neutral sites, a novel and rational set of criteria was developed based on the use of transcriptomic data and analysis of genomic architecture. To identify genes of negligible expression and importance, and to minimize the disruption of non-coding elements, we utilized previously published transcriptomic data of *R. opacus* grown under two distinct growth conditions and a complete genome assembly and annotation.[8, 38] The following list of criteria (Figure 4.4) was developed to identify neutral sites based on this multi-omics data such that integrations would not disrupt any open reading frames (ORFs) or regions upstream of ORFs, leading to a minimal effect on cell growth. 1) Transcriptomic data must identify two

consecutive ORFs with low or zero levels of transcription. 2) The ORFs must have non-essential or putative annotations (i.e. hypothetical protein). 3) An intergenic region of at least 200 bp must exist between the ORFs to avoid local context effects in case the nearby ORFs do have a function. 4) The ORFs must be convergent (5' to 3' followed by 3' to 5'). Regulatory regions typically occur upstream of an ORF (on the 5' side), and by selecting a region downstream (3') of both non-transcribed ORFs, the likelihood of disrupting potentially important non-coding DNA is reduced.

Using these criteria, four potential neutral sites were identified in the chromosome and an endogenous plasmid (Supplementary Table B.4) and are referred to as *R. opacus* chromosomal integration sites (ROCI-#) and *R. opacus* endogenous plasmid integration sites (ROP#I-#). Endogenous plasmid 8 was chosen for site selection based on our observation that it had the highest number of copies per chromosome (~5). A constitutively expressed EYFP gene along with a hygromycin B resistance marker was integrated into each of these sites. Site specific integration was confirmed using colony PCR (Supplementary Table B.5; Supplementary Figure B.5). *R. opacus* has a much lower transformation efficiency compared to a model organism like *E. coli*, and consequently only 2 to 4 colonies were recovered after electroporation with the linearized double stranded integration cassettes.[64, 65] However, all colonies tested had the correct PCR bands, signifying no false positive for all candidate neutral sites.

To determine site neutrality, a growth curve of the wild type strain and each recombinant strain was generated (Figure 4.5A), and the growth rate was calculated (Supplementary Table B.6). Three recombinant strains with integration into ROCI-2, ROCI-3, and ROP8I-1 had statistically indistinguishable growth rates compared to wild type (Student's t-test, $p < 0.05$), while one strain (ROCI-1) had an ~18% decrease in growth rate. All four sites are viable for integration, although only ROCI-2, ROCI-3, and ROP8I-1 are confirmed neutral sites. Thus, this targeted rational

approach to identifying neutral integration sites required fewer candidates to successfully determine multiple neutral locations than other reports using different screening metrics.[59-61] For example, among 16 candidate neutral sites in *Synechocystis* sp. PCC 6803, five integration sites were selected for growth tests after an initial assessment (e.g. RT-PCR).[61] A statistical analysis of the growth data showed that only two sites were neutral in terms of growth.

The fluorescence from each strain was also measured (Figure 4.5B), demonstrating successful gene function after recombination. The normalized fluorescence varied between the chromosomal sites, particularly ROCI-3 relative to ROCI-1 and ROCI-4. The location of the integration site, particularly in relation to the chromosome's origin of replication, has recently been found to influence gene copy numbers and expression levels in *E. coli* and *Bacillus subtilis*.[66-68] The fluorescence of ROP8I-1 was also significantly higher than that of the chromosomal sites, which is expected as Plasmid 8 was measured to have ~5 copies per chromosome. However, the fluorescence generated by ROP8I-1 is not five-fold higher than that of the chromosomal sites. Gene expression levels have been shown to be affected and vary based on genetic context, such as surrounding nucleotide sequences.[69, 70] Overall, the identification of these three neutral integration sites, in addition to the development of recombination plasmids that can be easily modified to insert any desired gene, will facilitate the development of *R. opacus* as a bioproduction strain.

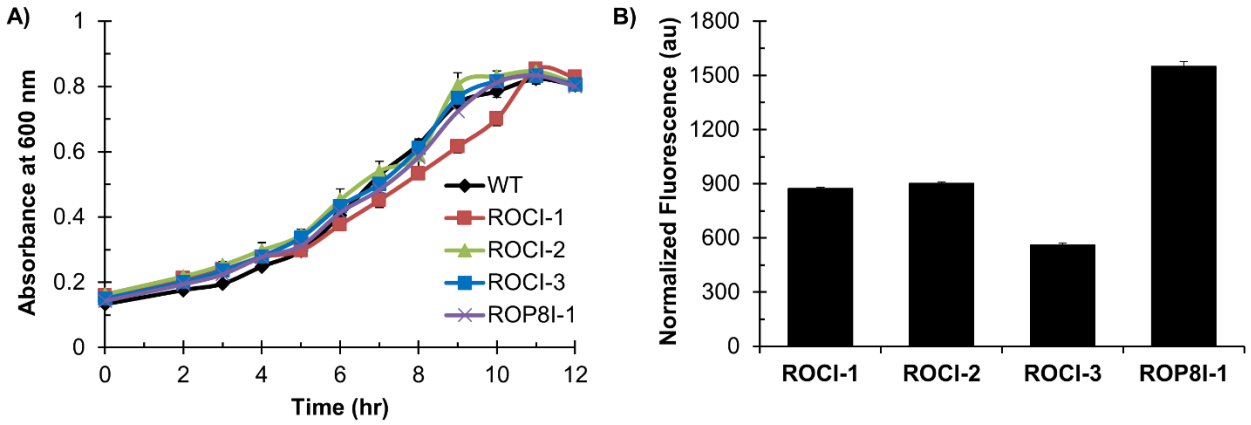


Figure 4.5. Neutral site characterization in *R. opacus*. Four candidate neutral sites were identified using the criteria described in Figure 4.4: three in the chromosome (LPD04028/LPD04029, LPD04690/LPD04691, and LPD03763/LPD03764) referred to as *R. opacus* chromosome integration sites 1 to 3 (ROCI-1 to ROCI-3); and one in endogenous plasmid 8 (LPD016184/LPD016185) referred to as *R. opacus* plasmid eight integration site 1 (ROP8I-1). A DNA cassette containing a constitutively expressed EYFP and HygR was integrated into each site. **A)** Growth curves of wild type *R. opacus* and strains containing integrations. The growth rates of ROCI-2, ROCI-3, and ROP8I-1 modified strains were statistically indistinguishable from the wild type growth rate (Students t-test, $p > 0.05$), while one recombinant strain (ROCI-1) had a ~18% growth rate reduction (Supplementary Table 6). **B)** Normalized fluorescence (see 4.5 Materials and Methods) produced from each strain. Values represent the average of three replicates grown in minimal media B, and error bars represent one standard deviation. Electrophoresis gel images confirming correct recombination events are shown in Supplementary Figure B.5.

4.3.6 CRISPRi-mediated gene repression

Genome recombineering can be useful for gene function analysis, but it is not an ideal tool if the goal is to only decrease gene expression rather than to delete or disrupt the gene. Two primary methods exist for prokaryotic gene repression, antisense RNA (asRNA) and CRISPR interference (CRISPRi), neither of which has been demonstrated in *Rhodococcus* spp.[71-74] As CRISPRi has been demonstrated in other Actinomycetales, it was chosen for our development efforts in *R. opacus*.[75-77] CRISPRi utilizes a catalytically dead version of the Cas9 nuclease (dCas9) to interfere with gene transcription.[78] The most widely used version of dCas9 originates from the AT rich *Streptococcus pyogenes* ($dCas9_{Spy}$).[72, 78] A previous study concluded that $dCas9_{Spy}$ did not function

efficiently in the related Actinomycetales *M. tuberculosis*, and in fact sensitized the cells to sub-minimum inhibitory concentrations of proteotoxic compounds.[75] After screening 11 Type IIA and Type IIC dCas9 homologs, they found that a codon optimized version of dCas9 from *Streptococcus thermophilus* ($dCas9_{Sth1}$) repressed gene transcription to the highest degree in *M. tuberculosis*.[75] The trade-off for $dCas9_{Sth1}$ is that the required PAM sequence is 7 base pairs compared to the $dCas9_{Spy}$ requirement of 3 base pairs, which drastically reduces the number of available target sites.[75] Notably, the $dCas9_{Sth1}$ PAM site is much more flexible than the rigid $dCas9_{Spy}$ PAM sequence, and Rock *et al*. demonstrated 19 different PAM sequences with 10-fold repression or higher.[75]

We placed the codon optimized $dCas9_{Sth1}$ under the arabinose inducible pBAD promoter on pAL5000 (S) and designed three sgRNAs (A, B, and C) based on rules discussed by Rock *et al.*[75] The sgRNAs were designed to target either the ribosome binding site (RBS) / 5' untranslated region (sgRNAs B and C) or the coding region (sgRNA A) of the constitutively expressed EYFP integrated into ROCI-2 (Supplementary Table B.7). Previous work with $dCas9_{Spy}$ has demonstrated that the repression ability is inversely correlated with the distance from the transcription start site (TSS).[72, 75, 78, 79] However, Rock *et al*. found that there was no correlation between the repression ability and the distance from the TSS when using $dCas9_{Sth1}$.[75] Originally, the sgRNAs were expressed using a pTet promoter optimized for *M. tuberculosis*, but after strains failed to demonstrate repression, the pTet promoter was replaced with the strong pConstitutive promoter. Eight concentrations of arabinose were used to demonstrate tunable repression of EYFP fluorescence (Figure 4.6), with percent repression ranges of 27-58%, 4-42% and 28-53% for sgRNAs A-C, respectively. The increased expression of $dCas9_{Sth1}$ did not lead to any growth defects between induction concentrations (Supplementary Figure B.7). This is the first reported

demonstration of a CRISPRi-based system for targeted and tunable gene repression in *Rhodococcus* spp. Our observations are also consistent with Rock *et al*.'s finding of the sgRNA binding position having no correlation with the repression ability, as sgRNA A targeted a site over 200 bp into the coding region, compared to sgRNAs B and C which both started binding within 15 bp of the TSS.[75] Future work would be a systematic investigation into the effect of diverse parameters (e.g. PAM sequences, target locations/strands, and sgRNA lengths/levels) on repression efficiency to increase the predictability of this gene repression tool in *R. opacus*.

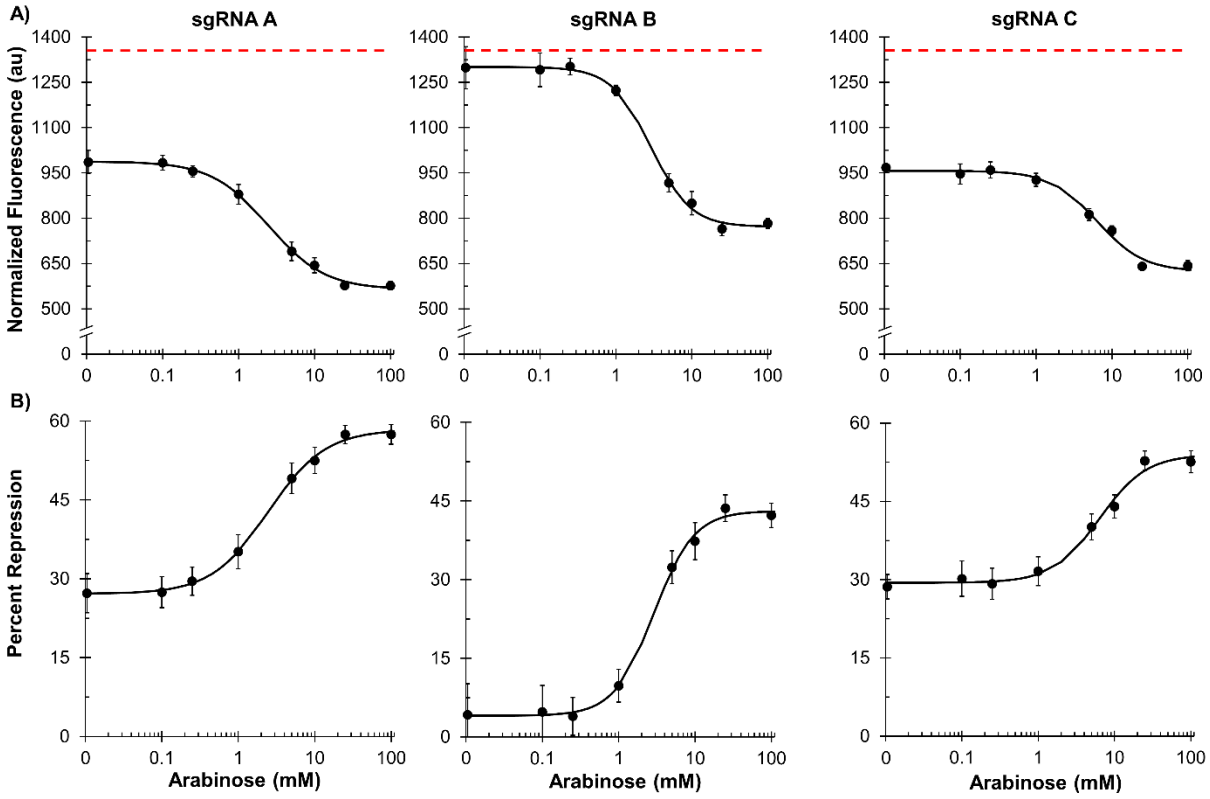


Figure 4.6. Tunable repression with CRISPRi in *R. opacus*. A codon optimized version of dCas9 from *Streptococcus thermophilus* ($dCas9_{Sth1}$) was placed under the control of the arabinose inducible pBAD promoter. The strong pConstitutive promoter drives sgRNA transcription. Three sgRNAs (A, B, and C) were designed to target a constitutively expressed EYFP gene integrated into ROCI-4. **A)** Normalized fluorescence (see 4.5 Materials and Methods) at early stationary phase in the presence of 0, 0.1. 0.25, 1, 5, 10, 25, and 100 mM arabinose. The red dashed line represents the positive control (EYFP only) value. **B)** The percent repression of each sgRNA (see 4.5 Materials and Methods). The solid black line in each case represents a fitted curve (see B.3 Supplementary Methods). Values represent the average of three replicates grown in minimal media

B with 4 g/L glucose, and error bars represent one standard deviation (A) or the propagated standard deviation (B; see B.3 Supplementary Methods).

4.4 Conclusion

R. opacus is a gram-positive Actinobacterium that has demonstrated promise as an industrially relevant strain for lignocellulosic biomass conversion due to its high aromatic tolerance and substantial number of catabolic genes for a broad range of feedstocks.[8, 11, 12, 39] Previously, this non-model strain had relatively few well-characterized genetic parts and tools available for engineering compared to even other Actinomycetales (e.g. *Mycobacterium* spp., *Streptomyces* spp., and *Corynebacterium* spp.), let alone model organisms like *E. coli* and *S. cerevisiae* (Table 4.1).[31, 47, 56, 75] This work has substantially expanded the molecular toolbox for gene expression by characterizing a constitutive promoter library, two antibiotic resistance markers, and the copy number of heterologous and endogenous plasmids. Furthermore, a curable plasmid backbone (pB264) was developed for temporary heterologous gene expression (e.g. recombinase expression). While engineering in *R. opacus* has been accomplished in the past, the actual expression level of heterologous enzymes has been unclear due to the use of uncharacterized parts.[22, 24-26] The characterization of promoters and plasmid vectors reported here, along with our recent work regarding chemical and metabolite sensors, provides a foundation for future metabolic engineering in *R. opacus*.[13] Furthermore, our characterization of the nine cryptic plasmids and observation of illegitimate integration events in *R. opacus* may support an interesting hypothesis regarding the role of large linear plasmids as hyper-recombinational gene storage devices in *Rhodococcus* and other Actinomycetales.[39]

Expanding beyond the development and characterization of fundamental genetic parts for heterologous gene expression, we also report methods for genome modification and gene

repression. For genome editing, we describe the development of a recombinase-based technique for site-specific genomic modification, in addition to a novel methodology for identification of neutral integration sites for stable heterologous gene expression. While site-specific recombinase-based techniques have been reported in other Actinomycetales, to our knowledge, this is the first use in *Rhodococcus*. Furthermore, our novel set of criteria for neutral site identification is applicable to other prokaryotes and will facilitate the rapid identification of non-deleterious candidate integration sites. Additionally, a CRISPRi platform was developed and demonstrated for tunable and targeted gene repression for the first time in *Rhodococcus*. CRISPRi has proved useful as a method for quickly and efficiently remodeling native metabolic pathways in prokaryotes.[77, 80] When combined with our tools for heterologous gene expression, CRISPRi will allow the lignocellulose consumption abilities of *R. opacus* to be paired with its extensive metabolic pathways for the production of valuable bioproducts.[8, 15, 26, 39] Notably, even though both of these systems were originally developed for the closely related *M. tuberculosis*, they still required significant optimization, reinforcing the challenges in engineering a non-model organism.[44, 75] The results of this, and other recent work, have contributed significantly towards making *R. opacus* a relevant strain for future industrial applications.

4.5 Materials and methods

4.5.1 Strains, plasmids and growth conditions

All plasmids were assembled in *E. coli* DH10B using either Gibson or Golden Gate assembly methods.[81-83] Kanamycin (20 μg/mL), gentamicin (10 μg/mL), chloramphenicol (34 μg/mL), or hygromycin B (200 μg/mL) was added as appropriate to *E. coli* cultures. Electrocompetent *E. coli* DH10B and *R. opacus* cells were transformed as previously described.[64, 83] Plasmid DNA was isolated using Zyppy Plasmid Miniprep Kit (Zymo), and PCR products were

extracted from electrophoresis gels using Zymoclean Gel DNA Recovery Kit (Zymo). All genetic part sequences, plasmids, and strains are listed in Supplementary Tables B.2, B.8, and B.9, respectively. Enzymes were purchased from New England Biolabs and Thermo Fischer Scientific.

The pConstitutive promoter sequence driving GFP+ underwent saturation mutagenesis at the -10 and -35 regions, either individually or concurrently, using degenerate oligonucleotides for PCR, followed by blunt end ligation to circularize the DNA.[84] Approximately 100 *R. opacus* colonies were screened in triplicate for fluorescence output. 25 promoters spanning ~45-fold difference in fluorescence output were selected (Figure 4.1), and the promoter sequences are shown in Supplementary Table B.1.

R. opacus was cultured in one of two defined minimal media (A or B; medium composition as previously described by DeLorenzo *et al*.)[13, 15] at 30 °C and 250 rpm using 50 mL glass culture tubes. Unless otherwise noted, 2 g/L glucose was used as the carbon source, and 1 g/L ammonium sulfate was used as the nitrogen source. For all cultures, a single colony was transferred from a tryptic soy broth (TSB) agar plate into 4.5 mL of minimal media and incubated for ~18 to 24 hrs, and then subcultured in 10 mL of media to create a larger volume of seed culture. Growth experiments were performed in triplicate at the 10 mL scale, with an initial optical density at 600 nm (OD_{600}) of ~0.2, unless otherwise noted. Kanamycin (50 μg/mL), gentamicin (10 μg/mL), chloramphenicol (15 μg/mL), and hygromycin B (50 μg/mL) were added as appropriate to *R. opacus* cultures. All chemicals were purchased from Sigma-Aldrich unless otherwise noted.

4.5.2 Growth and fluorescence measurements

The optical density (OD_{600}) was measured in VWR semi-micro polystyrene cuvettes at 600 nm using a Tecan Infinite 200 Pro plate reader. Cell fluorescence and absorbance at 600 nm (*Abs600*) were measured in black 96-well plates (Greiner Bio-One flat bottom, chimney well,

μclear) using a Tecan Infinite 200 Pro plate reader. An absorbance value at 600 nm can be converted into an optical density value (for *R. opacus* cultures) by using the experimentally determined relationship (OD_{600} = 1.975 × *Abs600*). The excitation and emission wavelengths for GFP+ were 488 and 530 nm, respectively.[13] The excitation and emission wavelengths for EYFP were 485 and 528 nm, respectively.[13] Fluorescence measurements were normalized using Equation 1, where $Fluo_{norm}$ is the normalized fluorescence, $Abs600_{sample}$ is the test strain absorbance at 600 nm, $Fluo_{sample}$ is the test strain fluorescence, $Fluo_{control}$ is the empty vector control strain fluorescence, and $Abs600_{control}$ is the empty vector control strain absorbance at 600 nm.[13]

Eq. 1)
$$Fluo_{norm} = \frac{Fluo_{sample}}{Abs600_{sample}} - \frac{Fluo_{control}}{Abs600_{control}}$$

4.5.3 Total DNA extraction and quantitative PCR

Total DNA (genomic DNA [gDNA] and plasmids) was extracted from 30 mL biological triplicates of *R. opacus* at an OD_{600} of ~1.0. Cultures were centrifuged at 3500 rcf and re-suspended in 1 mL of sterile deionized water. Cells were then frozen at -80 ^{o}C for one hour and then lysed at 100 ^{o}C for thirty minutes. The Zymo Quick-DNA Fungal/Bacterial Midiprep Kit was used for total DNA isolation. Final DNA concentrations were measured using a Nanodrop 2000 UV-Vis Spectrophotometer. All samples were diluted to 1 ng DNA/μL for further use.

qPCR was used to determine the absolute copy number ratio of the heterologous and endogenous plasmids relative to the chromosome, as previously described.[45] Primers were designed to amplify 140-224 base pairs of DNA from each target and are listed in Supplementary Table B.4. To determine the copy number ratio, an external standard dilution curve of cycle threshold (C_T) versus log copy number is required for each amplicon (Supplementary Figure B.1).

Amplified PCR product was used as an external standard and serially diluted from 10^9 to 10^5 copies of amplicon/µL. Equation 2 was used to initially dilute PCR product to 10^9 copies/µL.

Eq. 2) $$Amplicon\ Copy\ Number = \frac{6.02*10^{23}\left(\frac{copy}{mol}\right)*DNA\ amount\ (g)}{Amplicon\ length\ (bp)*660\left(\frac{g}{mol*bp}\right)}$$

qPCR was first performed in duplicate on the serially diluted PCR product using a Bio-Rad CFX96 real time thermocycler and SYBR Select Master Mix for CFX. 1 µL of mixed 5 mM primers and 1 µL of DNA were used for each reaction. Cycling conditions were 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s, 60 °C for 1 min, and fluorescence measurement. A threshold value of 3650 was set for all reactions. To confirm that no off-target amplification had occurred, a melting curve analysis was performed at the end of each PCR program and a single melting peak was verified for each amplicon. The efficiency of the PCR reaction was calculated using Equation 3 and a linear regression was performed to confirm linearity (Supplementary Table B.3).

Eq. 3) $$PCR\ Reaction\ Efficiency = 100 * (10^{\frac{-1}{standard\ curve\ slope}} - 1)$$

qPCR, as described above, was then performed on the extracted, biological triplicate DNA samples. 1 µL of mixed 5 mM primers and 1 µL of 1 ng DNA/µL were used for each reaction. The standard curves were used to determine the number of copies of each plasmid in the sample based on the measured sample C_T values. The copy number ratio was then determined by dividing the number of copies of each plasmid by the number of copies of the chromosome.

4.5.4 Identification of neutral sites

The criteria for identifying candidate neutral sites are displayed in Figure 4.4. The first step involved using previously collected transcriptomic data to find two sequential genes with zero or

very low levels of transcription.[8] The genetic architecture of the identified low expression gene pairs was analyzed using a previously published *R. opacus* genome (NCBI database Refseq, NZ_CP003949.1; referred to as LPD genome).[38] The desired orientation was for the two ORFs to be facing each other (sequential 5' to 3' and 3' to 5'). Furthermore, a gap of at least 200 bp between the respective stop codons was required. The LPD genome annotation of each gene was examined to avoid genes of critical function. All genes were further analyzed using NCBI blastn to confirm non-critical gene function (Supplementary Table B.4).

4.5.5 Gene integration

PCR was used to amplify a ~500 bp region up- and downstream of the selected integration site in the genome. These two pieces were assembled into a plasmid containing an *E. coli* origin of replication (p15a), a constitutively expressed EYFP, and a hygromycin B selection cassette (Supplementary Figure B.2A). This plasmid was then linearized by a PCR reaction such that the p15a origin was removed (Supplementary Figure B.2B). Approximately 1 μg of linearized DNA was transformed via electroporation, as previously described, into *R. opacus* competent cells harboring the constitutively expressed recombinases Che9c60 and Che9c61 (Supplementary Figure B.3).[13, 44] The electroporated cells were allowed to recover in TSB at 30 ºC for 6 hours, compared to the 3 hours normally allowed for transformations. Cells were then plated on a TSB agar plate containing hygromycin B. The vector containing the recombinase genes was cured after removal of selection pressure and a round of colony purification (Supplementary Figure B.4). Successful recombination at each integration site was confirmed by colony PCR (Supplementary Figure B.5).

4.5.6 CRISPRi

The codon optimized dCas9 from *S. thermophilus* ($dCas9_{Sth1}$) was placed under the control of the pBAD promoter. sgRNAs were designed based on rules described by Rock *et al.*[75] The tested sgRNAs were driven by the original strong constitutive promoter (pConstitutive). Both the pBAD-$Cas9_{Sth}$ and pConstitutive-sgRNA cassettes were placed onto pAL5000 (S). The three CRISPRi constructs (sgRNAs A, B, and C) were transformed into the strain containing BbaJ23104-EYFP integrated into ROCI-4. To increase the number of cell doublings, the culture media were modified to contain 4 g/L glucose and the initial inoculum density was reduced to an OD_{600} of 0.1. CRISPRi strains were induced with 0-100 mM arabinose. Fluorescence was measured at early stationary phase. Percent repression was calculated using Equation 4, where μ_I is the normalized fluorescence measured from the induced culture and μ_U is the normalized fluorescence measured from the positive control (EYFP only) culture.

Eq. 4) $$Percent\ Repression = 100 * (1 - \frac{\mu_I}{\mu_U})$$

4.6 Supporting information

Supplementary Tables B.1-10, Supplementary Figures B.1-7, and Supplementary Methods can be found in Appendix B.

4.7 References

1. Russell, J. J., Theriot, J. A., Sood, P., Marshall, W. F., Landweber, L. F., Fritz-Laylin, L., Polka, J. K., Oliferenko, S., Gerbich, T., Gladfelter, A., Umen, J., Bezanilla, M., Lancaster, M. A., He, S., Gibson, M. C., Goldstein, B., Tanaka, E. M., Hu, C.-K., and Brunet, A. (2017) Non-model model organisms, BMC Biol. 15.

2. Löbs, A.-K., Schwartz, C., and Wheeldon, I. (2017) Genome and metabolic engineering in non-conventional yeasts: Current advances and applications, Synthetic and Systems Biotechnology 2, 198-207.

3. Yan, Q., and Fong, S. S. (2017) Challenges and Advances for Genetic Engineering of Non-model Bacteria and Uses in Consolidated Bioprocessing, Frontiers in Microbiology 8, 2060.

4. Seppala, S., Wilken, S. E., Knop, D., Solomon, K. V., and O'Malley, M. A. (2017) The importance of sourcing enzymes from non-conventional fungi for metabolic engineering and biomass breakdown, Metab. Eng. 44, 45-59.

5. Balan, V., Chiaramonti, D., and Kumar, S. (2013) Review of US and EU initiatives toward development, demonstration, and commercialization of lignocellulosic biofuels, Biofuels, Bioproducts and Biorefining 7, 732-759.

6. Pu, Y., Zhang, D., Singh, P. M., and Ragauskas, A. J. (2008) The new forestry biofuels sector, Biofuels, Bioproducts and Biorefining 2, 58-73.

7. Beckham, G. T., Johnson, C. W., Karp, E. M., Salvachua, D., and Vardon, D. R. (2016) Opportunities and challenges in biological lignin valorization, Curr. Opin. Biotechnol. 42, 40-53.

8. Yoneda, A., Henson, W. R., Goldner, N. K., Park, K. J., Forsberg, K. J., Kim, S. J., Pesesky, M. W., Foston, M., Dantas, G., and Moon, T. S. (2016) Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630, Nucleic Acids Res 44, 2240-2254.

9. Lin, S. Y., and Dence, C. W. (1992) Methods in Lignin Chemistry, Springer.

10. Alvarez, H. M., Mayer, F., Fabritius, D., and Steinbuchel, A. (1996) Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630, Archives of microbiology 165, 377-386.

11. Kurosawa, K., Laser, J., and Sinskey, A. J. (2015) Tolerance and adaptive evolution of triacylglycerol-producing Rhodococcus opacus to lignocellulose-derived inhibitors, Biotechnology for Biofuels 8.

12. Holder, J. W., Ulrich, J. C., DeBono, A. C., Godfrey, P. A., Desjardins, C. A., Zucker, J., Zeng, Q., Leach, A. L. B., Ghiviriga, I., Dancel, C., Abeel, T., Gevers, D., Kodira, C. D., Desany, B., Affourtit, J. P., Birren, B. W., and Sinskey, A. J. (2011) Comparative and Functional Genomics of Rhodococcus opacus PD630 for Biofuels Development, PLoS Genet 7, e1002219.

13. DeLorenzo, D. M., Henson, W. R., and Moon, T. S. (2017) Development of Chemical and Metabolite Sensors for Rhodococcus opacus PD630, ACS synthetic biology 6, 1973-1978.

14. Ling, H., Teo, W., Chen, B., Leong, S. S., and Chang, M. W. (2014) Microbial tolerance engineering toward biochemical production: from lignocellulose to products, Curr. Opin. Biotechnol. 29, 99-106.

15. Hollinshead, W. D., Henson, W. R., Abernathy, M., Moon, T. S., and Tang, Y. J. (2016) Rapid metabolic analysis of Rhodococcus opacus PD630 via parallel 13C-metabolite fingerprinting, Biotechnol. Bioeng. 113, 91-100.

16. Lennen, R. M., and Pfleger, B. F. (2013) Microbial production of fatty acid-derived fuels and chemicals, Curr. Opin. Biotechnol. 24, 1044-1053.

17. Kruyer, N. S., and Peralta-Yahya, P. (2017) Metabolic engineering strategies to bio-adipic acid production, Curr. Opin. Biotechnol. 45, 136-143.

18. Sun, X., Lin, Y., Yuan, Q., and Yan, Y. (2014) Biological production of muconic acid via a prokaryotic 2,3-dihydroxybenzoic acid decarboxylase, ChemSusChem 7, 2478-2481.

19. Vardon, D. R., Franden, M. A., Johnson, C. W., Karp, E. M., Guarnieri, M. T., Linger, J. G., Salm, M. J., Strathmann, T. J., and Beckham, G. T. (2015) Adipic acid production from lignin, Energy & Environmental Science 8, 617-628.

20. Sheng, J., and Feng, X. (2015) Metabolic engineering of yeast to produce fatty acid-derived biofuels: bottlenecks and solutions, Front Microbiol 6, 554.

21. Wheeldon, I., Christopher, P., and Blanch, H. (2017) Integration of heterogeneous and biochemical catalysis for production of fuels and chemicals from biomass, Curr. Opin. Biotechnol. 45, 127-135.

22. Kurosawa, K., Wewetzer, S. J., and Sinskey, A. J. (2013) Engineering xylose metabolism in triacylglycerol-producing Rhodococcus opacus for lignocellulosic fuel production, Biotechnology for biofuels 6, 134.

23. Hetzler, S., and Steinbüchel, A. (2013) Establishment of Cellobiose Utilization for Lipid Production in Rhodococcus opacus PD630, Appl. Environ. Microbiol. 79, 3122-3125.

24. Kurosawa, K., Plassmeier, J., Kalinowski, J., Ruckert, C., and Sinskey, A. J. (2015) Engineering L-arabinose metabolism in triacylglycerol-producing Rhodococcus opacus for lignocellulosic fuel production, Metab. Eng. 30, 89-95.

25. Hetzler, S., and Steinbuchel, A. (2013) Establishment of cellobiose utilization for lipid production in Rhodococcus opacus PD630, Applied and environmental microbiology 79, 3122-3125.

26. Lanfranconi, M. P., and Alvarez, H. M. (2017) Rewiring neutral lipids production for the de novo synthesis of wax esters in Rhodococcus opacus PD630, J Biotechnol 260, 67-73.

27. Cress, B. F., Trantas, E. A., Ververidis, F., Linhardt, R. J., and Koffas, M. A. (2015) Sensitive cells: enabling tools for static and dynamic control of microbial metabolic pathways, Curr. Opin. Biotechnol. 36, 205-214.

28. Gilman, J., and Love, J. (2016) Synthetic promoter design for new microbial chassis, Biochem Soc Trans 44, 731-737.

29. Hetzler, S., Broker, D., and Steinbuchel, A. (2013) Saccharification of Cellulose by Recombinant Rhodococcus opacus PD630 Strains, Applied and environmental microbiology 79, 5159-5166.

30. Coussement, P., Bauwens, D., Maertens, J., and De Mey, M. (2017) Direct Combinatorial Pathway Optimization, ACS synthetic biology 6, 224-232.

31. Siegl, T., Tokovenko, B., Myronovskyi, M., and Luzhetskyy, A. (2013) Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes, Metabolic engineering 19, 98-106.

32. Ehrt, S., Guo, X. V., Hickey, C. M., Ryou, M., Monteleone, M., Riley, L. W., and Schnappinger, D. (2005) Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor, Nucleic acids research 33, e21.

33. Scholz, O., Thiel, A., Hillen, W., and Niederweis, M. (2000) Quantitative analysis of gene expression with an improved green fluorescent protein, European Journal of Biochemistry 267, 1565-1570.

34. Luo, Y., Zhang, L., Barton, K. W., and Zhao, H. (2015) Systematic Identification of a Panel of Strong Constitutive Promoters from Streptomyces albus, ACS synthetic biology 4, 1001-1010.

35. Li, S., Wang, J., Li, X., Yin, S., Wang, W., and Yang, K. (2015) Genome-wide identification and evaluation of constitutive promoters in streptomycetes, Microb Cell Fact 14, 172.

36. Seghezzi, N., Amar, P., Koebmann, B., Jensen, P. R., and Virolle, M. J. (2011) The construction of a library of synthetic promoters revealed some specific features of strong Streptomyces promoters, Appl. Microbiol. Biotechnol. 90, 615-623.

37. Kittell, B. L., and Helinski, D. R. (1993) Plasmid Incompatibility and Replication Control, In Bacterial Conjugation, pp 223-242, Springer, Boston, MA.

38. Chen, Y., Ding, Y., Yang, L., Yu, J., Liu, G., Wang, X., Zhang, S., Yu, D., Song, L., Zhang, H., Zhang, C., Huo, L., Huo, C., Wang, Y., Du, Y., Zhang, H., Zhang, P., Na, H., Xu, S., Zhu, Y., Xie, Z., He, T., Zhang, Y., Wang, G., Fan, Z., Yang, F., Liu, H., Wang, X., Zhang, X., Zhang, M. Q., Li, Y., Steinbuchel, A., Fujimoto, T., Cichello, S., Yu, J., and Liu, P. (2014) Integrated omics study delineates the dynamics of lipid droplets in Rhodococcus opacus PD630, Nucleic acids research 42, 1052-1064.

39. Larkin, M. J., Kulakov, L. A., and Allen, C. C. (2005) Biodegradation and Rhodococcus--masters of catabolic versatility, Curr. Opin. Biotechnol. 16, 282-290.

40. Sekine, M., Tanikawa, S., Omata, S., Saito, M., Fujisawa, T., Tsukatani, N., Tajima, T., Sekigawa, T., Kosugi, H., Matsuo, Y., Nishiko, R., Imamura, K., Ito, M., Narita, H., Tago, S., Fujita, N., and Harayama, S. (2006) Sequence analysis of three plasmids harboured in Rhodococcus erythropolis strain PR4, Environ Microbiol 8, 334-346.

41. Kalkus, J., Menne, R., Reh, M., and Schlegel, H. G. (1998) The terminal structures of linear plasmids from Rhodococcus opacus, Microbiology 144 (Pt 5), 1271-1279.

42. Ahsan, S., and Kabir, M. S. (2013) Linear plasmids and their replication, Stamford Journal of Microbiology 2, 1-5.

43. Desomer, J., Crespi, M., and Van Montagu, M. (1991) Illegitimate integration of non-replicative vectors in the genome of Rhodococcus fascians upon electrotransformation as an insertional mutagenesis system, Mol Microbiol 5, 2115-2124.

44. van Kessel, J. C., and Hatfull, G. F. (2007) Recombineering in Mycobacterium tuberculosis, Nat. Methods 4, 147-152.

45. Lee, C., Kim, J., Shin, S. G., and Hwang, S. (2006) Absolute and relative QPCR quantification of plasmid copy number in Escherichia coli, J Biotechnol 123, 273-280.

46. Bourn, W. R., Jansen, Y., Stutz, H., Warren, R. M., Williamson, A. L., and van Helden, P. D. (2007) Creation and characterisation of a high-copy-number version of the pAL5000 mycobacterial replicon, Tuberculosis (Edinb) 87, 481-488.

47. Kirchner, O., and Tauch, A. (2003) Tools for genetic engineering in the amino acid-producing bacterium Corynebacterium glutamicum, J Biotechnol 104, 287-299.

48. Pátek, M., and Nešvera, J. (2013) Promoters and Plasmid Vectors of Corynebacterium glutamicum, In Corynebacterium glutamicum (Yukawa, H., and Inui, M., Eds.), pp 51-88, Springer Berlin Heidelberg.

49. Lessard, P. A., O'Brien, X. M., Currie, D. H., and Sinskey, A. J. (2004) pB264, a small, mobilizable, temperature sensitive plasmid from Rhodococcus, BMC microbiology 4, 15.

50. Capecchi, M. R. (1989) Altering the genome by homologous recombination, Science 244, 1288-1292.

51. Immethun, C. M., Ng, K. M., DeLorenzo, D. M., Waldron-Feinstein, B., Lee, Y. C., and Moon, T. S. (2015) Oxygen-responsive genetic circuits constructed in synechocystis sp. PCC 6803, Biotechnol. Bioeng. 113, 433-442.

52. Wan, N., DeLorenzo, D. M., He, L., You, L., Immethun, C. M., Wang, G., Baidoo, E. E., Hollinshead, W., Keasling, J. D., Moon, T. S., and Tang, Y. J. (2017) Cyanobacterial carbon metabolism: Fluxome plasticity and oxygen dependence, Biotechnol. Bioeng. 114, 1593-1602.

53. Immethun, C. M., DeLorenzo, D. M., Focht, C. M., Gupta, D., Johnson, C. B., and Moon, T. S. (2017) Physical, chemical, and metabolic state sensors expand the synthetic biology toolbox for Synechocystis sp. PCC 6803, Biotechnol. Bioeng. 114, 1561-1569.

54. Tyo, K. E., Jin, Y. S., Espinoza, F. A., and Stephanopoulos, G. (2009) Identification of gene disruptions for increased poly-3-hydroxybutyrate accumulation in Synechocystis PCC 6803, Biotechnol. Prog. 25, 1236-1243.

55. Zhang, L., Ou, X., Zhao, G., and Ding, X. (2008) Highly efficient in vitro site-specific recombination system based on streptomyces phage phiBT1 integrase, J Bacteriol 190, 6392-6397.

56. Phelan, R. M., Sachs, D., Petkiewicz, S. J., Barajas, J. F., Blake-Hedges, J. M., Thompson, M. G., Reider Apel, A., Rasor, B. J., Katz, L., and Keasling, J. D. (2017) Development of Next Generation Synthetic Biology Tools for Use in Streptomyces venezuelae, ACS synthetic biology 6, 159-166.

57. Fogg, P. C., Colloms, S., Rosser, S., Stark, M., and Smith, M. C. (2014) New applications for phage integrases, J Mol Biol 426, 2703-2716.

58. Muyrers, J. P., Zhang, Y., Buchholz, F., and Stewart, A. F. (2000) RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners, Genes Dev 14, 1971-1982.

59. Schwartz, C., Shabbir-Hussain, M., Frogue, K., Blenner, M., and Wheeldon, I. (2017) Standardized Markerless Gene Integration for Pathway Engineering in Yarrowia lipolytica, ACS synthetic biology 6, 402-409.

60. Ng, A. H., Berla, B. M., and Pakrasi, H. B. (2015) Fine-Tuning of Photoautotrophic Protein Production by Combining Promoters and Neutral Sites in the Cyanobacterium Synechocystis sp. Strain PCC 6803, Appl Environ Microbiol 81, 6857-6863.

61. Pinto, F., Pacheco, C. C., Oliveira, P., Montagud, A., Landels, A., Couto, N., Wright, P. C., Urchueguia, J. F., and Tamagnini, P. (2015) Improving a Synechocystis-based photoautotrophic chassis through systematic genome mapping and validation of neutral sites, DNA Res. 22, 425-437.

62. Johnsson, P., Morris, K. V., and Grander, D. (2014) Pseudogenes: a novel source of trans-acting antisense RNAs, Methods Mol Biol 1167, 213-226.

63. Bird, C. P., Stranger, B. E., and Dermitzakis, E. T. (2006) Functional variation and evolution of non-coding DNA, Curr Opin Genet Dev 16, 559-564.

64. Kalscheuer, R., Arenskotter, M., and Steinbuchel, A. (1999) Establishment of a gene transfer system for Rhodococcus opacus PD630 based on electroporation and its application for recombinant biosynthesis of poly(3-hydroxyalkanoic acids), Applied microbiology and biotechnology 52, 508-515.

65. Dower, W. J., Miller, J. F., and Ragsdale, C. W. (1988) High efficiency transformation of E. coli by high voltage electroporation, Nucleic Acids Res 16, 6127-6145.

66. Bryant, J. A., Sellars, L. E., Busby, S. J., and Lee, D. J. (2014) Chromosome position effects on gene expression in Escherichia coli K-12, Nucleic Acids Res 42, 11383-11392.

67. Bassalo, M. C., Garst, A. D., Halweg-Edwards, A. L., Grau, W. C., Domaille, D. W., Mutalik, V. K., Arkin, A. P., and Gill, R. T. (2016) Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli, ACS Synth Biol 5, 561-568.

68. Sauer, C., Syvertsson, S., Bohorquez, L. C., Cruz, R., Harwood, C. R., van Rij, T., and Hamoen, L. W. (2016) Effect of Genome Position on Heterologous Gene Expression in Bacillus subtilis: An Unbiased Analysis, ACS synthetic biology 5, 942-947.

69. Davis, J. H., Rubin, A. J., and Sauer, R. T. (2011) Design, construction and characterization of a set of insulated bacterial promoters, Nucleic acids research 39, 1131-1141.

70. Zong, Y., Zhang, H. M., Lyu, C., Ji, X., Hou, J., Guo, X., Ouyang, Q., and Lou, C. (2017) Insulated transcriptional elements enable precise design of genetic circuits, Nat Commun 8, 52.

71. Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression, Nature Protocols 8, 2180-2196.

72. Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression, Cell 152, 1173-1183.

73. Alessandra, S., Alessandro, T., Flavio, S., and Alejandro, H. (2008) Artificial antisense RNAs silence lacZ in E. coli by decreasing target mRNA concentration, BMB reports 41, 568-574.

74. Hoynes-O'Connor, A., and Moon, T. S. (2016) Development of Design Rules for Reliable Antisense RNA Behavior in E. coli, ACS Synthetic Biology 5, 1441-1454.

75. Rock, J. M., Hopkins, F. F., Chavez, A., Diallo, M., Chase, M. R., Gerrick, E. R., Pritchard, J. R., Church, G. M., Rubin, E. J., Sassetti, C. M., Schnappinger, D., and Fortune, S. M. (2017) Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform, Nature microbiology 2, 16274.

76. Choudhary, E., Thakur, P., Pareek, M., and Agarwal, N. (2015) Gene silencing by CRISPR interference in mycobacteria, Nature communications 6, 6267.

77. Cleto, S., Jensen, J. V. K., Wendisch, V. F., and Lu, T. K. (2016) Corynebacterium glutamicum Metabolic Engineering with CRISPR Interference (CRISPRi), ACS Synthetic Biology 5, 375-385.

78. Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression, Nature protocols 8, 2180-2196.

79. Deaner, M., Mejia, J., and Alper, H. S. (2017) Enabling Graded and Large-Scale Multiplex of Desired Genes Using a Dual-Mode dCas9 Activator in Saccharomyces cerevisiae, ACS synthetic biology 6, 1931-1943.

80. Kim, S. K., Seong, W., Han, G. H., Lee, D. H., and Lee, S. G. (2017) CRISPR interference-guided multiplex repression of endogenous competing pathway genes for redirecting metabolic flux in Escherichia coli, Microb Cell Fact 16, 188.

81. Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability, PloS one 3, e3647.

82. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases, Nature methods 6, 343-345.

83. Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., 3rd, Burland, V., Mau, B., Petrosino, J. F., Qin, X., Muzny, D. M., Ayele, M., Gibbs, R. A., Csorgo, B., Posfai, G., Weinstock, G. M., and Blattner, F. R. (2008) The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse, Journal of bacteriology 190, 2597-2606.

84. Georgescu, R., Bandara, G., and Sun, L. (2003) Saturation Mutagenesis, In Directed Evolution Library Creation (Arnold, F., and Georgiou, G., Eds.), pp 75-83, Humana Press.

85. Ranes, M. G., Rauzier, J., Lagranderie, M., Gheorghiu, M., and Gicquel, B. (1990) Functional analysis of pAL5000, a plasmid from Mycobacterium fortuitum: construction of a "mini" mycobacterium-Escherichia coli shuttle vector, J Bacteriol 172, 2793-2797.

86. Zhang, Y., Praszkier, J., Hodgson, A., and Pittard, A. J. (1994) Molecular analysis and characterization of a broad-host-range plasmid, pEP2, Journal of bacteriology 176, 5718-5728.

87. Nesvera, J., Patek, M., Hochmannova, J., Abrhamova, Z., Becvarova, V., Jelinkova, M., and Vohradsky, J. (1997) Plasmid pGA1 from Corynebacterium glutamicum codes for a gene product that positively influences plasmid copy number, Journal of bacteriology 179, 1525-1532.

88. Archer, J. A., and Sinskey, A. J. (1993) The DNA sequence and minimal replicon of the Corynebacterium glutamicum plasmid pSR1: evidence of a common ancestry with plasmids from C. diphtheriae, J Gen Microbiol 139, 1753-1759.

89. Xiong, X., Wang, X., and Chen, S. (2012) Engineering of a xylose metabolic pathway in Rhodococcus strains, Applied and environmental microbiology 78, 5483-5491.

90. Mahenthiralingam, E., Draper, P., Davis, E. O., and Colston, M. J. (1993) Cloning and sequencing of the gene which encodes the highly inducible acetamidase of Mycobacterium smegmatis, J Gen Microbiol 139, 575-583.

91. Timm, J., Lim, E. M., and Gicquel, B. (1994) Escherichia coli-mycobacteria shuttle vectors for operon and gene fusions to lacZ: the pJEM series, Journal of bacteriology 176, 6749-6753.

92. Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T., and Jacobs, W. R., Jr. (1990) Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis, Mol Microbiol 4, 1911-1919.

93. Chiba, K., Hoshino, Y., Ishino, K., Kogure, T., Mikami, Y., Uehara, Y., and Ishikawa, J. (2007) Construction of a pair of practical Nocardia-Escherichia coli shuttle vectors, Jpn J Infect Dis 60, 45-47.

94. Tauch, A., Kirchner, O., Loffler, B., Gotker, S., Puhler, A., and Kalinowski, J. (2002) Efficient electrotransformation of corynebacterium diphtheriae with a mini-replicon derived from the Corynebacterium glutamicum plasmid pGA1, Current microbiology 45, 362-367.

95. Dong, L., Nakashima, N., Tamura, N., and Tamura, T. (2004) Isolation and characterization of the Rhodococcus opacus thiostrepton-inducible genes tipAL and tipAS: application for recombinant protein expression in Rhodococcus, FEMS microbiology letters 237, 35-40.

96. Vesely, M., Patek, M., Nesvera, J., Cejkova, A., Masak, J., and Jirku, V. (2003) Host-vector system for phenol-degrading Rhodococcus erythropolis based on Corynebacterium plasmids, Applied microbiology and biotechnology 61, 523-527.

97. Hanisch, J., Waltermann, M., Robenek, H., and Steinbuchel, A. (2006) The Ralstonia eutropha H16 phasin PhaP1 is targeted to intracellular triacylglycerol inclusions in Rhodococcus opacus PD630 and Mycobacterium smegmatis mc2155, and provides an anchor to target other proteins, Microbiology 152, 3271-3280.

Chapter 5: Selection of stable reference genes for RT-qPCR in *Rhodococcus opacus* PD630

Reprinted with permission from DeLorenzo, D.M. and Moon, T.S. Selection of stable reference genes for RT-qPCR in *Rhodococcus opacus* PD630. *Scientific Reports* (2018). 8, 6019.

The genetic parts developed as part of this dissertation enable the ability to controllably express or repress either heterologous or endogenous genes in *R. opacus*. The expression level of most genes, however, can be difficult to estimate if there is no easily quantifiable metric (e.g., fluorescence from a reporter protein). In this chapter, a methodology for directly quantifying the mRNA concentration of any gene—reverse transcription quantitative polymerase chain reaction (RT-qPCR)—will be established in *R. opacus* through the identification of stable reference genes for data normalization. I performed all experiments and wrote the manuscript.

5.1 Abstract

Rhodococcus opacus PD630 is a gram-positive bacterium with promising attributes for the conversion of lignin into valuable fuels and chemicals. To develop an organism as a cellular factory, it is necessary to have a deep understanding of its metabolism and any heterologous pathways being expressed. For the purpose of quantifying gene transcription, reverse transcription quantitative PCR (RT-qPCR) is the gold standard due to its sensitivity and reproducibility. However, RT-qPCR requires the use of reference genes whose expression is stable across distinct growth or treatment conditions to normalize the results. Unfortunately, no in-depth analysis of stable reference genes has been conducted in *Rhodococcus*, inhibiting the utilization of RT-qPCR in *R. opacus*. In this work, ten candidate reference genes, chosen based on previously collected

RNA sequencing data or literature, were examined under four distinct growth conditions using three mathematical programs (BestKeeper, Normfinder, and geNorm). Based on this analysis, the minimum number of reference genes required was found to be two, and two separate pairs of references genes were identified as optimal normalization factors for when ribosomal RNA is either present or depleted. This work represents the first validation of reference genes for *Rhodococcus*, providing a valuable starting point for future research.

5.2 Introduction

Rhodococcus opacus PD630 (hereafter *R. opacus*) is a gram-positive, oleaginous bacterium that possesses beneficial traits for the conversion of lignin into valuable fuels and chemicals[1-4]. Some of these features include a number of catabolic pathways for lignin-derived aromatic compound consumption and a high native tolerance towards these and other inhibitory lignocellulosic biomass breakdown products[2-5]. Furthermore, *R. opacus* can consume multiple types of carbon sources simultaneously[6], thereby facilitating higher rates of feedstock conversion. Additionally, this organism can direct a large fraction of its cellular resources to the production of biofuel precursors (up to ~78% triacylglycerol [TAG] of cell dry weight)[1]. *R. opacus* has been previously engineered to facilitate lignocellulose conversion[3,4,7,8] and a substantial genetic toolbox has recently been developed[9,10]. However, a deep understanding of this organism's metabolism and any heterologous pathways being expressed is required to maximize its potential.

A number of technologies exist for examining an organism's gene expression (i.e. the transcriptome), which is the first step to a systems level understanding. One such technology is the microarray, which allows for gene expression profiling[11]. RNA sequencing (RNA-Seq) is a newer technology that has become the default method for examining the entire transcriptome of an organism[12]. However, it can add additional costs if only several genes are of interest, is limited

when mRNA concentrations are low (although this is changing with the advent of single cell sequencing[13]), and generally still requires corroboration via additional quantitative methods[12]. One such complimentary method is reverse transcription quantitative PCR (RT-qPCR), which is considered the gold standard of mRNA quantification due to its high sensitivity, reproducibility, speed, ability to examine numerous samples simultaneously, and large dynamic range[14,15]. Both microarrays and RT-qPCR require the use of an internal standard, optimally a gene that is stably expressed across the tested growth or treatment conditions, to normalize expression data between samples and conditions[16].

Unfortunately, no in-depth analysis of stable reference genes (RGs) has been performed in *Rhodococcus*, limiting the ability to quantitatively analyze gene expression. One qPCR study in *Rhodococcus equi* even stated that no reference gene was included in their experiment and that such an inclusion could have improved their work[17]. We could find only two examples of reference genes previously reported in *Rhodococcus*. The first reference gene was a gene encoding sigma factor A (*sigA*) in *Rhodococcus* sp. RHA1[18], although no justification for this choice was provided. The second reference gene was a gene encoding DNA Polymerase IV, which was used in *Rhodococcus* sp. RHA1, *Rhodococcus jostii*, and *Rhodococcus erythropolis*, and this choice was justified based on a microarray experiment performed in *Rhodococcus* sp. RHA1[19,20]. Both of these reference genes were used in isolation and their characterization was incomplete, which fails to satisfy the current minimum information guidelines for publication of quantitative PCR experiments (i.e. MIQE guidelines stating that the minimum number of reference genes needs to be quantitatively determined and that one gene is not generally sufficient for normalization)[21,22].

In this work, we identified ten candidate reference genes (RGs) and examined the stability of their expression in *R. opacus* across four distinct growth conditions using three mathematical

models (BestKeeper[23], NormFinder[24], and geNorm[16,25]). Additionally, the minimum number of required reference genes was identified. Two different sets of genes were identified as optimal normalization factors (NFs) depending on whether ribosomal RNA (rRNA) is either present or depleted. This work facilitates the utilization of RT-qPCR in *R. opacus*, in addition to providing a valuable preliminary set of reference genes for further future validation in other *Rhodococcus* spp.

5.3 Results and discussion

5.3.1 Choice of candidate reference genes

Two methods were utilized for the selection of candidate reference genes (RGs). The primary approach used our previously published transcriptomic data collected from *R. opacus* grown in a minimal salts medium with either glucose or phenol to identify stably expressed genes[3]. We selected nine genes as candidates (RG1 to RG9; Table 5.1) whose DeSeq2 normalized transcript level did not vary significantly between the two growth conditions, whose DeSeq2 expression value was greater than 750, and whose coding region is at least 350 bp in length[3,26]. The secondary approach utilized a literature review which found that *sigA* had been previously used as a RG in *Rhodococcus* sp. RHA1[18] and that a DNA Polymerase IV gene has been previously used in *Rhodococcus jostii*, *Rhodococcus erythropolis*, and *Rhodococcus* sp. RHA1 as a RG[19,20]. *sigA* was ruled out due to no justification for its selection as a RG being provided[18]. As PD630_RS27310 is annotated as a DNA Polymerase IV gene in *R. opacus*, it was selected as the tenth candidate RG (RG10; Table 5.1).

Table 5.1. List of candidate reference genes (RG). The amplicon size, PCR efficiency percentage, minimum and maximum threshold cycle (C_T) values observed across all tested growth conditions, the standard deviation of C_T values across all tested growth conditions (determined by BestKeeper), and the C_T value of the no template control (NTC) are listed for each RG. See Supplementary Figure D.1 for standard curves used to calculate PCR efficiency. See Supplementary Figure D.2 for melting curve analysis. See Supplementary Table D.1 for oligonucleotide sequences and melting and annealing temperatures.

Reference gene (RG)	Gene number	Gene annotation	Amplicon size (bp)	PCR efficiency	Min C_T	Max C_T	C_T Std Dev	NTC C_T
RG1	PD630_ RS22865	Pup-protein ligase	157	94%	23.44	25.23	0.40	N.d.
RG2	PD630_ RS20570	Hypothetical protein	87	94%	25.86	28.20	0.67	N.d.
RG3	PD630_ RS03840	23S rRNA	87	95%	8.74	10.16	0.33	N.d.
RG4	PD630_ RS15810	Polyribonucleotide nucleotidyltransferase	115	101%	22.93	27.41	1.37	N.d.
RG5	PD630_ RS25785	L,D-transpeptidase Mb0493	90	101%	24.56	25.96	0.37	N.d.
RG6	PD630_ RS13910	NAD(P)H dehydrogenase	82	100%	25.47	29.01	1.13	N.d.
RG7	PD630_ RS25530	ATP-dependent Clp protease ATP-binding subunit ClpX	94	92%	24.93	25.77	0.21	N.d.
RG8	PD630_ RS01395	16S rRNA	101	92%	10.48	11.74	0.29	N.d.
RG9	PD630_ RS37755	rRNA small subunit methyltransferase G	102	92%	28.32	30.33	0.46	N.d.
RG10	PD630_ RS27310	DNA polymerase IV	71	94%	28.34	30.61	0.61	N.d.

N.d. = no amplification detected

5.3.2 RT-qPCR primer characterization and data collection

RT-qPCR primers were designed based on previously published suggestions[22], in addition to the specific criteria discussed in the Supplementary Materials. The size and specificity of each amplicon was confirmed using agarose gel electrophoresis and melting curve analysis (Figure 5.1 and Supplementary Figure D.2), and the no template controls (NTCs) demonstrated non-detectable levels of amplification (Table 5.1 and Supplementary Figure D.2). To calculate the PCR amplification efficiency, a ten-fold serial dilution of template DNA (PCR amplified product) was performed and followed by qPCR (Supplementary Figure D.1). A linear regression analysis was performed on the resultant C_T values to confirm the linearity of each serial dilution and to calculate the PCR amplification efficiency (Table 5.1 and Supplementary Figure D.1). As all serial dilutions had an acceptable R^2 of at least 0.97 and PCR efficiencies ranged from 92 to 101%, all primer sets were deemed suitable for qPCR (Table 5.1 and Supplementary Figure D.2).

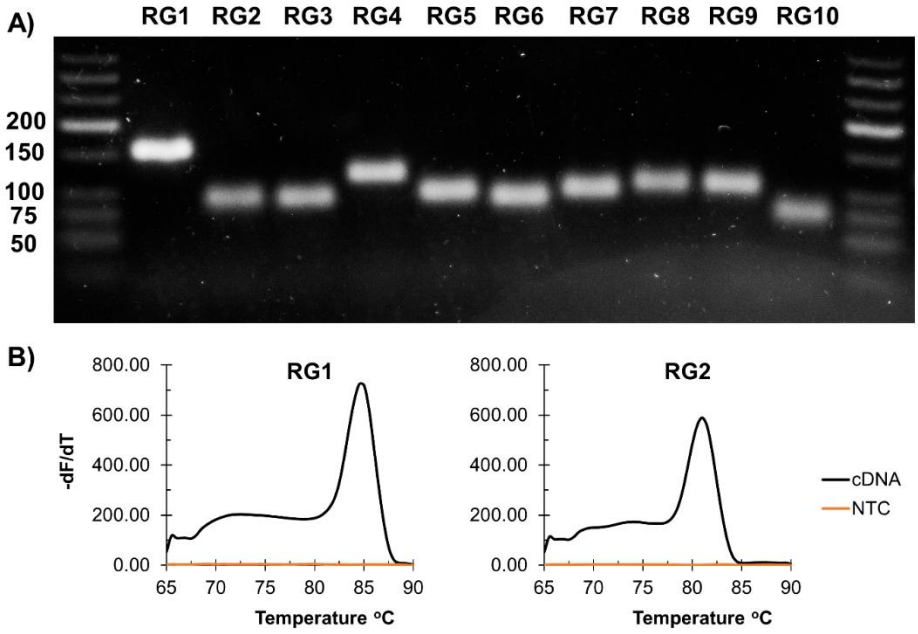


Figure 5.1. Confirmation of specificity of primers used in RT-qPCR analysis. **A)** Contrast enhanced image of electrophoresis gel confirming amplicon size and primer specificity using RT-

qPCR amplification product. The sizes for nucleotide ladder are indicated to left of bands (50 to 200 bp). The size of each amplicon is denoted in Table 5.1. The original non-enhanced gel image, in addition to the corresponding gel image of no template controls (NTCs), can be found in Supplementary Figure D.5. **B)** Representative melt curves post RT-qPCR for RG1 and RG2 as well as the corresponding NTCs (see Supplementary Figure D.2 for all melt curves).

R. opacus was cultured in four distinct growth conditions, and RNA was extracted from biological triplicate cultures for candidate RG expression stability analysis. The growth conditions are described in full in the Materials and Methods, but in summary they consist of a minimal salts medium with glucose and either high nitrogen (HN) or low nitrogen (LN), a rich tryptic soy broth medium (TSB), or a minimal salts medium with phenol and high nitrogen (PHE). Each medium was selected based on growth conditions that *R. opacus* is likely to experience during general research endeavors and based on the diverse predicted changes in metabolic topology required for catabolism of each respective feedstock. Glucose was chosen as a representative sugar feedstock as it is frequently provided in *R. opacus* cell cultures[3,9,27]. High and low nitrogen concentrations were examined because *R. opacus* undergoes a metabolic flux shift to produce large quantities of industrially relevant lipids during nitrogen deprivation[1,28,29]. TSB was selected as it contains an array of amino acids, which are utilized by a diverse set of metabolic pathways[30]. Phenol was selected as a representative aromatic compound that may be found in depolymerized lignin[3,4] and because it requires a different subset of metabolic pathways than glucose and amino acids[6].

RT-qPCR was then performed on the cDNA generated from the triplicate RNA samples collected from the HN, LN, TSB, and PHE growth conditions (Figure 5.2). The raw C_T values ranged from 9.2 to 30.2 for RG1 through RG10 across all growth conditions. The minimum and maximum C_T values, in addition to the standard deviation, observed for each candidate RG across all growth conditions are listed in Table 5.1. A visual appraisal of Figure 5.2 can provide initial

insight into candidate RG expression stability, as some RGs demonstrated a tight clustering of C_T values (e.g. RG1, RG3, RG5, RG7, RG8, and RG9) while others demonstrated a spread in C_T values (e.g. RG2, RG4, RG6, and RG10) across growth conditions.

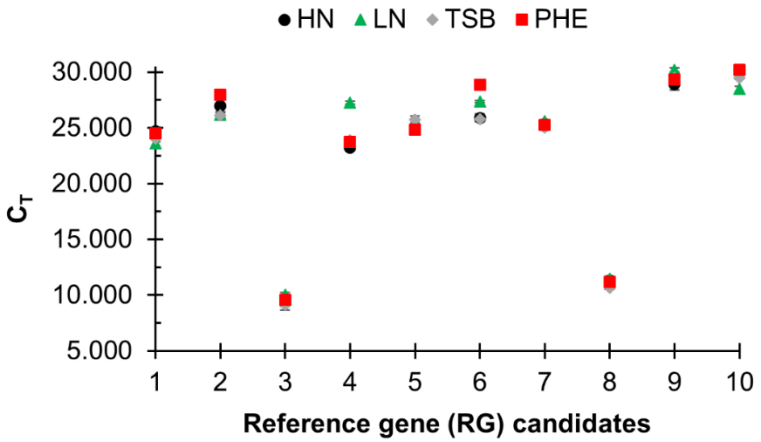


Figure 5.2. CT values for ten candidate reference genes in *R. opacus*. RNA was extracted from biological triplicates of *R. opacus* grown in four distinct growth conditions (HN [circle], LN [triangle], TSB [diamond], and PHE [square]; see 5.5 Materials and Methods for full description). RT-qPCR was 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.

5.3.3 Expression stability of candidate reference genes across distinct growth conditions

Candidate RG expression stability was quantitatively examined using three different statistical programs: BestKeeper[23], NormFinder[24], and geNorm[16,25]. Each of these programs generates a gene expression stability coefficient (r-value for BestKeeper, stability value for Normfinder, and M value for geNorm) and rank order for the candidate RGs (see 5.5 Materials and Methods for full description; Figure 5.3). Additionally, candidate RGs were ranked based on

their C_T standard deviation calculated by BestKeeper (Supplementary Table D.2). A C_T standard deviation greater than 1 is considered unstable[23].

All analyses identified the same three candidate RGs, albeit in different rank orders, as being the most stably expressed across all four growth conditions. According to BestKeeper, the top three most stable RGs from 1st to 3rd, based on the smallest standard deviation of C_T values, were RG7, RG8, and RG3 (Supplementary Table D.2). The top three RGs with the lowest BestKeeper r-value from 1st to 3rd were RG3, RG7, and RG8 (Figure 5.3A1). Normfinder determined that the top three most stable RGs from 1st to 3rd, based on its stability value calculation, were RG7, RG3, and RG8 (Figure 5.3B1). Finally, geNorm determined that the top three most stable RGs from 1st to 3rd, based on its M value, were RG7, RG8, and RG3 (Figure 5.3C1).

Two of the top three most stable candidate RGs encode rRNAs (RG3 and RG8), which is consistent with previous works in other organisms that identified an rRNA gene as a stable RG[31,32]. However, rRNA is often depleted from mRNA samples being prepared for RNA-Seq[33]. If RT-qPCR is to be used to corroborate RNA-Seq results, other non-rRNA RGs need to be identified. All analyses were repeated on the non-rRNA candidate RGs, and two genes (RG7 and RG9) appeared in the top 3 rankings of three analyses, while another (RG5) appeared in the top 3 rankings of two analyses (Figure 5.3). The BestKeeper r-value identified the top three non-rRNA RGs from 1st to 3rd as RG6, RG7, and RG9 (Figure 5.3A2). As RG6 had a C_T standard deviation greater than 1 (Supplementary Table D.2), it was ruled out as a candidate. Normfinder identified the top three non-rRNA RGs from 1st to 3rd as RG7, RG5, and RG9 (Figure 5.3B2). geNorm identified the top three non-rRNA RGs from 1st to 3rd as RG9, RG5, and RG7 (Figure 5.3C2).

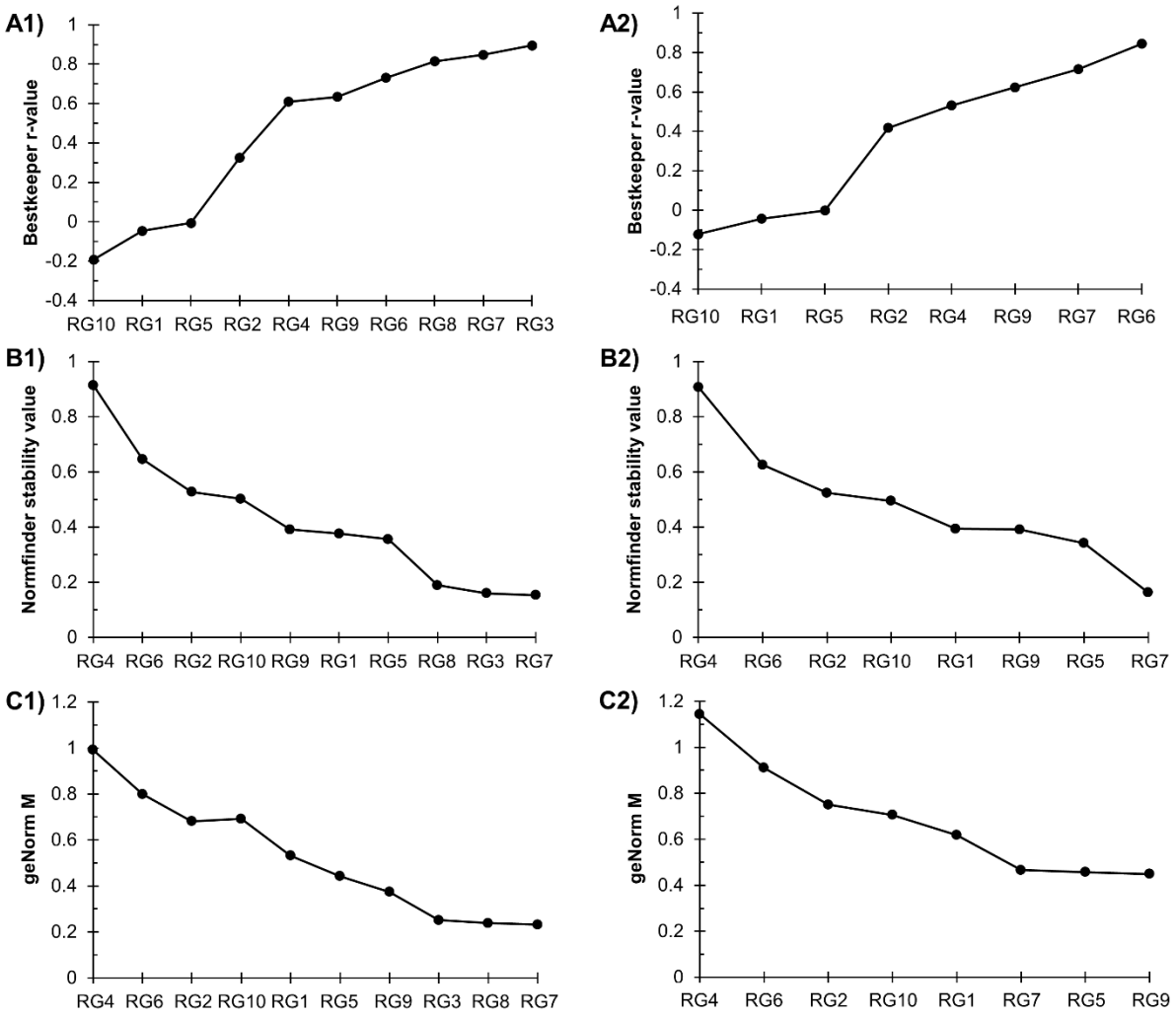


Figure 5.3. Rankings of candidate reference genes. Genes were ranked from the least stable (on the left) to the most stable (on the right). Analysis was performed after pooling C_T data across all four growth conditions. The designation of 1 means that the analyses were performed on all ten candidate RGs. The designation of 2 means that the analyses were performed on the eight non-rRNA candidate RGs. **A)** Genes were ranked according to their BestKeeper r-value. BestKeeper r-value significance can be found in Supplementary Tables D.3 and D.4. **B)** Genes were ranked according to their Normfinder stability value. **C)** Genes were ranked according to their geNorm M value.

5.3.4 Minimum required number of reference genes

Analyses of candidate RG expression stability identified several consistently transcribed genes across the four distinct growth conditions, but it was still unknown how many RGs were required for optimal normalization of expression data. The geNorm software is also capable of producing a V value (see Materials and Methods), which produces a quantitative suggestion for the ideal number of required RGs. By calculating the pairwise variation of V_n/V_{n+1} (V value), where n is the number of RGs, the benefit of using n versus n+1 RGs for normalization can be quantified. A V value below 0.15 signifies that there is no additional benefit from adding another RG to the normalization factor (NF)[16,25]. A V value was calculated for all ten candidate RGs and the non-rRNA RG subset, and both had a V_2/V_3 value below 0.15 (0.087 and 0.145, respectively), meaning that there is no benefit of using three RGs over two RGs in the NF (Figure 5.4 and Supplementary Figure D.3).

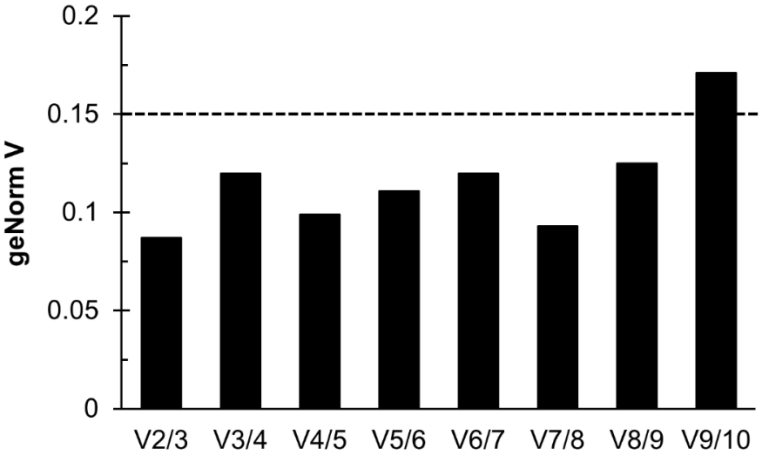


Figure 5.4. Minimum number of reference genes. The pair-wise variation Vn/Vn+1, where n represents the number of RGs used in the normalization factor (NF), was calculated by geNorm to determine the minimum number of RGs required for normalization. A geNorm V value below 0.15 signifies that no additional benefit is gained from increasing the number of reference genes from n to n+1. The dashed line represents a V value of 0.15.

5.3.5 Validation of the selected reference genes

To examine the effect of utilizing different combinations of RGs in the NF, a normalized expression analysis was performed using expression data from PD630_RS15810 (RG4) grown under the four distinct growth conditions (Figure 5.5). A box plot visualizing the non-normalized expression data for PD630_RS15810 was generated to show the general trends in expression level changes across the four growth conditions (Figure 5.5A). The normalized relative fold changes in expression going from the HN condition to either the LN, TSB, or PHE condition were calculated using REST 2009 with one of six NFs, including RG6, RG10, RG3/RG7, RG3/RG8, RG7/RG8, and RG3/RG7/RG8 (Figure 5.5). Confidence intervals of 95% were calculated by REST 2009 and used for comparison between different NFs. RG6 was chosen as an example of a poor RG candidate, while RG10 was chosen as DNA Polymerase IV had been used as a RG in other *Rhodococcus* spp.[19,20]. All two-component combinations (NF_2) of RG3, RG7, and RG8 were examined as it was unclear whether there were differences between them, in addition to a NF comprising all three to confirm that there was no significant difference between a NF_2 and a NF_3.

The results of the normalized expression analysis revealed that there was no significant difference in the normalized relative fold change ratio going from the HN condition to either the LN, TSB, or PHE condition when using a NF_2 comprising any two selections of RG3, RG7, and RG8 or a NF_3 comprising all three genes. As the NF_2 comprising RG7 and RG8 (RG7/RG8) had the smallest 95% confidence interval, it was chosen as the optimal pair of RGs for future use in *R. opacus* under the tested growth conditions. RG7/RG8 normalization revealed that PD630_RS15810 expression with 95% confidence was downregulated between 0.054 to 0.088-fold when going from HN to LN, 0.355 to 0.575-fold when going from HN to TSB, and 0.494 to 0.916-fold when going from HN to PHE (Figure 5.5).

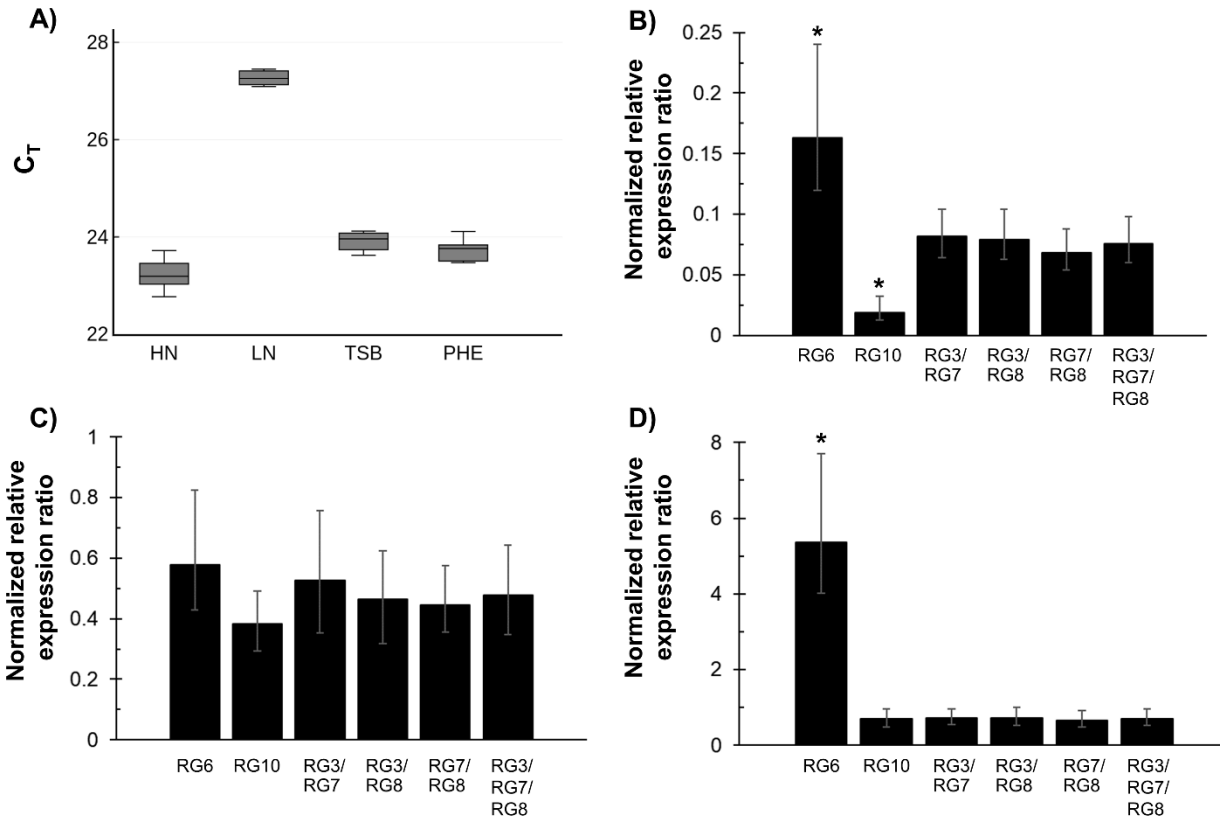


Figure 5.5. Effect of reference gene choice on RT-qPCR normalization. **A)** Box plots of averaged PD630_RS15810 expression data (C_T) for all four growth conditions (HN, LN, TSB, and PHE). Each gray box represents the first through third quartiles, the solid black line represents the median, and the whiskers represent the minimum and maximum values. **B-D)** The normalized relative expression ratio of PD630_LPD05540 going from HN to either LN (B), TSB (C), or PHE (D). The expression data was normalized with a NF, including RG6, RG10, RG3/RG7, RG3/RG8, RG7/RG8, and RG3/RG7/RG8 using REST 2009. Error bars represent the 95% confidence interval (CI) as calculated by REST 2009. Stars indicate that a 95% CI range falls outside of the 95% CI range of the RG3/RG7/RG8 (and RG7/RG8) normalized ratio.

The effect of non-optimal RGs was also examined in comparison to the NF_3 comprising RG3, RG7 and RG8. When PD630_RS15810 expression was normalized with RG6, there was a significantly different 95% confidence range ($p < 0.05$) of 0.120 to 0.240-fold downregulation when going from HN to LN, a comparable 0.430 to 0.824-fold downregulation going from HN to TSB, and a significantly different ($p < 0.05$) 4.009 to 7.703-fold upregulation when going from

HN to PHE (Figure 5.5). When PD630_RS15810 expression was normalized with the literature-based selection RG10, there was a significantly different 95% confidence range ($p < 0.05$) of 0.013 to 0.032-fold downregulation when going from HN to LN, a comparable 0.294 to 0.490-fold downregulation going from HN to TSB, and a comparable 0.480 to 0.971-fold downregulation when going from HN to PHE. These results demonstrate that a poorly selected RG can substantially alter the observed change in gene expression (e.g. RG6), and that RGs should be examined in all relevant growth conditions prior to use as they may be stable in some instances but not in others (e.g. RG10).

To determine an optimal NF for a scenario where rRNA has been depleted, a normalized expression analysis was performed again on PD630_RS15810 using the top three non-rRNA candidate RGs (RG5, RG7, and RG9). The normalized expression ratio of all NF_2 combinations of these three genes, in addition the NF_3 comprising all of them, was examined in comparison to the best NF_2 comprising RG7 and RG8 (Supplementary Figure D.4). The NFs including RG5/ RG7, RG5/RG9, and RG5/RG7/RG9 all produced 95% confidence normalized expression fold change ranges that were significantly different ($p < 0.05$), compared to the expression fold change range produced by RG7/RG8 when going from HN to TSB (Supplementary Figure D.4C). The NF comprising RG7 and RG9 was comparable to RG7/RG8 in all scenarios, with 95% confidence normalized expression ranges of PD630_RS15810 similarly changing 0.072 to 0.127-fold when going from HN to LN, 0.534 to 0.909-fold when going from HN to TSB, and 0.550 to 1.096-fold when going from HN to PHE.

5.4 Conclusion

This study represents the first in-depth analysis of stable reference genes in *Rhodococcus* and identified two comparable normalization factors in *R. opacus* for samples with or without

rRNA. The expression of ten candidate reference genes was examined across four distinct growth conditions (HN, LN, TSB, and PHE) using three mathematical models (BestKeeper, Normfinder, and geNorm) to identify and rank gene expression stability. Based on the geNorm V value and corroboration via a normalized gene expression analysis using REST 2009, only two reference genes are required for an optimal normalization factor in *R. opacus*. For samples containing rRNA, the two best reference genes were RG7 (PD630_RS25530; ATP-dependent Clp protease ATP-binding subunit ClpX) and RG8 (PD630_RS01395; 16S rRNA). For samples depleted in rRNA, the best two reference genes were RG7 and RG9 (PD630_RS37755; rRNA small subunit methyltransferase G). The diversity of growth conditions tested in this work bestows confidence in our selections of reference genes for *R. opacus*, in addition to providing a valuable starting point in the choice of reference genes when studying other *Rhodococcus* spp.

5.5 Materials and methods

5.5.1 Strain and culture conditions

Rhodococcus opacus PD630 (DSMZ 44193) was grown in either a minimal salts medium as previously described (minimal media recipe B)[9], with one of three distinct combinations of carbon and nitrogen sources (discussed in detail in the Gene expression studies section), or a rich tryptic soy broth (TSB) medium. Cultures were incubated at 30 ^{o}C and 250 rotations per minute (rpm). For the growth experiment, all cultures were grown in triplicate 10 mL cultures with an initial optical density at 600 nm (OD_{600}) of 0.2. Chemicals were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise noted. MIQE guidelines were applied as appropriate to the design and execution of experiments[21,22].

5.5.2 Candidate reference gene selection and primer design

The selection of candidate reference genes was based on either previously published transcriptomic data for *R. opacus*[3] or literature for other *Rhodococcus* spp.[19,20]. RG1 through RG9 were selected as potentially stable genes based on analysis of RNA-Seq data performed on *R. opacus* grown in a minimal salts medium with either glucose or phenol[3]. RG10 was chosen based on its gene annotation as DNA Polymerase IV, as other *Rhodococcus* spp. studies have used a DNA Polymerase IV gene as a reference gene[19,20]. RT-qPCR primers (Integrated DNA Technologies) were designed based on literature suggestions[21,22] and guidelines discussed in the Supplementary Materials.

5.5.3 Primer amplification efficiency and specificity

To calculate primer amplification efficiency, a serial dilution of template DNA and a linear regression analysis of the resultant qPCR results were required. PCR product was obtained by using 0.5 μL of GoTaq G2 polymerase (Promega), 10 μL of GoTaq buffer, 1.5 μL of a mixture containing the forward and reverse primers (5 μM each), 0.5 μL of genomic DNA (gDNA), 1 μL of 10 mM dNTPs (GBiosciences), and 36.5 μL of H_2O; and thermocycler conditions of 95 °C for 2 min, followed by 35 cycles of 95 °C for 15 s, 62 °C for 30 s, and 72 °C for 30 s. PCR products were gel extracted using the Zymoclean DNA Gel Recovery kit (Zymo Research) and prepared for qPCR using the DNA Clean and Concentrator kit (Zymo Research). Five rounds of a 10-fold serial dilution were performed on the purified PCR product. qPCR was performed in duplicate on the serially diluted PCR product using a Bio-Rad CFX96 real time thermocycler, TempPlate 96-well semi-skirt 0.1 mL PCR plates (USA Scientific), and Power SYBR Green PCR Master Mix (Applied Biosystems). 10 μL of Power SYBR Green PCR Master Mix, 0.5 μL of a mixture containing the forward and reverse primers (5 μM each), 1 μL of PCR product, and 8.5 μL of H_2O

were used for each reaction. Cycling conditions were 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s, 62 °C for 20 s, 72 °C for 20 s, and a fluorescence measurement. A fluorescence threshold value of 750 was set for all reactions. To confirm that no off-target amplification had occurred, a melting curve analysis was performed at the end of each qPCR program and a single melting peak was verified for each amplicon. A linear regression was performed on the threshold cycle (C_T) values observed from each serial dilution to confirm linearity (R^2) and to calculate the equation of the line, which was used to estimate the efficiency of the PCR (Equation 1; Supplementary Figure D.1).

Eq. 1) $$\boldsymbol{PCR\ Efficiency\ Percentage = 100 * ([10^{\frac{-1}{standard\ curve\ slope}}] - 1)}$$

5.5.4 Gene expression studies

To examine gene expression stability, *R. opacus* was grown in four distinct growth conditions. The first condition was a rich TSB medium. The other three conditions were minimal salts media with carbon and nitrogen sources as follows: glucose with high nitrogen (HN; 2 g/L glucose and 1 g/L ammonium sulfate), glucose with low nitrogen (LN; 2 g/L glucose and 0.05 g/L ammonium sulfate), and phenol with high nitrogen (PHE; 0.75 g/L phenol + 1 g/L ammonium sulfate). A 5 mL seed culture in a 50 mL glass tube containing the minimal salts medium with 2 g/L glucose and 1 g/L ammonium sulfate was used for the HN, LN, and TSB growth conditions, while the PHE seed culture also had 0.3 g/L phenol added to acclimate the cells to the inhibitory aromatic. To remove remaining glucose and ammonium sulfate from the seed cultures, samples were centrifuged at 3500 relative centrifugal force (rcf), washed with minimal media containing no carbon or nitrogen sources, centrifuged again at 3500 rcf, and re-suspended in their final growth condition media. When the cultures reached mid-exponential phase, the cells were centrifuged at

3500 rcf and re-suspended in DNA/RNA Shield (Zymo Research), per the instructions. Samples were stored at -80 ^{o}C until needed for further use.

5.5.5 RNA extraction and cDNA synthesis

RNA was extracted from the triplicate biological samples stored at -80 ^{o}C in the DNA/RNA Shield using the Quick-RNA MiniPrep Plus kit (Zymo Research), per the instructions. The optional ZR BashingBead Lysis Tubes (0.1 and 0.5 mm beads; Zymo Research) were used to lyse the cells. All samples were treated with the TURBO DNA-free kit (Ambion) to remove any gDNA present in the samples and then purified using the RNA Clean and Concentrator kit (Zymo Research). To confirm that all gDNA was depleted, PCR was performed by using the previously described GoTaq protocol with primers targeting the genome. Gel electrophoresis was performed to confirm that no PCR amplification bands existed. Any samples with persisting gDNA were re-treated with the TURBO DNA-free kit, purified again, and examined via PCR and gel electrophoresis. RNA concentration and purity were quantified using a NanoDrop 2000c (all samples had a 260 nm/280 nm absorbance ratio of 2 to 2.1). 1 ug of RNA per sample was converted to cDNA using the AffinityScript QPCR cDNA synthesis kit (Agilent Technologies).

5.5.6 RT-qPCR

RT-qPCR was performed in technical triplicate on the biological triplicate RNA extracts using a Bio-Rad CFX96 and Power SYBR Green PCR Master Mix (Applied Biosystems). 10 μL of Power SYBR Green PCR Master Mix, 0.5 μL of a mixture containing the forward and reverse primers (5 μM each), 1 μL of cDNA, and 8.5 μL of H_2O were used for each reaction. Cycling conditions were 95 °C for 10 min, followed by 35 cycles of 95 °C for 15 s, 62 °C for 20 s, 72 ^{o}C for 20 s, and a fluorescence measurement. A threshold value of 750 was set for all reactions. To confirm that no off-target amplification had occurred, a melting curve analysis was performed at

the end of each qPCR program and a single melting peak was verified for each amplicon (Supplementary Figure D.2). Additionally, a negative control (no template control; NTC) was included for each primer pair (Table 5.1 and Supplementary Figure D.2). Gel electrophoresis was also performed post RT-qPCR to confirm the existence of a single amplicon (Figure 5.1A and Supplementary Figure D.5). Gel images were captured using a DigiDoc-It imaging system and its accompanying Doc-ItLS software. The original Figure 5.1A gel image was cropped, and contrast and exposure settings for the whole image were modified using Adobe Lightroom to improve image interpretation. The original image can be found as Supplementary Figure D.5.

5.5.7 Reference gene expression stability analysis

The expression level stability of the ten candidate RGs across the four distinct growth conditions was assessed and ranked using three commonly used software tools: BestKeeper[23], NormFinder version 0.953[24], and geNorm (built into qBase+ [Biogazelle])[16,25] (Figures 5.3 and 5.4; Supplementary Tables D.2 and D.3). Additionally, all analyses were performed again using just candidate RGs 1, 2, 4-7, 9, and 10 to accommodate a situation where rRNA (RG3 and 8) has been depleted from the sample (Figure 5.3, Supplementary Figures D.3 and D.4, and Supplementary Table D.4).

BestKeeper is a Microsoft Excel based tool that calculates the standard deviation and coefficient of variance of the C_T value for each candidate RG, and creates an index based on the geometric mean of the C_T values, which it uses to facilitate pair-wise comparisons against and between RGs. A Pearson correlation coefficient (r-value) is calculated for each RG, with higher r-values representing greater stability (max value of 1)[23]. The technical triplicate C_T values for each biological replicate were averaged prior to entry into BestKeeper and all four growth conditions were pooled and analyzed together. Candidate RGs were ranked on both their C_T value standard

deviation (values > 1.0 deemed unstable[23]; Supplementary Table D.2) and their r-value (Figures 5.3A1 and 5.3A2, and Supplementary Tables D.3 and D.4).

Normfinder is also a Microsoft Excel based tool that utilizes a model-based approach to compare intra- and inter-group variation between candidate RGs and then generates a stability value and ranking for each candidate gene[24]. The lower the stability value, the more stable the gene's expression. The technical triplicate C_T values for each biological replicate were averaged and samples from all four growth conditions were pooled. Values were then converted to a log-scale per Equation 2, where E is the amplicon specific PCR efficiency, prior to entry into Normfinder[24]. Candidate RGs were ranked on their stability value (Figures 5.3B1 and 5.3B2).

Eq. 2) $$\boldsymbol{Log\ transformed\ value = E^{-C_{T\ avg}}}$$

geNorm, formerly a Microsoft Excel based tool that has been incorporated into the qBase+ software package (Biogazelle), examines the pairwise variation between RGs and creates a stability M value, where lower values represent more stably expressed genes[16,25]. The program iterates its calculations wherein it eliminates the gene with the highest M value and then recalculates the stability of the remaining genes. Thus, geNorm creates a ranking of RG stability (Figures 5.3C1 and 5.3C2). A geNorm M value lower than 1 is generally considered stable for heterogeneous growth conditions[22]. geNorm can also determine the minimum number of RGs required for optimal RT-qPCR normalization. Various normalization factors (NF) are calculated by first taking the geometric mean of the two most stable RGs' C_T values and then generating additional factors by adding the next best RG until all RGs are used. A pairwise variation (V_n/V_{n+1}) V value is then calculated by comparing the benefit of going from NF_n to NF_{n+1}, where n is a number of RGs (Figure 5.4 and Supplementary Figure D.3). Once the V value is less than 0.15, there is no additional benefit of going from n to n+1[16,25].

5.5.8 Validation of reference gene selection

The effect of different combinations of the top candidate RGs was examined by creating multiple different NFs and normalizing the same set of expression data using REST 2009 (Qiagen). The relative expression ratio data of PD630_RS15810 going from the HN condition to either the LN, TSB, or PHE condition was normalized by a NF_2 comprising any two of the top three RGs identified across all analyses (RG3, RG7, and RG8), in addition to a NF_3 comprising all three. Additionally, RG6 was used as an example of an unsuitable RG, while RG10 was used based on literature references for other *Rhodococcus* spp[19,20]. 95% confidence intervals were generated by REST 2009. If 95% confidence intervals failed to overlap, they were labelled as significantly different ($p < 0.05$). As two of the top three RGs identified encode rRNAs (RG3 and RG8), which are frequently depleted when performing RNA-Seq, additional NFs comprising non-rRNA RGs (RG5, RG7, and RG9) were examined in comparison to the overall top NF combination (RG7 and RG8; Supplementary Figure D.4).

5.6 Supplementary materials

The Supplementary Materials can be found in Appendix D and consist of Supplementary Table D.1 through D.4, Supplementary Figure D.1 through D.5, and RT-qPCR primer design criteria.

5.7 References

1. Alvarez, H. M.; Mayer, F.; Fabritius, D.; Steinbuchel, A., Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630. Arch Microbiol 1996, 165 (6), 377-86.

2. Holder, J. W.; Ulrich, J. C.; DeBono, A. C.; Godfrey, P. A.; Desjardins, C. A.; Zucker, J.; Zeng, Q.; Leach, A. L. B.; Ghiviriga, I.; Dancel, C.; Abeel, T.; Gevers, D.; Kodira, C. D.; Desany, B.; Affourtit, J. P.; Birren, B. W.; Sinskey, A. J., Comparative and Functional Genomics of Rhodococcus opacus PD630 for Biofuels Development. PLoS Genet 2011, 7, e1002219.

3. Yoneda, A.; Henson, W. R.; Goldner, N. K.; Park, K. J.; Forsberg, K. J.; Kim, S. J.; Pesesky, M. W.; Foston, M.; Dantas, G.; Moon, T. S., Comparative transcriptomics elucidates adaptive

phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630. Nucleic Acids Res 2016, 44 (5), 2240-54.

4. Kurosawa, K.; Laser, J.; Sinskey, A. J., Tolerance and adaptive evolution of triacylglycerol-producing Rhodococcus opacus to lignocellulose-derived inhibitors. Biotechnology for Biofuels 2015, 8 (76).

5. Lin, S. Y.; Dence, C. W., Methods in Lignin Chemistry. Springer: 1992.

6. Hollinshead, W. D.; Henson, W. R.; Abernathy, M.; Moon, T. S.; Tang, Y. J., Rapid metabolic analysis of Rhodococcus opacus PD630 via parallel 13C-metabolite fingerprinting. Biotechnol. Bioeng. 2016, 113 (1), 91-100.

7. Kurosawa, K.; Plassmeier, J.; Kalinowski, J.; Ruckert, C.; Sinskey, A. J., Engineering L-arabinose metabolism in triacylglycerol-producing Rhodococcus opacus for lignocellulosic fuel production. Metab. Eng. 2015, 30, 89-95.

8. Kurosawa, K.; Wewetzer, S. J.; Sinskey, A. J., Engineering xylose metabolism in triacylglycerol-producing Rhodococcus opacus for lignocellulosic fuel production. Biotechnology for Biofuels 2013, 6 (1), 134.

9. DeLorenzo, D. M.; Henson, W. R.; Moon, T. S., Development of Chemical and Metabolite Sensors for Rhodococcus opacus PD630. ACS synthetic biology 2017, 6 (10), 1973-1978.

10. DeLorenzo, D. M.; Rottinghaus, A. G.; Henson, W. R.; Moon, T. S., Molecular toolkit for gene expression control and genome modification in Rhodococcus opacus PD630. ACS synthetic biology 2018, Just Accepted.

11. Schena, M.; Shalon, D.; Davis, R. W.; Brown, P. O., Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995, 270 (5235), 467-70.

12. Wang, Z.; Gerstein, M.; Snyder, M., RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009, 10 (1), 57-63.

13. Zhu, S.; Qing, T.; Zheng, Y.; Jin, L.; Shi, L., Advances in single-cell RNA sequencing and its applications in cancer research. Oncotarget 2017, 8 (32), 53763-53779.

14. Bustin, S. A., Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J Mol Endocrinol 2002, 29 (1), 23-39.

15. Heid, C. A.; Stevens, J.; Livak, K. J.; Williams, P. M., Real time quantitative PCR. Genome Res. 1996, 6 (10), 986-94.

16. Vandesompele, J.; De Preter, K.; Pattyn, F.; Poppe, B.; Van Roy, N.; De Paepe, A.; Speleman, F., Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002, 3 (7), research0034.1-11.

17. Madrigal, R. G.; Shaw, S. D.; Witkowski, L. A.; Sisson, B. E.; Blodgett, G. P.; Chaffin, M. K.; Cohen, N. D., Use of Serial Quantitative PCR of the vapA Gene of Rhodococcus equi in Feces for Early Detection of R. equi Pneumonia in Foals. J Vet Intern Med 2016, 30 (2), 664-70.

18. Mohn, W. W.; Wilbrink, M. H.; Casabon, I.; Stewart, G. R.; Liu, J.; van der Geize, R.; Eltis, L. D., Gene cluster encoding cholate catabolism in Rhodococcus spp. J Bacteriol 2012, 194 (24), 6712-9.

19. Goncalves, E. R.; Hara, H.; Miyazawa, D.; Davies, J. E.; Eltis, L. D.; Mohn, W. W., Transcriptomic assessment of isozymes in the biphenyl pathway of Rhodococcus sp. strain RHA1. Appl Environ Microbiol 2006, 72 (9), 6183-93.

20. Szokol, J.; Rucka, L.; Simcikova, M.; Halada, P.; Nesvera, J.; Patek, M., Induction and carbon catabolite repression of phenol degradation genes in Rhodococcus erythropolis and Rhodococcus jostii. Appl. Microbiol. Biotechnol. 2014, 98 (19), 8267-79.

21. Bustin, S. A.; Benes, V.; Garson, J. A.; Hellemans, J.; Huggett, J.; Kubista, M.; Mueller, R.; Nolan, T.; Pfaffl, M. W.; Shipley, G. L.; Vandesompele, J.; Wittwer, C. T., The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. 2009, 55 (4), 611-22.

22. Taylor, S.; Wakem, M.; Dijkman, G.; Alsarraj, M.; Nguyen, M., A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines. Methods 2010, 50 (4), S1-5.

23. Pfaffl, M. W.; Tichopad, A.; Prgomet, C.; Neuvians, T. P., Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. Biotechnol. Lett. 2004, 26 (6), 509-15.

24. Andersen, C. L.; Jensen, J. L.; Orntoft, T. F., Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 2004, 64 (15), 5245-50.

25. Hellemans, J.; Mortier, G.; De Paepe, A.; Speleman, F.; Vandesompele, J., qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol 2007, 8 (2), R19.

26. Love, M. I.; Huber, W.; Anders, S., Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014, 15 (12), 550.

27. Kurosawa, K.; Boccazzi, P.; de Almeida, N. M.; Sinskey, A. J., High-cell-density batch fermentation of Rhodococcus opacus PD630 using a high glucose concentration for triacylglycerol production. J. Biotechnol. 2010, 147, 212-218.

28. Chen, Y.; Ding, Y.; Yang, L.; Yu, J.; Liu, G.; Wang, X.; Zhang, S.; Yu, D.; Song, L.; Zhang, H.; Zhang, C.; Huo, L.; Huo, C.; Wang, Y.; Du, Y.; Zhang, H.; Zhang, P.; Na, H.; Xu, S.; Zhu, Y.; Xie, Z.; He, T.; Zhang, Y.; Wang, G.; Fan, Z.; Yang, F.; Liu, H.; Wang, X.; Zhang, X.; Zhang, M. Q.; Li, Y.; Steinbuchel, A.; Fujimoto, T.; Cichello, S.; Yu, J.; Liu, P., Integrated omics study delineates the dynamics of lipid droplets in Rhodococcus opacus PD630. Nucleic Acids Res 2014, 42 (2), 1052-64.

29. Goswami, L.; Tejas Namboodiri, M. M.; Vinoth Kumar, R.; Pakshirajan, K.; Pugazhenthi, G., Biodiesel production potential of oleaginous Rhodococcus opacus grown on biomass gasification wastewater. Renewable Energy 2017, 105, 400-406.

30. Bender, D. A., Amino Acid Metabolism. 3rd ed.; Wiley-Blackwell: 2012; p 478.

31. Edwards, K. J.; Saunders, N. A., Real-time PCR used to measure stress-induced changes in the expression of the genes of the alginate pathway of Pseudomonas aeruginosa. J Appl Microbiol 2001, 91 (1), 29-37.

32. Pinto, F.; Pacheco, C. C.; Oliveira, P.; Montagud, A.; Landels, A.; Couto, N.; Wright, P. C.; Urchueguia, J. F.; Tamagnini, P., Improving a Synechocystis-based photoautotrophic chassis

through systematic genome mapping and validation of neutral sites. DNA Res. 2015, 22 (6), 425-37.

33. O'Neil, D.; Glowatz, H.; Schlumpberger, M., Ribosomal RNA depletion for efficient use of RNA-seq capacity. Curr Protoc Mol Biol 2013, Chapter 4, Unit 4 19.

Chapter 6: Elucidation of the role of the β-ketoadipate pathway and transporter genes in aromatic catabolism in *Rhodococcus opacus* PD630.

This work is a modified subset reprinted with permission from Henson, W.R.*, Campbell, T.*, DeLorenzo, D.M.*, Gao, Y., Berla, B., Kim, S.J., Foston, M., Moon, T.S., Dantas, G. Multi-omic elucidation of aromatic catabolism in adaptively evolved *Rhodococcus opacus*. Metabolic Engineering (2018). 49, 69-83. * = co-first authorship

The development of a genetic toolbox for an organism is just the first of many steps towards successfully engineering it. In this chapter, a handful of genetic parts and methodologies described in previous chapters will be applied towards elucidating the role of two aromatic degradation routes and several transporters identified in *R. opacus*. Understanding how this organism catabolizes lignin-derived aromatic compounds is critical for its application of upgrading biomass into value-added products. I engineered the five knockout strains used in this study, performed growth assays related to β-ketoadipate pathway knockout strains, and wrote the related sections of the manuscript.

6.1 Abstract

Lignin utilization has been identified as a key factor in biorefinery profitability. However, lignin depolymerization generates heterogeneous aromatic mixtures that inhibit microbial growth and conversion of lignocellulose to biochemicals. *Rhodococcus opacus* is a promising aromatic-catabolizing, oleaginous bacterium, but mechanisms for its aromatic tolerance and utilization remain under-characterized. To better understand aromatic catabolism, gene knockout experiments were performed to confirm the primary degradation routes of five aromatic compounds. Additionally, the specificity of three transporters that were upregulated based on

transcriptomic analysis were examined through gene knockouts. These results provide an improved understanding of aromatic bioconversion and facilitate development of *R. opacus* as a biorefinery host.

6.2 Introduction

Lignocellulosic biomass is a potential source of renewable bio-based fuels and chemicals. A large fraction of lignocellulose (~70-85%) is cellulose and hemicellulose, which can be depolymerized into monomeric sugars, such as glucose and xylose [1-2]. The remaining 15-30% of lignocellulose is lignin, a highly cross-linked, heterogeneous, and recalcitrant aromatic polymer [3]. Microbial conversion of lignocellulose holds promise as a new way to produce a variety of novel products and petrochemical replacements (e.g., drop-in fuels) [4-13]. Current biomass pretreatment approaches, which aim to increase sugar extractability from lignocellulose, also release aromatics from lignin that negatively affect microbial product titers, yields, and productivities [14-15]. Current processes separate the carbohydrate fraction from the lignin fraction, and the separated lignin is either discarded or burned for process heat or on-site electricity generation [16]. However, removal of lignin-derived inhibitors from sugar streams is still challenging and costly, and techno-economic analyses have identified the co-utilization of lignin as an important factor in biorefinery profitability [17-18]. Therefore, efforts are underway to develop microbial strains that can tolerate and convert inhibitory, lignin-derived compounds into value-added products [19-28].

Rhodococcus opacus PD630 (hereafter *R. opacus*) is a promising bacterial strain for producing valuable products from lignocellulose. *R. opacus* is a gram-positive soil bacterium that has been shown to have an inherently high tolerance to aromatic compounds [29-30]. Additionally, it accumulates high levels (up to ~78% as cell dry weight) of the biofuel precursor triacylglycerol (TAG) and demonstrates a moderately high growth rate [30-31]. *R. opacus* can natively consume, or

has been engineered to consume, hexose and pentose sugars present in lignocellulosic feedstocks [31-32]. *R. opacus* can also consume aromatic compounds found in depolymerized lignin, such as phenol, 4-hydroxybenzoate, and vanillate [33-35], making it uniquely qualified to utilize all three primary depolymerized components of lignocellulose.

The inherent heterogeneity of lignin is a major barrier to lignin valorization, and economical strategies require rapid conversion of complex aromatic mixtures to products [36-39]. Microbial degradation of diverse aromatic compounds is typically accomplished by biological funnelling, where compounds are converted into common metabolites (e.g., catechol and protocatechuate), prior to their catabolism via a central aromatic degradation pathway, such as the β-ketoadipate pathway [39-40]. In this work, gene knockout experiments in *R. opacus* confirmed the degradation route of five aromatic compounds through the β-ketoadipate pathway and the importance of upregulated transporters for growth on aromatic compounds.

6.3 Results and discussion

6.3.1 Examining the degradation routes of five aromatic compounds in *R. opacus*

The funnelling pathways convert aromatic compounds into two intermediates, CAT or PCA, for degradation by the β-ketoadipate pathway (Figure 6.1). Previous analyses of *Rhodococcus* strains suggest that VAN and HBA are converted to PCA, while GUA, BEN, and PHE are converted to CAT [41]. To confirm whether the five tested aromatic compounds (PHE, GUA, HBA, VAN, and BEN) were catabolized via the CAT or PCA branch of the β-ketoadipate pathway, critical genes from both branches were knocked out, and growth assays were performed. β-ketoadipate gene cluster #3, which contains the CAT degradation pathway, was knocked out by a disruption of RS30730 (cis,cis-muconate cycloisomerase; ΔCAT), while β-ketoadipate gene cluster #1, which contains the PCA degradation pathway, was knocked out by a disruption of

RS25360 (3-carboxy-cis,cis-muconate cycloisomerase; ΔPCA). The knockout of the PCA degradation pathway was complicated by the fact that other central β-ketoadipate pathway genes required for CAT degradation are located further downstream in β-ketoadipate gene cluster #1 (Figure 6.1A). To ensure transcription of those downstream genes, a constitutive promoter was integrated into the genome. The ΔCAT strain was unable to grow using PHE, GUA, and BEN as sole carbon sources, while the ΔPCA strain was unable to grow using HBA and VAN as sole carbon sources (Figure 6.1B, 6.1C). Together, these results confirm that PHE, GUA, and BEN are funnelled through the CAT branch, and HBA and VAN are funnelled through the PCA branch of the β-ketoadipate pathway in *R. opacus*.

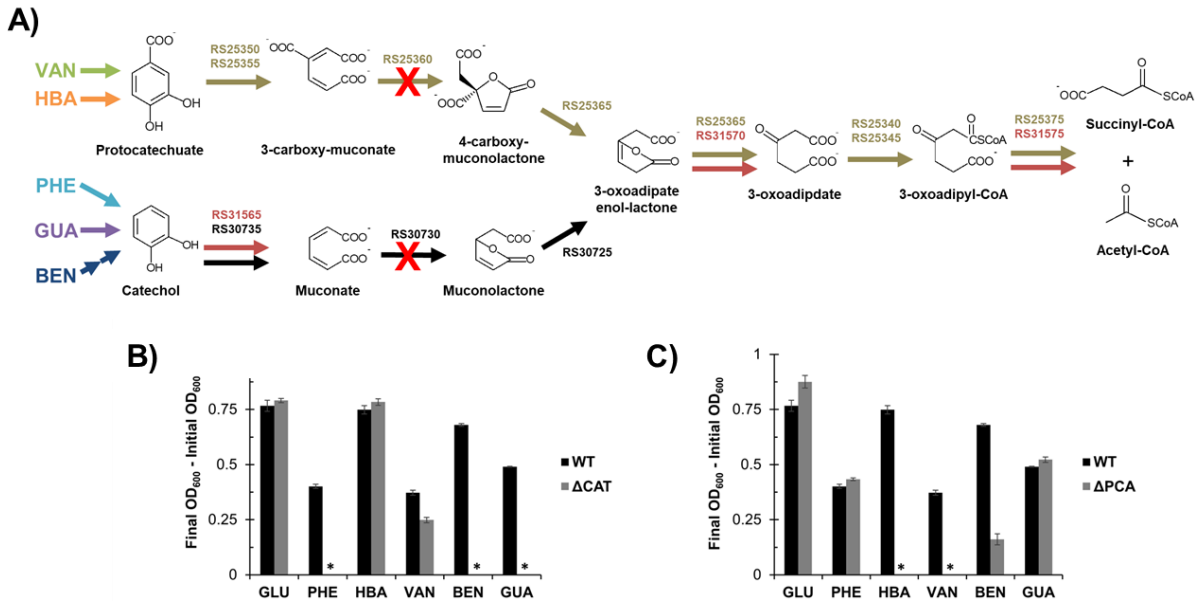


Figure 6.1. Identification of β-ketoadipate degradation branch degradation routes for five aromatic compounds in *R. opacus* PD630. **A)** Pathway map showing genes involved in β-ketoadipate pathway gene clusters in *R. opacus*. Gene codes are from the NCBI reference sequence NZ_CP003949.1. Red crosses represent gene knockouts performed. **B)** Difference between initial OD_{600} and final OD_{600} after 24 hours of growth for the RS30730 knockout (ΔCAT) strain and the WT strain using different sole carbon sources. **C)** Difference between initial OD_{600} and final OD_{600} after 24 hours of growth for the RS25360 knockout (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.

6.3.2 Transporters for lignin model compounds.

Transporter expression control may be related to aromatic tolerance mechanisms. The putative shikimate transporter (RS31355), which was significantly upregulated in other phenol-adapted *R. opacus* strains [34], was also upregulated in PHE (281 to 299-fold) and the mixture (87 to 295-fold) for both WT and PVHG6 relative to the glucose condition at both time points. An MFS transporter (RS33590) was moderately upregulated in response to PCA precursors VAN (137 to 193-fold) and HBA (10 to 15-fold), and the mixture (45 to 115-fold), but was minimally upregulated by CAT precursors PHE, GUA, and BEN (1.0 to 3.7-fold) in WT and PVHG6 at both time points (Table S11). Another MFS transporter (RS30810) located adjacent to the BEN degradation cluster was upregulated in BEN (1330 to 1720-fold) and the mixture (958 to 2620-fold) for both WT and PVHG6 at the first time point. The MFS transporter RS30810 was also moderately upregulated by CAT precursor GUA (57 to 62-fold), but only minimally upregulated by PCA precursors VAN (1.6 to 7.5-fold) and HBA (1.1 to 8.7-fold) in both strains compared to the glucose condition. The branch-dependent upregulation of MFS transporters RS33590 and RS30810 is similar to what was observed for the β-ketoadipate pathway. Overall, we observed multiple transporters whose expression was responsive to lignin model compounds.

To explore the role of transporters in aromatic tolerance and utilization, we compared the growth of three putative transporter knockout mutants (RS31355, RS33590, and RS30810) to that of the WT strain using different carbon sources (Figure 6.2). The growth of these knockout mutants and the WT strain were compared using 0.75 g/L PHE, 1.25 g/L GUA, 1.75 g/L VAN, 2 g/L HBA, 5 g/L BEN, and the combined mixture (2.5 g/L total aromatics) as sole carbon sources. The putative shikimate transporter (RS31355) knockout mutant had lower cell densities (OD_{600}) after 48 hours than the WT strain using PHE and the mixture as sole carbon sources, which matches

previous reports of RS31355 having a role in PHE transport (Yoneda et al., 2016). The putative vanillate transporter (RS33590) knockout mutant had lower OD_{600} values than the WT strain using VAN, PHE, and the mixture as sole carbon sources. The putative benzoate transporter (RS30810) knockout mutant had lower OD_{600} values than the WT strain using BEN, VAN, GUA, PHE, and the mixture as sole carbon sources (Figure 6.2). None of the transporter knockout mutants exhibited growth impairments on HBA (Figure 6.2). The impaired growth of the putative aromatic transporter knockout mutants on some compounds suggests transporter specificity and that the influx or efflux of aromatic compounds plays an important role in aromatic tolerance and utilization (Figure 6.2).

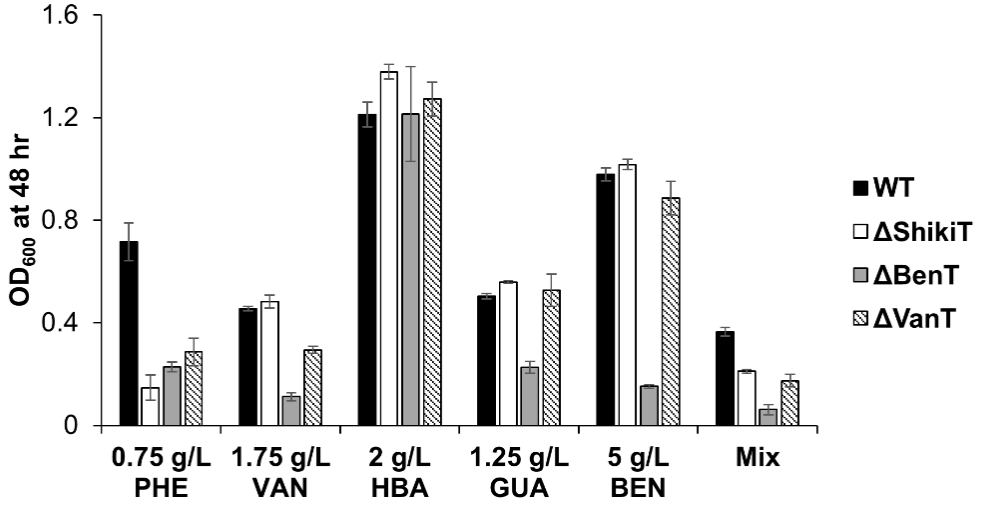


Figure 6.2. Growth of transporter knockout mutants compared to that of the WT strain using aromatic carbon sources. Bars represent the average of three biological replicates and error bars represent one standard deviation. PHE = phenol, VAN = vanillate, HBA = 4-hydroxybenzoate, GUA = guaiacol, BEN = sodium benzoate, and Mix = 0.5 g/L of PHE, VAN, HBA, GUA and BEN (2.5 g/L total aromatics). ΔShikiT = RS31355 knockout mutant, ΔBenT = RS30810 knockout mutant, and ΔVanT = RS33590 knockout mutant.

6.4 Materials and methods

6.4.1 Chemicals and strains

Chemicals were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise indicated, and *Rhodococcus opacus* PD630 (DSMZ 44193) was used as the ancestral strain (WT) for comparison to all mutated strains. For all growth experiments (unless otherwise indicated), cells were grown at 30 °C and 250 rpm in one of two defined minimal salts media (A or B; medium composition as previously described [41]). Culture media were filter-sterilized using a 0.22 µm filter, and pH was adjusted to 7.0 using 6M HCl or 2M NaOH. Carbon and nitrogen sources were added as sterile-filtered stock solutions. *R. opacus* strains were maintained on tryptic soy broth (TSB) plates supplemented with 1.5% agar.

6.4.2 Generation of *R. opacus* knockout mutants

Knockout mutants were generated via homologous recombination in the wild type *R. opacus* strain as described previously [42]. Briefly, a helper plasmid expressing a pair of recombinases (pDD120) was introduced via electroporation into WT *R. opacus*. Electrocompetent cells were made from this strain as previously described [41]. An integration vector containing an antibiotic resistance marker and DNA regions homologous to each target gene (~500 bp per homologous arm) was then introduced into the strain harboring the recombinase helper plasmid by electroporation [42]. Successful recombination was confirmed by colony PCR. See Table S16 for plasmids utilized for knockouts and Table S17 for all strains generated. Plasmid DNA was isolated from *Escherichia coli* DH10B for electroporation into *R. opacus* using the Zyppy Plasmid Miniprep Kit (Zymo).

6.4.3 Growth of *R. opacus* knockout mutants

For the β-ketoadipate pathway knockout experiment, the WT, ΔCAT, and ΔPCA strains were initially cultured in 2 mL of minimal medium B containing 1 g/L ammonium sulfate, 4 g/L glucose, and either no aromatic compound, 0.2 g/L PHE, 0.25 g/L GUA, 0.5 g/L HBA, 0.25 g/L VAN, or 0.5 g/L BEN. Cells were then centrifuged, washed with minimal medium B containing no carbon source, and re-suspended to an initial OD_{600} of 0.2 in 10 mL of minimal medium B containing 1 g/L ammonium sulfate and either 1 g/L glucose, 0.4 g/L PHE, 0.5 g/L GUA, 1.0 g/L HBA, 0.5 g/L VAN, or 1 g/L BEN as sole carbon sources. For the transporter knockout experiment, all strains were initially cultured in 2 mL of minimal medium A supplemented with 0.2 g/L of sodium benzoate, phenol, guaiacol, 4-hydroxybenzoate, and vanillate (1 g/L total aromatics) and 1 g/L ammonium sulfate for 24 hours, and then subcultured into 10 mL of the same media for 24 hours to an OD_{600} of ~0.5. Cells were then diluted to an OD_{600} of 0.05 in 10 mL of minimal media A supplemented with 1 g/L ammonium sulfate and either 0.75 g/L PHE, 1.75 g/L VAN, 2 g/L HBA, 1.25 g/L GUA, 5 g/L BEN or a mixture of 0.5 g/L of each compound (2.5 g/L total aromatics). The optical density at 600 nm was measured for all cultures after 48 hours of growth.

6.5 References

1. Sun, Y.; Cheng, J., Hydrolysis of lignocellulosic materials for ethanol production: a review. Bioresour. Technol. 2002, 83 (1), 1-11.

2. Wheeldon, I.; Christopher, P.; Blanch, H., Integration of heterogeneous and biochemical catalysis for production of fuels and chemicals from biomass. Curr. Opin. Biotechnol. 2017, 45, 127-135.

3. Pandey, M. P.; Kim, C. S., Lignin Depolymerization and Conversion: A Review of Thermochemical Methods. Chem. Eng. Technol. 2011, 34, 29-41.

4. Bhat, A. H.; Dasan, Y. K.; Khan, I., Extraction of Lignin from Biomass for Biodiesel Production. In Agricultural Biomass Based Potential Materials, Springer, Cham: 2015; pp 155-179.

5. Hahn-Hagerdal, B.; Galbe, M.; Gorwa-Grauslund, M. F.; Liden, G.; Zacchi, G., Bio-ethanol-the fuel of tomorrow from the residues of today. Trends Biotechnol. 2006, 24 (12), 549-56.

6. Xu, P.; Qiao, K.; Ahn, W. S.; Stephanopoulos, G., Engineering Yarrowia lipolytica as a platform for synthesis of drop-in transportation fuels and oleochemicals. Proc. Natl. Acad. Sci. U. S. A. 2016, 113 (39), 10848-53.

7. Lee, J. W.; Yi, J.; Kim, T. Y.; Choi, S.; Ahn, J. H.; Song, H.; Lee, M. H.; Lee, S. Y., Homo-succinic acid production by metabolically engineered Mannheimia succiniciproducens. Metab. Eng. 2016, 38, 409-417.

8. Yang, J. E.; Park, S. J.; Kim, W. J.; Kim, H. J.; Kim, B. J.; Lee, H.; Shin, J.; Lee, S. Y., One-step fermentative production of aromatic polyesters from glucose by metabolically engineered Escherichia coli strains. Nat. Commun. 2018, 9 (1), 79.

9. Zhou, Y. J.; Buijs, N. A.; Zhu, Z.; Qin, J.; Siewers, V.; Nielsen, J., Production of fatty acid-derived oleochemicals and biofuels by synthetic yeast cell factories. Nat. Commun. 2016, 7, 11709.

10. Zhou, Y. J.; Buijs, N. A.; Zhu, Z.; Gomez, D. O.; Boonsombuti, A.; Siewers, V.; Nielsen, J., Harnessing Yeast Peroxisomes for Biosynthesis of Fatty-Acid-Derived Biofuels and Chemicals with Relieved Side-Pathway Competition. J. Am. Chem. Soc. 2016, 138 (47), 15368-15377.

11. Jones, J. A.; Vernacchio, V. R.; Sinkoe, A. L.; Collins, S. M.; Ibrahim, M. H. A.; Lachance, D. M.; Hahn, J.; Koffas, M. A. G., Experimental and computational optimization of an Escherichia coli co-culture for the efficient production of flavonoids. Metab. Eng. 2016, 35, 55-63.

12. Qiao, K.; Wasylenko, T. M.; Zhou, K.; Xu, P.; Stephanopoulos, G., Lipid production in Yarrowia lipolytica is maximized by engineering cytosolic redox metabolism. Nat. Biotechnol. 2017, 35 (2), 173-177.

13. Pereira, B.; Li, Z. J.; De Mey, M.; Lim, C. G.; Zhang, H.; Hoeltgen, C.; Stephanopoulos, G., Efficient utilization of pentoses for bioproduction of the renewable two-carbon compounds ethylene glycol and glycolate. Metab. Eng. 2016, 34, 80-87.

14. Pienkos, P. T.; Zhang, M., Role of pretreatment and conditioning processes on toxicity of lignocellulosic biomass hydrolysates. Cellulose 2009, 16, 743-762.

15. Ibraheem, O.; Ndimba, B. K., Molecular adaptation mechanisms employed by ethanologenic bacteria in response to lignocellulose-derived inhibitory compounds. Int. J. Biol. Sci. 2013, 9 (6), 598-612.

16. Ragauskas, A. J.; Beckham, G. T.; Biddy, M. J.; Chandra, R.; Chen, F.; Davis, M. F.; Davison, B. H.; Dixon, R. A.; Gilna, P.; Keller, M.; Langan, P.; Naskar, A. K.; Saddler, J. N.; Tschaplinski, T. J.; Tuskan, G. A.; Wyman, C. E., Lignin valorization: improving lignin processing in the biorefinery. Science (New York, N.Y.) 2014, 344 (6185), 1246843.

17. Balan, V.; Chiaramonti, D.; Kumar, S., Review of US and EU initiatives toward development, demonstration, and commercialization of lignocellulosic biofuels. Biofuels, Bioprod. Biorefin. 2013, 7 (6), 732-759.

18. Valdivia, M.; Galan, J. L.; Laffarga, J.; Ramos, J.-L., Biofuels 2020: Biorefineries based on lignocellulosic materials. Microb. Biotechnol. 2016, 9 (5), 585-594.

19. Jin, Y. S.; Cate, J. H., Metabolic engineering of yeast for lignocellulosic biofuel production. Curr. Opin. Chem. Biol. 2017, 41, 99-106.

20. Blazeck, J.; Hill, A.; Liu, L.; Knight, R.; Miller, J.; Pan, A.; Otoupal, P.; Alper, H. S., Harnessing Yarrowia lipolytica lipogenesis to create a platform for lipid and biofuel production. Nat. Commun. 2014, 5, 3131.

21. Vardon, D. R.; Franden, M. A.; Johnson, C. W.; Karp, E. M.; Guarnieri, M. T.; Linger, J. G.; Salm, M. J.; Strathmann, T. J.; Beckham, G. T., Adipic acid production from lignin. Energy Environ. Sci. 2015, 8 (2), 617-628.

22. Bugg, T. D. H.; Rahmanpour, R., Enzymatic conversion of lignin into renewable chemicals. Curr. Opin. Chem. Biol. 2015, 29 (Supplement C), 10-17.

23. Dunlop, M. J., Engineering microbes for tolerance to next-generation biofuels. Biotechnol. Biofuels 2011, 4, 32.

24. Keating, D. H.; Zhang, Y.; Ong, I. M.; McIlwain, S.; Morales, E. H.; Grass, J. A.; Tremaine, M.; Bothfeld, W.; Higbee, A.; Ulbrich, A.; Balloon, A. J.; Westphall, M. S.; Aldrich, J.; Lipton, M. S.; Kim, J.; Moskvin, O. V.; Bukhman, Y. V.; Coon, J. J.; Kiley, P. J.; Bates, D. M.; Landick, R., Aromatic inhibitors derived from ammonia-pretreated lignocellulose hinder bacterial ethanologenesis by activating regulatory circuits controlling inhibitor efflux and detoxification. Front. Microbiol. 2014, 5, 402.

25. Henske, J. K.; Gilmore, S. P.; Knop, D.; Cunningham, F. J.; Sexton, J. A.; Smallwood, C. R.; Shutthanandan, V.; Evans, J. E.; Theodorou, M. K.; O'Malley, M. A., Transcriptomic characterization of Caecomyces churrovis: a novel, non-rhizoid-forming lignocellulolytic anaerobic fungus. Biotechnol. Biofuels 2017, 10, 305.

26. Le, R. K.; Das, P.; Mahan, K. M.; Anderson, S. A.; Wells, T., Jr.; Yuan, J. S.; Ragauskas, A. J., Utilization of simultaneous saccharification and fermentation residues as feedstock for lipid accumulation in Rhodococcus opacus. AMB Express 2017, 7 (1), 185.

27. Fischer, C. R.; Klein-Marcuschamer, D.; Stephanopoulos, G., Selection and optimization of microbial hosts for biofuels production. Metab. Eng. 2008, 10 (6), 295-304.

28. Pham, H. L.; Wong, A.; Chua, N.; Teo, W. S.; Yew, W. S.; Chang, M. W., Engineering a riboswitch-based genetic platform for the self-directed evolution of acid-tolerant phenotypes. Nat. Commun. 2017, 8 (1), 411.

29. Tsitko, I. V.; Zaitsev, G. M.; Lobanok, A. G.; Salkinoja-Salonen, M. S., Effect of aromatic compounds on cellular fatty acid composition of Rhodococcus opacus. Appl. Environ. Microbiol. 1999, 65 (2), 853-5.

30. Alvarez, H. M.; Mayer, F.; Fabritius, D.; Steinbuchel, A., Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630. Arch. Microbiol. 1996, 165 (6), 377-86.

31. Kurosawa, K.; Boccazzi, P.; de Almeida, N. M.; Sinskey, A. J., High-cell-density batch fermentation of Rhodococcus opacus PD630 using a high glucose concentration for triacylglycerol production. J. Biotechnol. 2010, 147 (3-4), 212-8.

32. Kurosawa, K.; Wewetzer, S. J.; Sinskey, A. J., Engineering xylose metabolism in triacylglycerol-producing Rhodococcus opacus for lignocellulosic fuel production. Biotechnol. Biofuels 2013, 6 (1), 134.

33. Kosa, M.; Ragauskas, A. J., Bioconversion of lignin model compounds with oleaginous Rhodococci. Appl. Microbiol. Biotechnol. 2012, 93 (2), 891-900.

34. Yoneda, A.; Henson, W. R.; Goldner, N. K.; Park, K. J.; Forsberg, K. J.; Kim, S. J.; Pesesky, M. W.; Foston, M.; Dantas, G.; Moon, T. S., Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630. Nucleic Acids Res. 2016, 44 (5), 2240-54.

35. Xie, S.; Sun, Q.; Pu, Y.; Lin, F.; Sun, S.; Wang, X.; Ragauskas, A. J.; Yuan, J. S., Advanced Chemical Design for Efficient Lignin Bioconversion. ACS Sustainable Chem. Eng. 2017, 5 (3), 2215-2223.

36. Rodriguez, A.; Salvachúa, D.; Katahira, R.; Black, B. A.; Cleveland, N. S.; Reed, M.; Smith, H.; Baidoo, E. E. K.; Keasling, J. D.; Simmons, B. A.; Beckham, G. T.; Gladden, J. M., Base-Catalyzed Depolymerization of Solid Lignin-Rich Streams Enables Microbial Conversion. ACS Sustainable Chem. Eng. 2017, 5 (9), 8171-8180.

37. Van den Bosch, S.; Renders, T.; Kennis, S.; Koelewijn, S. F.; Van den Bossche, G.; Vangeel, T.; Deneyer, A.; Depuydt, D.; Courtin, C. M.; Thevelein, J. M.; Schutyser, W.; Sels, B. F., Integrating lignin valorization and bio-ethanol production: on the role of Ni-Al2O3 catalyst pellets during lignin-first fractionation. Green Chem. 2017, 19 (14), 3313-3326.

38. Shuai, L.; Amiri, M. T.; Questell-Santiago, Y. M.; Heroguel, F.; Li, Y.; Kim, H.; Meilan, R.; Chapple, C.; Ralph, J.; Luterbacher, J. S., Formaldehyde stabilization facilitates lignin monomer production during biomass depolymerization. Science (New York, N.Y.) 2016, 354 (6310), 329-333.

39. Linger, J. G.; Vardon, D. R.; Guarnieri, M. T.; Karp, E. M.; Hunsinger, G. B.; Franden, M. A.; Johnson, C. W.; Chupka, G.; Strathmann, T. J.; Pienkos, P. T.; Beckham, G. T., Lignin valorization through integrated biological funneling and chemical catalysis. Proc. Natl. Acad. Sci. U. S. A. 2014, 111, 12013-12018.

40. Harwood, C. S.; Parales, R. E., The beta-ketoadipate pathway and the biology of self-identity. Annu. Rev. Microbiol. 1996, 50, 553-90.

41. DeLorenzo, D. M.; Henson, W. R.; Moon, T. S., Development of Chemical and Metabolite Sensors for Rhodococcus opacus PD630. ACS synthetic biology 2017, 6 (10), 1973-1978.

42. DeLorenzo, D. M.; Rottinghaus, A. G.; Henson, W. R.; Moon, T. S., Molecular Toolkit for Gene Expression Control and Genome Modification in Rhodococcus opacus PD630. ACS synthetic biology 2018, 7 (2), 727-738.

Chapter 7: Conclusion and future directions

7.1 Conclusion

The discovery of fossil fuels and their by-products has allowed humanity to rapidly industrialize and achieve many notable accomplishments over the past two and a half centuries. As the human population increases, however, fossil fuel consumption continues to increase while reserves diminish. Furthermore, research has definitively linked greenhouse gas emissions from the burning of fossil fuels with detrimental changes in atmospheric composition. As we improve our understanding of these changes, society must face the harmful environmental side-effects of large-scale carbon burning. Alternative sources of fuel and petroleum-derived products are required to help displace those currently produced from hydrocarbon deposits. One such alternative is renewable plant biomass, which, when depolymerized, forms easily fermentable sugars along with aromatic compounds similar to those found in petroleum. Biomass breakdown products—particularly those derived from the lignin fraction—exhibit a large diversity of monomers and oligomers that make chemical separations difficult.

A biological catalyst can be employed to funnel a variety of different biomass breakdown products into single value-added products using their vast catabolic pathways and central metabolism. *Rhodococcus opacus* PD630 was chosen as a suitable biological catalyst due to its ability to consume an array of compounds found in depolymerized biomass, such as aromatics derived from lignin and hexose and pentose sugars derived from the cellulose and hemicellulose fractions, respectively. Lignocellulose-derived feedstocks can readily be funneled into the central metabolite acetyl CoA and then into lipid metabolism, wherein *R. opacus* generates high levels of triacylglycerides (TAGs). TAGs are readily converted into biodiesel via a simple

transesterification reaction. Furthermore, *R. opacus* has a high native tolerance to aromatic compounds, which are generally toxic and inhibitory to most microbes, allowing higher substrate loads and greater product titers.

The work described in this dissertation promotes the use of *R. opacus* as a microbial chassis for the conversion of biomass-derived products into biofuels, or other value-added products, and provides insight into its aromatic catabolism in a number of ways. These advances include: 1) the development of strong constitutive promoters for the overexpression of heterologous genes, 2) the development of chemical and metabolite sensors for inducible and dynamic gene expression, 3) the characterization of native and endogenous plasmid backbones and resistance markers, 4) a heterologous T7 RNA polymerase platform for high levels of gene transcription, 5) the demonstration of genetic logic circuits for programmable gene expression, 6) a recombinase-based platform for gene knockouts and insertions, 7) a CRISPR interference (CRISPRi) platform for targeted gene repression, 8) the identification of stable reference genes for RT-qPCR applications, 9) insight into aromatic degradation through the β-ketoadipate pathway, and 10) insight into the role of aromatic transporters. Taken together, this work greatly advances the ability to engineer *R. opacus* for any desired application, in addition to providing understanding into its catabolism of aromatic compounds.

7.2 Future directions

7.2.1 Investigating aromatic-related degradation pathways, gene regulation, and tolerance mechanisms in *R. opacus*

The utilization of a biological catalyst to "funnel" a mixture of biomass breakdown products through an organism's metabolism into a single, high-value product is one approach to

upgrading a low-value, heterogeneous feedstock.[1] *R. opacus* is one of several prokaryotic and eukaryotic organisms (e.g., *Pseudomonas putida*[2], *Escherichia coli*[3], *Trichosporon cutaneum*[4]) that could be used as a biological funnel. One of the primary requirements for this function is that the organism be able to import and consume the majority of the lignin breakdown products in the mixture. The mechanisms that facilitate these processes in *R. opacus* are understudied, as only three transporters—those described in Chapter 6—and a handful of aromatic degradation pathways have been characterized.[5-10] Thus far, *R. opacus* has been demonstrated to consume the lignin model compounds 4-hydroxybenzoate[5], benzoate[5], phenol[5], vanillate[5], guaiacol[5], catechol[5], protocatechuate[5], ferulic acid[11], and trans-*p*-coumaric acid[12], in addition to several types of actual lignin breakdown products, such as switchgrass pyrolysis oil[13], alkali-treated corn stover and poplar wood [14-16], and depolymerized kraft lignin[17]. Very recently, an engineered strain of *R. opacus* that secretes a heterologous laccase was demonstrated to grow on insoluble kraft lignin.[18] Further research into the types of aromatics that *R. opacus* can consume and the mechanisms that allow it to do so (i.e., compound import, funneling pathways, redox shifts) should be conducted.

Knowing that *R. opacus* is capable of consuming aromatics derived from lignin is just the first step in utilizing this host as a biological catalyst. While transcriptomic and proteomic data has been analyzed to identify gene candidates for many aromatic funneling pathways, gene knockouts still need to be performed to confirm these degradation routes, such as was done in Chapter 6 with the catechol and protocatechuate branches of the β-ketoadipate pathway.[5, 18] Furthermore, as the current reference genomes available for *R. opacus* have been discovered to contain many inaccuracies in gene annotation, a more comprehensive bioinformatic analysis based on protein homology should be performed, as has been done in *Pseudomonas putida*. These pathways can then also be confirmed in *R. opacus* through gene knockouts.[19] To fill in researchers knowledge

of other critical pathways and funneling enzymes in *R. opacus*, additional transcriptomic and proteomic data should be collected on a much wider variety of lignin model compounds, in addition to various forms of depolymerized lignin. To maximize the potential of this host, a comprehensive knowledge of all its transporters, catabolic, and other aromatic-related genes should be investigated.

If considering a cost-effective and efficient industrial process, *R. opacus* must not only be able to consume lignin-derived feedstocks but also perform this function at a high rate. On a single lignin model compound (phenol), *R. opacus* has demonstrated record level microbial degradation rates at the highest reported concentrations for a free cell.[20] A mixture of lignin model compounds proved more problematic though, as *R. opacus* has been observed to consume aromatics in a preferential order, rather than simultaneously.[5] The source of this consumption hierarchy could arise through either enzyme kinetics (i.e., varying transport and degradation rates for each compound) or via transcriptional, post-transcriptional, translational, or post-translational regulation. In *P. putida*, another viable biological catalyst for lignin valorization, a single catabolite repression control (Crc) protein was found to inhibit the consumption of 4-hydroxybenzoate and vanillate when either glucose or acetate were present.[21] When this gene was deleted, higher rates of conversion of these compounds was measured. Knockout studies of regulators identified in *R. opacus* to be upregulated in previous transcriptomic studies conducted on mixtures of lignin model compounds could lead to similar discoveries. Furthermore, small RNA and proteomic studies conducted when *R. opacus* is grown on single aromatic compounds and mixtures of compounds could provide insight into post-transcriptional, translational and post-translational regulation mechanisms, which to date has not been studied in *R. opacus*.

Regarding post-transcriptional regulation, recent work has revealed that in the related organism *Mycobacteria tuberculosis,* small regulatory RNA (sRNA) act in a functionally distinct manner than in model organisms, such as *E. coli*, to modulate translational rates.[22] Rather than depending on a protein to assist in sRNA:mRNA binding, as occurs with the Hfq protein in *E. coli*, these mycobacterial sRNA are protein independent and use looping structures to facilitate gene repression. Small RNA have been previously detected in *R. opacus*, but no research has been performed to further investigate the mechanism or function of these non-coding RNAs.[20] Additional studies into native sRNA could provide both insight into an entire level of regulatory mechanisms previously ignored in *R. opacus* and also enable the development of synthetic RNA regulators for tunable gene repression.

The final topic related to lignin conversion that should be investigated once catabolic pathways and their regulation are better understood is the tolerance mechanisms that *R. opacus* employs in the presence of toxic lignin-derived compounds. Recent work has found that wild type *R. opacus* modulates its lipidome in the presence of phenol, including changes such as fewer double bonds in multiple lipid species and increased levels of phosphatidylinositols.[23] Another study found that *R. opacus* mutants adapted on mixtures of aromatics were able to generate higher levels of triacylglycerol than the wild type strain.[5, 20] These lipid changes are hypothesized to alter the ability of phenolics to cross the cellular membrane, although it is unclear if the changes are inhibiting or promoting transfer. It is also not known which genes in particular are responsible for the observed changes in the lipidome. Past studies have identified lists of possible gene targets, whether they be genes that were up- or downregulated in the presence of aromatics or genes that were mutated during adaptive evolution on aromatics, but no follow-up work has yet been performed to narrow down these lists to the key genes.

Another suspected aromatic tolerance mechanism in *R. opacus* is its catabolic capabilities. As *R. opacus* consumes inhibitory aromatics inside or nearby the cell, the effective concentration of those compounds is lowered. The more efficiently this process occurs, the higher the observed tolerance. This is supported by the fact that different lineages of *R. opacus* mutants adapted on mixtures of aromatics had recurrent mutations in genes related to the reduction-oxidation (redox) state within the cell. In particular, genes for superoxide dismutase and cytochrome ubiquinol oxidase were frequently mutated, with protein models predicting reduced activity.[5] The activity of superoxide dismutase was tested between one adapted strain (PVHG6) and wild type *R. opacus*, and the mutated version was found to have substantially lower activity. The role of superoxide dismutase is to remove superoxide radicals from the cell. Superoxide radicals, however, are required as part of the aromatic ring cleaving step. Thus, an increase in the prevalence of superoxide radicals could lead to improvements in aromatic consumption, facilitating increases in tolerance. Further research via gene knockouts and knockdowns, gene overexpressions, and the re-creation of single nucleotide polymorphisms (SNPs) observed in adapted strains should be performed for gene targets identified in both the lipidome and redox state tolerance mechanisms to better understand the sources of tolerance in *R. opacus*.

7.2.2 Engineering branched-chain fatty acid production in *R. opacus*

The engineering of lipid metabolism in common lab microbes, such as *E. coli* and *Saccharomyces cerevisiae*, has primarily focused on altering metabolic pathways to lead to the overproduction of straight chain fatty acids.[24-27] *R. opacus* is well-known for being an oleaginous organism innately capable of high levels of TAG accumulation (up to 76% cell dry weight when grown on gluconate).[28] TAGs are comprised of three fatty acid esters attached to a glycerol and are a valuable product due to the ease in which they can be converted into biodiesel via a simple

transesterification reaction that generates straight chain free fatty acids.[29] However, fatty acids that contain branching have a number of improved properties relevant to biofuel—valuable military-grade jet fuel, in particular—including a lower freezing point and better cold-flow.[27, 30-31] Furthermore, these properties can be tuned by altering the number and location of the branch points.[31] Other valuable applications of branched-chain fatty acids include petroleum replacements for use as plasticizers, surfactants, and flavorings.[32]

Branched-chain fatty acid synthesis has been engineered in several model organisms, such as *E. coli* and *S. cerevisiae*, using glucose as a feedstock.[27, 32] The benefit of using *R. opacus* for branched-chain fatty acid production is that low-grade lignin can be used as a feedstock instead of the more valuable glucose. In addition, previous lipid analysis of *R. opacus* strains have found varying levels of native branched-chain fatty acids, up to 34% when grown on aromatics.[28, 33-34] This innate ability to produce branched-chain fatty acids, in addition to being an oleaginous organism, could result in a reduction in engineering efforts to generate high product titers compared to other potential aromatic consuming microbes. Furthermore, as branched-chain fatty acids can lead to more valuable products that traditional biofuels, an industrial process utilizing this strain would more easily achieve a positive economic return.

7.2.3 Domestication of *R. opacus*

Microbial cell culture in a laboratory or industrial setting significantly contrasts with what an organism is exposed to in its native habitat. For instance, *R. opacus* is natively a soil bacterium, but it is cultured in liquid media. Additionally, native cellular mechanisms that are beneficial in the wild, such as the ability to recombine DNA, is a nuisance when an engineered DNA construct is inadvertently disrupted. Microbial strains have been previously domesticated to a lab setting similar to how higher-order organisms, such as dogs, have been domesticated for human-

associated life.[35] Domestication to improve cell growth can be achieved simply by serially culturing the microbe in the desired media over many generations. For instance, *R. opacus* strains passaged on a mixture of aromatic compounds were found to have lost up to two of the nine endogenous plasmids that the wild type harbors.[5] This loss of two large plasmids (172 kilo base pairs [kbps] and 97 kbps) could improve *R. opacus* growth by reducing the metabolic cost of DNA repair and replication. These plasmids are thought to harbor genes that are non-essential for cell growth but could be beneficial to the cell under extreme conditions.[36] In a controlled setting, these genes are likely unnecessary and the removal of all nine endogenous plasmids could lead to an improvement in the growth rate of *R. opacus*. Furthermore, origins of replication and maintenance systems (i.e., toxin-antitoxin systems) could be isolated from these removed endogenous plasmids to create a suite of synthetic plasmid backbones.

Passive domestication of a microbe is not the only option, as rational engineering can be applied to make a better performing lab strain. For instance, the common *E. coli* DH10B cloning strain was modified from its ancestor *E. coli* K-12 MG1655 in several ways to improve plasmid assembly and stability. To reduce recombination events that disrupt plasmid DNA, the *recA* recombinase gene was deleted from the genome. To avoid possible issues with plasmid propagation, native restriction/modification systems that could digest introduced plasmid DNA were also knocked out.[37] To further reduce metabolic burden, unnecessary genes, such as redundant gene duplications that add genetic bloat to the genome, could be removed from the *R. opacus* genome. In a K-12 strain of *E. coli*, 15% of the genome was deleted leading to improvements in growth and allowing for simplified engineering efforts.[38] Any efforts to domesticate *R. opacus* into a more stable host would be beneficial if the strain is to be used in an industrial setting in the future.

7.2.4 Improving *R. opacus* consumption on actual lignin breakdown products

The eventual goal for *R. opacus* is to use this host as a biological catalyst to convert low-value lignin breakdown products into higher value compounds, such as lipids. Thus far, it has been shown that *R. opacus* is capable of growth on a variety of aromatic compounds, in addition to several forms of real-world lignin.[5, 15, 20, 39-40] The adaptive capacity of *R. opacus* has also been demonstrated, as the growth and consumption of different inhibitory compounds found in depolymerized lignocellulose (e.g., furans, hydroxymethylfurfural, phenolics, organic acids) has been improved through evolutionary selection.[5, 20, 41] However, no adaption of *R. opacus* on real-world lignin has been performed. Performing directed evolution using real lignin as a feedstock, which contains a greater diversity of compounds than previously published carbon sources, would be the most beneficial step to improving *R. opacus* as a chassis for lignin valorization. Performing a multi-omics analysis on these adapted strains could provide a wealth of insight into the catabolic and tolerance mechanisms that *R. opacus* possesses and provide even more gene targets that can be investigated using the tools developed in this dissertation.

7.3 References

1. Linger, J. G.; Vardon, D. R.; Guarnieri, M. T.; Karp, E. M.; Hunsinger, G. B.; Franden, M. A.; Johnson, C. W.; Chupka, G.; Strathmann, T. J.; Pienkos, P. T.; Beckham, G. T., Lignin valorization through integrated biological funneling and chemical catalysis. *Proceedings of the National Academy of Sciences* 2014, *111*, 12013-12018.

2. Ravi, K.; García-Hidalgo, J.; Gorwa-Grauslund, M. F.; Lidén, G., Conversion of lignin model compounds by Pseudomonas putida KT2440 and isolates from compost. *Appl. Microbiol. Biotechnol.* 2017, *101* (12), 5059-5070.

3. Wu, W.; Dutta, T.; Varman, A. M.; Eudes, A.; Manalansan, B.; Loque, D.; Singh, S., Lignin Valorization: Two Hybrid Biochemical Routes for the Conversion of Polymeric Lignin into Value-added Chemicals. *Sci Rep* 2017, *7* (1), 8420.

4. Liu, W.; Wang, Y.; Yu, Z.; Bao, J., Simultaneous saccharification and microbial lipid fermentation of corn stover by oleaginous yeast Trichosporon cutaneum. *Bioresour Technol* 2012, *118*, 13-8.

5. Henson, W. R.; Campbell, T.; DeLorenzo, D. M.; Gao, Y.; Berla, B.; Kim, S. J.; Foston, M.; Moon, T. S.; Dantas, G., Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus. *Metab. Eng.* 2018, *49*, 69-83.

6. Di Canito, A.; Zampolli, J.; Orro, A.; D'Ursi, P.; Milanesi, L.; Sello, G.; Steinbüchel, A.; Di Gennaro, P. J. B. G., Genome-based analysis for the identification of genes involved in o-xylene degradation in Rhodococcus opacus R7. 2018, *19* (1), 587.

7. Martínková, L.; Uhnáková, B.; Pátek, M.; Nešvera, J.; Křen, V., Biodegradation potential of the genus Rhodococcus. *Environment International* 2009, *35* (1), 162-177.

8. Eulberg, D.; Kourbatova, E. M.; Golovleva, L. A.; Schlömann, M., Evolutionary relationship between chlorocatechol catabolic enzymes from Rhodococcus opacus 1CP and their counterparts in proteobacteria: sequence divergence and functional convergence. *Journal of bacteriology* 1998, *180* (5), 1082-1094.

9. Zaitsev, G. M.; Uotila, J. S.; Tsitko, I. V.; Lobanok, A. G.; Salkinoja-Salonen, M. S., Utilization of Halogenated Benzenes, Phenols, and Benzoates by Rhodococcus opacus GM-14. *Applied and environmental microbiology* 1995, *61* (12), 4191-4201.

10. Kitagawa, W.; Kimura, N.; Kamagata, Y., A novel p-nitrophenol degradation gene cluster from a gram-positive bacterium, Rhodococcus opacus SAO101. *Journal of bacteriology* 2004, *186* (15), 4894-4902.

11. Plaggenborg, R.; Overhage, J.; Loos, A.; Archer, J. A.; Lessard, P.; Sinskey, A. J.; Steinbuchel, A.; Priefert, H., Potential of Rhodococcus strains for biotechnological vanillin production from ferulic acid and eugenol. *Appl. Microbiol. Biotechnol.* 2006, *72* (4), 745-55.

12. Wang, B.; Rezenom, Y. H.; Cho, K.-C.; Tran, J. L.; Lee, D. G.; Russell, D. H.; Gill, J. J.; Young, R.; Chu, K.-H., Cultivation of lipid-producing bacteria with lignocellulosic biomass: Effects of inhibitory compounds of lignocellulosic hydrolysates. *Bioresour. Technol.* 2014, *161*, 162-170.

13. Wei, Z.; Zeng, G.; Kosa, M.; Huang, D.; Ragauskas, A. J., Pyrolysis oil-based lipid production as biodiesel feedstock by Rhodococcus opacus. *Appl. Biochem. Biotechnol.* 2015, *175* (2), 1234-1246.

14. Le, R. K.; Wells Jr, T.; Das, P.; Meng, X.; Stoklosa, R. J.; Bhalla, A.; Hodge, D. B.; Yuan, J. S.; Ragauskas, A. J., Conversion of corn stover alkaline pre-treatment waste streams into biodiesel via Rhodococci. *RSC Advances* 2017, *7* (7), 4108-4115.

15. He, Y.; Li, X.; Ben, H.; Xue, X.; Yang, B., Lipid Production from Dilute Alkali Corn Stover Lignin by Rhodococcus Strains. *ACS Sustainable Chemistry & Engineering* 2017, *5* (3), 2302-2311.

16. Li, X.; He, Y.; Zhang, L.; Xu, Z.; Ben, H.; Gaffrey, M. J.; Yang, Y.; Yang, S.; Yuan, J. S.; Qian, W.-J.; Yang, B., Discovery of potential pathways for biological conversion of poplar wood into lipids by co-fermentation of Rhodococci strains. *Biotechnology for Biofuels* 2019, *12* (1), 60.

17. Zhao, C.; Xie, S.; Pu, Y.; Zhang, R.; Huang, F.; Ragauskas, A. J.; Yuan, J. S., Synergistic enzymatic and microbial lignin conversion. *Green Chemistry* 2016, *18* (5), 1306-1312.

18. Xie, S.; Sun, S.; Lin, F.; Li, M.; Pu, Y.; Cheng, Y.; Xu, B.; Liu, Z.; da Costa Sousa, L.; Dale, B. E.; Ragauskas, A. J.; Dai, S. Y.; Yuan, J. S., Mechanism-Guided Design of Highly

Efficient Protein Secretion and Lipid Conversion for Biomanufacturing and Biorefining. *Advanced Science 0* (0), 1801980.

19. Jimenez, J. I.; Minambres, B.; Garcia, J. L.; Diaz, E., Genomic analysis of the aromatic catabolic pathways from Pseudomonas putida KT2440. *Environ Microbiol* 2002, *4* (12), 824-41.

20. Yoneda, A.; Henson, W. R.; Goldner, N. K.; Park, K. J.; Forsberg, K. J.; Kim, S. J.; Pesesky, M. W.; Foston, M.; Dantas, G.; Moon, T. S., Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630. *Nucleic Acids Res* 2016, *44* (5), 2240-54.

21. Johnson, C. W.; Abraham, P. E.; Linger, J. G.; Khanna, P.; Hettich, R. L.; Beckham, G. T., Eliminating a global regulator of carbon catabolite repression enhances the conversion of aromatic lignin monomers to muconate in Pseudomonas putida KT2440. *Metab Eng Commun* 2017, *5*, 19-25.

22. Mai, J.; Rao, C.; Watt, J.; Sun, X.; Lin, C.; Zhang, L.; Liu, J., Mycobacterium tuberculosis 6C sRNA binds multiple mRNA targets via C-rich loops independent of RNA chaperones. *Nucleic Acids Res* 2019.

23. Henson, W. R.; Hsu, F. F.; Dantas, G.; Moon, T. S.; Foston, M., Lipid metabolism of phenol-tolerant Rhodococcus opacus strains for lignin bioconversion. *Biotechnology for Biofuels* 2018, *11*, 339.

24. Runguphan, W.; Keasling, J. D., Metabolic engineering of Saccharomyces cerevisiae for production of fatty acid-derived biofuels and chemicals. *Metab. Eng.* 2014, *21*, 103-13.

25. Xu, P.; Li, L.; Zhang, F.; Stephanopoulos, G.; Koffas, M., Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proc Natl Acad Sci U S A* 2014, *111* (31), 11299-304.

26. Jawed, K.; Mattam, A. J.; Fatma, Z.; Wajid, S.; Abdin, M. Z.; Yazdani, S. S., Engineered Production of Short Chain Fatty Acid in Escherichia coli Using Fatty Acid Synthesis Pathway. *PLoS ONE* 2016, *11* (7), e0160035.

27. Bentley, G. J.; Jiang, W.; Guaman, L. P.; Xiao, Y.; Zhang, F., Engineering Escherichia coli to produce branched-chain fatty acids in high percentages. *Metab. Eng.* 2016, *38*, 148-158.

28. Waltermann, M.; Luftmann, H.; Baumeister, D.; Kalscheuer, R.; Steinbuchel, A., Rhodococcus opacus strain PD630 as a new source of high-value single-cell oil? Isolation and characterization of triacylglycerols and other storage lipids. *Microbiology* 2000, *146 (Pt 5)*, 1143-9.

29. Cho, H. U.; Park, J. M., Biodiesel production by various oleaginous microorganisms from organic wastes. *Bioresour Technol* 2018, *256*, 502-508.

30. Haushalter, R. W.; Kim, W.; Chavkin, T. A.; The, L.; Garber, M. E.; Nhan, M.; Adams, P. D.; Petzold, C. J.; Katz, L.; Keasling, J. D., Production of anteiso-branched fatty acids in Escherichia coli; next generation biofuels with improved cold-flow properties. *Metab. Eng.* 2014, *26*, 111-118.

31. Knothe, G.; Dunn, R. O., A Comprehensive Evaluation of the Melting Points of Fatty Acids and Esters Determined by Differential Scanning Calorimetry. *J. Am. Oil Chem. Soc.* 2009, *86* (9), 843-856.

32. Yu, A. Q.; Pratomo Juwono, N. K.; Foo, J. L.; Leong, S. S. J.; Chang, M. W., Metabolic engineering of Saccharomyces cerevisiae for the overproduction of short branched-chain fatty acids. *Metab. Eng.* 2016, *34*, 36-43.

33. Tsitko, I. V.; Zaitsev, G. M.; Lobanok, A. G.; Salkinoja-Salonen, M. S., Effect of aromatic compounds on cellular fatty acid composition of Rhodococcus opacus. *Appl Environ Microbiol* 1999, *65* (2), 853-5.

34. Alvarez, H. M.; Mayer, F.; Fabritius, D.; Steinbuchel, A., Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630. *Arch Microbiol* 1996, *165* (6), 377-86.

35. Eydallin, G.; Ryall, B.; Maharjan, R.; Ferenci, T., The nature of laboratory domestication changes in freshly isolated Escherichia coli strains. *Environ Microbiol* 2014, *16* (3), 813-28.

36. DeLorenzo, D. M.; Rottinghaus, A. G.; Henson, W. R.; Moon, T. S., Molecular Toolkit for Gene Expression Control and Genome Modification in Rhodococcus opacus PD630. *ACS synthetic biology* 2018, *7* (2), 727-738.

37. Durfee, T.; Nelson, R.; Baldwin, S.; Plunkett, G., 3rd; Burland, V.; Mau, B.; Petrosino, J. F.; Qin, X.; Muzny, D. M.; Ayele, M.; Gibbs, R. A.; Csorgo, B.; Posfai, G.; Weinstock, G. M.; Blattner, F. R., The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. *J Bacteriol* 2008, *190* (7), 2597-606.

38. Posfai, G.; Plunkett, G., 3rd; Feher, T.; Frisch, D.; Keil, G. M.; Umenhoffer, K.; Kolisnychenko, V.; Stahl, B.; Sharma, S. S.; de Arruda, M.; Burland, V.; Harcum, S. W.; Blattner, F. R., Emergent properties of reduced-genome Escherichia coli. *Science* 2006, *312* (5776), 1044-6.

39. Ravi, K.; Abdelaziz, O. Y.; Nöbel, M.; García-Hidalgo, J.; Gorwa-Grauslund, M. F.; Hulteberg, C. P.; Lidén, G., Bacterial conversion of depolymerized Kraft lignin. *Biotechnology for Biofuels* 2019, *12* (1), 56.

40. Wei, Z.; Zeng, G.; Huang, F.; Kosa, M.; Huang, D.; Ragauskas, A. J., Bioconversion of oxygen-pretreated Kraft lignin to microbial lipid with oleaginous Rhodococcus opacus DSM 1069. *Green Chemistry* 2015, *17* (5), 2784-2789.

41. Kurosawa, K.; Laser, J.; Sinskey, A. J., Tolerance and adaptive evolution of triacylglycerol-producing Rhodococcus opacus to lignocellulose-derived inhibitors. *Biotechnology for Biofuels* 2015, *8* (76).

Appendix A: Supplementary information for development of chemical and metabolite sensors for *Rhodococcus opacus* PD630

A.1 Supplementary data tables

Supplementary Table A.1. List of genetic elements.

Part	Nucleotide Sequence
Kanamycin Resistance Gene Cassette[1]	tcagaagaactcgtcaagaaggcgatagaaggcgatgcgctgcgaatcgggag cggcgataccgtaaagcacgaggaagcggtcagcccattcgccgccaagctct tcagcaatatcacgggtagccaacgctatgtcctgatagcggtccgccacacc cagccggccacagtcgatgaatccagaaaagcggccattttccaccatgatat tcggcaagcaggcatcgccatgggtcacgacgagatcctcgccgtcgggcatc cgcgccttgagcctggcgaacagttcggctggcgcgagcccctgatgctcttc gtccagatcatcctgatcgacaagaccggcttccatccgagtacgtgctcgct cgatgcgatgtttcgcttggtggtcgaatgggcaggtagccggatcaagcgta tgcagccgccgcattgcatcagccatgatggatactttctcggcaggagcaag gtgagatgacaggagatcctgccccggcacttcgcccaatagcagccagtccc ttcccgcttcagtgacaacgtcgagcacagctgcgcaaggaacgcccgtcgtg gccagccacgatagccgcgctgcctcgtcttggagttcattcagggcaccgga caggtcggtcttgacaaaaagaaccgggcgcccctgcgctgacagccggaaca cggcggcatcagagcagccgattgtctgttgtgcccagtcatagccgaatagc ctctccacccaagcggccggagaacctgcgtgcaatccatcttgttcaatcat gcgaaacgatcctcatcctgtctcttgatcagatcttgatcccctgcgccatc agatccttggcggcaagaaagccatccagtttactttgcagggcttcccaacc ttaccagagggcgccccagctggcaattccggttcgcttgctgtccataaaac cgcccagtctagctatcgccatgtaagcccactgcaagctacctg
pAL5000 – Origin of Replication[1]	ttagaacagcggtggattgtcggcttcgttgtgggccttttgagccgcttcct gttctgccgcacgctctttcctcgcccgatagccgagtcgcttaacggtgtcc agatgcagcccgaaatgtttggccgtttgcggccaagagtggccctcgtcgtc gtgataggcgcggatgcgttcgcggcgtgcagcctgctcggcgagccactcgc tgcgttcctgcgccacgagccggacgacgtggcgttcggatagtccggtgatt cgagcgccttcggcggcggtcacgcgccgcttttgcggacagtcggctgccg gttgtagccgtcgctgtagccgtcgctcatagcaatgcctccatggctgacgc ggactttgcgcgccgcgcaactgtgctcgccgccgtgcgcgctgctgcgccct tccgcgagatggccgactggcgcgcactgagtgtggcctcgtagaccacgatc ccgtccgcccaaatgcgcgacttggttgtgatccaacgccaaatgctgttggc gatggcgcggacctcgctgtccggtagcggtccgggacacacgtcgttgcacg ggaactcggcgtttcgcgcgtggcactcggcatagatcgcgcggccgagtccg tccacgttccgggtcggcaggtagatccgcatgagggcgggacgataggccca caacctgacggaatcgaacagtgcgcaattccgccctagcggcgtcggagccg ctttgtacgtggtctgctgacgccagcgcggcggtggcatgttcgcgccgagc tcggcctcgatgtggctgagtgtgtagagatctgagtggagccattccgtttc ccaggcgatgtggccggggtttttggtcatgaggcctgagtaactgcggtcgc cgtccacggcgcgccgaaggccttcggcgcacgccgccatgtatgcgagcggc

	ttacgccgcgcgtattcggtgcgtggaacaggggcgttgagtgcccacactgc gtgtgcgtggccgttggcgcgattgcccacgatcgcgttgggcagcggatggg acccccgggcgctgagcgctcggagcgctgcgtctggatggtctacgtccacg accagcaggtttgccagcgctgttgggttcgcctcgatgtaccggcggcctag ggccgacgcgcggctttggcggtagatcccctcgagcagatcgtcgcttgcca gcggccagtacggcagccagagctgctcaaattcgtcggcgacgtggctca
pMB1 – Origin of Replication[1]	cgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaa atcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccag gcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgct taccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaat gctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggc tgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaacta tcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagcca ctggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttg aagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgc tctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggca aacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacg cgcagaaaaaaaggatctcaagaagatcctttgatct
Strong Constitutive Promoter + UTR (for Figure 1.1)[2]	tgtgcgggctctaacacgtcctagtatggtaggatgagcaacatttcgacgcc gagagattcgccgcccgaaatgagcacgatccgcatgcttaattaagaaggag atatacat
***araC* Cassette + pBAD Promoter + UTR[3]**	ttatgacaacttgacggctacatcattcacttttttcttcacaaccggcacgga actcgctcgggctggccccggtgcatttttttaaatacccgcgagaaatagagt tgatcgtcaaaaccaacattgcgaccgacggtggcgataggcatccgggtggt gctcaaaagcagcttcgcctggctgatacgttggtcctcgcgccagcttaaga cgctaatccctaactgctggcggaaaagatgtgacagacgcgacggcgacaag caaacatgctgtgcgacgctggcgatatcaaaattgctgtctgccaggtgatc gctgatgtactgacaagcctcgcgtacccgattatccatcggtggatggagcg actcgttaatcgcttccatgcgccgcagtaacaattgctcaagcagatttatc gccagcagctccgaatagcgcccttccccttgcccggcgttaatgatttgccc aaacaggtcgctgaaatgcggctggtgcgcttcatccgggcgaaagaaccccg tattggcaaatattgacggccagttaagccattcatgccagtaggcgcgcgga cgaaagtaaacccactggtgataccattcgcgagcctccggatgacgaccgta gtgatgaatctctcctggcgggaacagcaaaatatcacccggtcggcaaacaa attctcgtccctgatttttcaccaccccctgaccgcgaatggtgagattgaga atataacctttcattcccagcggtcggtcgataaaaaaatcgagataaccgtt ggcctcaatcggcgttaaacccgccaccagatgggcattaaacgagtatcccg gcagcaggggatcattttgcgcttcagccatacttttcatactcccgccattc agagaagaaaccaattgtccatattgcatcagacattgccgtcactgcgtctt ttactggctcttctcgctaaccaaaccggtaaccccgcttattaaaagcattc tgtaacaaagcgggaccaaagccatgacaaaaacgcgtaacaaaagtgtctat aatcacggcagaaaagtccacattgattatttgcacggcgtcacactttgcta tgccatagcatttttatccataagattagcggatcctacctgacgctttttat cgcaactctctactgtttctccatacagcggataaagtagcaaagagaaggag gttagga
LPD06575 Upstream Region[4]	cgcaaagcggatcacctccgcgctcagcggatcgacgcgccgttcccacggtc gtactttcgtccgtacctcatcagcgccccggacaggaagggaggctccgacc
LPD06568 Upstream Region[4]	ccaccagcgattcgtgtacgtattcggactcggggttgacatttcaatgtgac tgcgcgcacagtagtggtgtacgcacagcgtgatggtagtacgcagagcgtac agaatgcaagaccctcggcgcgtgaacacccaccgcgtcgcccaaaactcgat atcagcccattccgctccctcccagcccaaggagaaacag

LPD06699 Upstream Region[4]	agaaatcccttcggcgctactccgggtgtgatgtgcgtcgctgattcggaaaa gtgtgtacgagtcgtgggggctccggaaccgccgattcggtgaaattcgcgcg gtcgagatagacgaaacggcagtaccacgcatctgcgccttcttgacaacgcg gaaacgctatggctaccgtcacattattcaaccagcggacatgcgttcggtat gcgaacaggatctgtccggtaccaacctagacgggacatcgga
LPD06740 Upstream Region[4]	aactcccgaccttagcatcgcctgtgactcacgtcactgcaagtgccggcgac ccccgtccggcgaacccgtttcgcgttccggatcgcgccttcgcaatgcggat cgacggactgtgatgcccgtcatacgttgtgctggtccccattcgcgggggtt accgcacgaatcgtgcagaaaggcggtcaggagaa
LPD03031 Upstream Region[4]	cggacgctcttcgacacgcctaggaaacatgcaggaaacctcacatcggcccg gctgaaaagtgaggatcagttgacgttgcggtcaccgaccgcagcaccgccgc aacaccggttttctacctttgggtctcccacacaacgaggaggcccggg
pTet promoter (optimized for *Mycobacterium* sp.)[5]	tctgaccagggaaaatagccctctgacctggggatttgacttccctatcagtg atagagataatctgggagtccctatcagtgatagagaAGGCGGtatcgat
***R. opacus* optimized pConstitutive + UTR + *tetR* (derived from Rock et al., 2017)[5]**	gtgatagattgcgatgcgtgtgctcggttgaacctctctgtcagatcaggagg acttcgcatgtcgcggttggataagtccaaggtcatcaactcggcgctggagc tgctgaacgaggtcggtatcgagggactgaccacgcgcaagttggcccaaaag ctcggcgtggaacaaccgaccctctactggcacgtgaagaacaagcgcgcgct gctggatgcgttggccatcgagatgctggatcgccatcacacgcacttttgcc ccctggagggcgagagctggcaggattttctgcggaacaacgcgaagtcgttc cggtgcgccctcctgagccatcgggatggcgccaaagtccacttgggcacccg cccgaccgagaagcaatacgagaccctggagaaccagctggccttcctctgtc aacagggtttttcgttggagaacgccctgtatgccctgtccgccgtgggacac ttcaccttgggctgcgtgttggaggatcaggaacatcaggtggcgaaagagga acgggagacccccacgacggattcgatgcccccgttgttgcgccaggcgatcg aactgtttgaccaccagggcgccgaacccgccttttgttcgggctggagctg atcatctgcgggctggaaaagcaactgaagtgcgaaagcggttcgtga
***amiA, amiC, amiD,* and *amiS* + pAcet Promoter + UTR[6]**	gaactccgttgtagtgcttgtggtggcatccgtggcgcggccgcggtaccaga tctttaaatctagataaagaagtgacgcggtctcaagcgtcgagcgtcgtcag cgtgtcgaggatgtcgaagtcgtagccgtcggcgctggcgatgtagacctgct ggtcgaattgactgtcgcgcatacacatcgggccccggggcccgtcgaacccg acatcgtgcgcggatgccatcaggtccggtatctcgggggagtgggcccgctg gaagatggcctcgagcgcaacgagaccctcgtaacaggattcggccatcgcgt tgagcggtggcgcgtcggcgccgtagcgggcgacgtagctgcccatcaggtcc atggcacccgcggtggccagtgaactgaagtacgccgcggcgacatagaggtt ttcggtggagccggcgccgctggccagcagcatgttctcctccatcagcgggc tgaaccgcgccatgcggtcgtgcccgccggcgcgcgcgaactcgcggttgaac aacacggcgtcctggccgacgagcagcatcaacacggcctgcgcccccgacgc gatggccttgcggaccggtgcgcggaaatcgtcggtgccgtacgggacgtaga tctcccgtctgagctcgaggtccagatctcggcagtacgcgcgggcggccgcg gcggaacggcgcggccagatgtagtcatcgccgaccaggcaccaggaccggat gccgaagtggtcgcgcagccaggcgagcgcgggcgcgatctggatctgcggtg tctcgcctgtgcagaacacgcccggtgtgcgttcaccgccctcgtacaacgag gtgtagacgtacgggatgcggtcgcggaccaccggggagatgcggttgcgcac ggccgagatgtgccagccggtcacggcgtcgagaccgtgacctcgcaaccggt cggcgacggtccgggcgacgtcgtcgccgggccgtccgccgtcgagcacctcg atggtgaccttgcggccctgcaggccgcctcggtcgttgacctccttggccgc gagctcggccacggcctcgcacgaaggcgcgaagattcccgctggcccttgaa gcggaatcaccagcccgacgcggaactcaacctcgccgtcctgcactccagat

	caccgtcgatcccgtgtagtctgcgcttcaaagctttctagcagaaataattc attctgaacagaccccgccgtcgacacgaggagacacccaccatggccgccgg acagcagcgccgccccaacctcctgctgccgttggtgcgtctgacccacctcg cggagtcggcgatcgaacgcgtgctcgcggactcgtcgctcaagatcgaggac tggcgcgtgctcgacgagttggccggacggcgcaccgtgcccatgagcgatct cgcgcaggccacgctgatcacgggtccgactctcaccagaaccgtcgatcgcc ttgtgtcgcaagggatcatctaccggactgccgatctgcatgaccgccggcgg gtgctcgtggcgttgaccccgcgggggcggacgctgcgcaaccgcctggtgga cgcggtagccgaggccgagtgtgcggcttttgaatcgtgcgggctggacgtcg accagttgcgcgaactcgtcgacaccacctcgaatttgacttcgtaaccaccc gcgcccggccggcgttcacccttgacttttattttcatctggatatatttcgg gtgaatggaaaggggtgaccatgccgacctacacattccgttgttcccactgc ggtcccttcgatctcacctgcgcgatctccgagcgcgatgcggcggcgacctg tccggagtgccggacgccggcgcgccgggtcttcggttcggtagggctgacga cattcaccgcgggacatcaccgcgcattcgacgcggcgtccgcgagcgccgaa agtcccacggtggtgaagtcgattcccgcaggcgcggaccgcccgcgggcccc gcgccgcaatcccggtctaccgagtctgccgaggtactagcgacatgggtggc gtcgggctcttctacgtgggtgcggtgctcatcatcgacgggctgatgctgct gggccgcatcagcccacgaggcgcaacaccgctgaacttcttcgtcggcggac tgcaggtggtgacgcctacggtgctgatcctgcagtccggcggagacgcggcc gtgatcttcgcggcctccgggctctacctgttcggcttcacctacctgtgggt ggccatcaacaacgtgaccgactgggacggagaaggtctcggatggttctcgc tgttcgtcgcgatcgccgcactcggctactcgtggcacgcgttcaccgccgag gccgacccggcgttcggggtgatctggctgctgtgggcagtgctgtggttcat gctgttcctgctgctcggcctggggcacgacgcactggggcccgccgtcgggt tcgtcgcggtggccgaaggcgtgatcaccgccgccgtgccggccttcctgatc gtgtcgggcaactgggaaaccggcccgctccccgccgcggtcatcgccgtgat cggttttgccgcagttgttctcgcataccccatcgggcgccgtctcgcagcgc cgtcagtcaccaaccctccaccggccgcgctcgcggccaccacccgataagag aaagggagtccacatatgtaacggatccagctgcagaattcgaagcttatcga tgtcgacgtagttacgagatcggcggccgcat
CFP[7]	atgactagcaaaagaagcaaaggtgaagaactgttcactggtgttgttccaat tctggttgaactggatggtgatgttaatggtcacaaattttctgtctctggtg agggtgaaggtgatgcaacctacggtaaactgaccctgaaatttatttgcact actggtaaactgcctgttccgtggccaaccctggtcactactctgacttgggg tgttcaatgctttgctcgttacccagatcacatgaaacagcatgacttttca agtctgccatgccggaaggttatgttcaggaacgtactatcttttcaaagat gacggtaactacaagacccgtgctgaagtcaagtttgaaggtgataccctggt taatcgtatcgagctgaaaggtattgattttaaagaagatggtaacattctgg gtcacaaactggaatacaacgctatttctgataatgtatacatcactgctgac aaacaaaagaatggtatcaaagctaatttcaaaattcgtcacaacattgaaga tggtagcgttcaactggcagaccattatcaacaaaatactccaattggcgatg gccctgtcctgctgccagacaaccattacctgtccacccaatctcgtctgtct aaagatccgaacgaaaagcgcgatcacatggtcctgctggagtttgtaaccgc tgctggtattaccctgggcatggatgaactgtataaataatag
RFP[7]	atggcgagtagcgaagacgttatcaaagagttcatgcgtttcaaagttcgtat ggaaggttccgttaacggtcacgagttcgaaatcgaaggtgaaggtgaaggtc gtccgtacgaaggtacccagaccgctaaactgaaagttaccaaaggtggtccg ctgccgttcgcttgggacatcctgtccccgcagttccagtacggttccaaagc ttacgttaaacacccggctgacatcccggactacctgaaactgtccttcccgg aaggtttcaaatgggaacgtgttatgaacttcgaagacggtggtgttgttacc

	gttacccaggactcctccctgcaagacggtgagttcatctacaaagttaaact gcgtggtaccaacttcccgtccgacggtccggttatgcagaaaaaaaccatgg gttgggaagcttccaccgaacgtatgtacccggaagacggtgctctgaaaggt gaaatcaaaatgcgtctgaaactgaaagacggtggtcactacgacgctgaagt taaaaccacctacatggctaaaaaaccggttcagctgccgggtgcttacaaaa ccgacatcaaactggacatcacctcccacaacgaagactacaccatcgttgaa cagtacgaacgtgctgaaggtcgtcactccaccggtgcttaa
mCherry[8]	atggtgagcaagggcgaggaggataacatggccatcatcaaggagttcatgcg cttcaaggtgcacatggagggctccgtgaacggccacgagttcgagatcgagg gcgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtg accaagggtggccccctgcccttcgcctgggacatcctgtcccctcagttcat gtacggctccaaggcctacgtgaagcaccccgccgacatccccgactacttga agctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcgaggac ggcggcgtggtgaccgtgacccaggactcctccctgcaggacggcgagttcat ctacaaggtgaagctgcgcggcaccaacttcccctccgacggccccgtaatgc agaagaagaccatgggctgggaggcctcctccgagcggatgtaccccgaggac ggcgccctgaagggcgagatcaagcagaggctgaagctgaaggacggcggcca ctacgacgctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgc ccggcgcctacaacgtcaacatcaagttggacatcacctcccacaacgaggac tacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcgg catggacgagctgtacaagtaa
GFP+[9]	atggctagcaaaggagaagaacttttcactggagttgtcccaattcttgttga attagatggtgatgttaatgggcacaaattttctgtcagtggagagggtgaag gtgatgctacatacggaaagcttacccttaaatttatttgcactactggaaaa ctacctgttccatggccaacacttgtcactactttgacctatggtgttcaatg cttttcccgttatccggatcatatgaaacggcatgacttttcaagagtgcca tgcccgaaggttatgtacaggaacgcactatatctttcaaagatgacgggaac tacaagacgcgtgctgaagtcaagtttgaaggtgataccccttgttaatcgtat cgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaac tcgagtacaactataactcacacaatgtatacatcacggcagacaaacaaaag aatggaatcaaagctaacttcaaaattcgccacaacattgaagatggctccgt tcaactagcagaccattatcaacaaaatactccaattggcgatggccctgtcc ttttaccagacaaccattacctgtcgacacaatctgccctttcgaaagatccc aacgaaaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggat tacacatggcatggatgagctctacaaataa
sfGFP[10]	atgcgtaaaggcgaagagctgttcactggtgtcgtccctattctggtggaact ggatggtgatgtcaacggtcataagttttccgtgcgtggcgagggtgaaggtg acgcaactaatggtaaactgacgctgaagttcatctgtactactggtaaactg ccggtaccttggccgactctggtaacgacgctgacttatggtgttcagtgctt tgctcgttatccggaccatatgaagcagcatgacttcttcaagtccgccatgc cggaaggctatgtgcaggaacgcacgatttcctttaaggatgacggcacgtac aaaacgcgtgcggaagtgaaatttgaaggcgataccctggtaaaccgcattga gctgaaaggcattgactttaaagaagacggcaatatcctgggccataagctgg aatacaatttaacagccacaatgtttacatcaccgccgataaacaaaaaaat ggcattaaagcgaattttaaaattcgccacaacgtggaggatggcagcgtgca gctggctgatcactaccagcaaaacactccaatcggtgatggtcctgttctgc tgccagacaatcactatctgagcacgcaaagcgttctgtctaaagatccgaac gagaaacgcgatcatatggttctgctggagttcgtaaccgcagcgggcatcac gcatggtatggatgaactgtacaaatga
EYFP[3]	atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcga gctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagg

	gcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaag ctgcccgtgccctggcccaccctcgtgaccaccttcggctacggcctgcaatg cttcgcccgctaccccgaccacatgaagctgcacgacttcttcaagtccgcca tgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaac tacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcat cgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagc tggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaag aacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgt gcagctcgccgaccactaccagcagaacaccccccatcggcgacggccccgtgc tgctgcccgacaaccactacctgagctaccagtccgccctgagcaaagacccc aacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggat cactctcggcatggacgagctgtacaaggctgcgaattga

Supplementary Table A.2. Summary of plasmids.

Plasmid Name	Functional Insert	Origin of Replication	Antibiotic Resistance	Length (bp)
pDD56	pBAD.EYFP	pAL5000/pMB1	Kanamycin	6013
pDD57	pConstitutive.GFP+	pAL5000/pMB1	Kanamycin	4776
pDD65	Empty Vector	pAL5000/pMB1	Kanamycin	3659
pDD86	pAcet.GFP+	pAL5000/pMB1	Kanamycin	7451
pDD87	pLPD03031.GFP+	pAL5000/pMB1	Kanamycin	4818
pDD89	pLPD02697.GFP+	pAL5000/pMB1	Kanamycin	5026
pDD90	pLPD06364.GFP+	pAL5000/pMB1	Kanamycin	4904
pDD91	pLPD06132.GFP+	pAL5000/pMB1	Kanamycin	5009
pDD93	pLPD06131.GFP+	pAL5000/pMB1	Kanamycin	4780
pDD94	pLPD02061.GFP+	pAL5000/pMB1	Kanamycin	5020
pDD95	pLPD03169.GFP	pAL5000/pMB1	Kanamycin	4821
pDD101	pConstitutive.EYFP	pAL5000/pMB1	Kanamycin	4782
pDD102	pConstitutive.RFP	pAL5000/pMB1	Kanamycin	4736
pDD103	pConstitutive.mCherry	pAL5000/pMB1	Kanamycin	4769
pDD104	pConstitutive.CFP	pAL5000/pMB1	Kanamycin	4790
pDD105	pConstitutive.sfGFP	pAL5000/pMB1	Kanamycin	4775
pDD122	pTet.mCherry + pConstitutive.TetR	pAL5000/pMB1	Kanamycin	6225
pDD131	pTet.mCherry + pConstitutive (optimized).TetR	pAL5000/pMB1	Kanamycin	6225
pRH033	pLPD06740.GFP+	pAL5000/pMB1	Kanamycin	4857
pRH035	pLPD06568.GFP+	pAL5000/pMB1	Kanamycin	4861
pRH036	pLPD06575.GFP+	pAL5000/pMB1	Kanamycin	4768
pRH037	pLPD06699.GFP+	pAL5000/pMB1	Kanamycin	4917

Supplementary Table A.3. Summary of strains.

Strain Name	Genus	Species	Strain	Plasmid Contained
DMD056	*Escherichia*	*coli*	DH10B	Empty Vector pAL5000.pMB1.Kanamycin
DMD057	*Escherichia*	*coli*	DH10B	pBAD.EYFP.pAL5000.pMB1.Kanamycin
DMD066	*Rhodococcus*	*opacus*	PD630	pBAD.EYFP.pAL5000.pMB1.Kanamycin
DMD081	*Rhodococcus*	*opacus*	PD630	Empty Vector pAL5000.pMB1.Kanamycin
DMD082	*Rhodococcus*	*opacus*	PD630	pLPD06575.GFP+.pAL5000.pMB1.Kanamycin
DMD083	*Rhodococcus*	*opacus*	PD630	pLPD06568.GFP+.pAL5000.pMB1.Kanamycin
DMD084	*Rhodococcus*	*opacus*	PD630	pLPD06740.GFP+.pAL5000.pMB1.Kanamycin
DMD086	*Rhodococcus*	*opacus*	PD630	pLPD06699.GFP+.pAL5000.pMB1.Kanamycin
DMD097	*Escherichia*	*coli*	DH10B	Max.Const.GFP+.pAL5000
DMD099	*Escherichia*	*coli*	DH10B	pAcet.GFP+.pAL5000.pMB1.Kanamycin
DMD100	*Rhodococcus*	*opacus*	PD630	pAcet.GFP+.pAL5000.pMB1.Kanamycin
DMD105	*Escherichia*	*coli*	DH10B	pLPD03031.GFP+.pAL5000.pMB1.Kanamycin
DMD107	*Escherichia*	*coli*	DH10B	pLPD02697.GFP+.pAL5000.pMB1.Kanamycin
DMD108	*Escherichia*	*coli*	DH10B	pLPD06364.GFP+.pAL5000.pMB1.Kanamycin
DMD109	*Escherichia*	*coli*	DH10B	pLPD06132.GFP+.pAL5000.pMB1.Kanamycin
DMD111	*Escherichia*	*coli*	DH10B	pLPD06131.GFP+.pAL5000.pMB1.Kanamycin
DMD112	*Escherichia*	*coli*	DH10B	pLPD02061.GFP+.pAL5000.pMB1.Kanamycin
DMD113	*Escherichia*	*coli*	DH10B	pLPD03169.GFP.pAL5000.pMB1.Kanamycin
DMD116	*Rhodococcus*	*opacus*	PD630	pLPD03031.GFP+.pAL5000.pMB1.Kanamycin
DMD118	*Rhodococcus*	*opacus*	PD630	pLPD02697.GFP+.pAL5000.pMB1.Kanamycin
DMD119	*Rhodococcus*	*opacus*	PD630	pLPD06364.GFP+.pAL5000.pMB1.Kanamycin
DMD120	*Rhodococcus*	*opacus*	PD630	pLPD06132.GFP+.pAL5000.pMB1.Kanamycin
DMD122	*Rhodococcus*	*opacus*	PD630	pLPD06131.GFP+.pAL5000.pMB1.Kanamycin
DMD123	*Rhodococcus*	*opacus*	PD630	pLPD02061.GFP+.pAL5000.pMB1.Kanamycin
DMD124	*Rhodococcus*	*opacus*	PD630	pLPD03169.GFP.pAL5000.pMB1.Kanamycin
DMD135	*Escherichia*	*coli*	DH10B	Max.Const.EYFP.pAL5000
DMD136	*Escherichia*	*coli*	DH10B	Max.Const.RFP.pAL5000
DMD137	*Escherichia*	*coli*	DH10B	Max.Const.mCherry.pAL5000
DMD138	*Escherichia*	*coli*	DH10B	Max.Const.CFP.pAL5000
DMD139	*Escherichia*	*coli*	DH10B	Max.Const.sfGFP.pAL5000
DMD146	*Rhodococcus*	*opacus*	PD630	Max.Const.GFP+.pAL5000
DMD147	*Rhodococcus*	*opacus*	PD630	Max.Const.EYFP.pAL5000
DMD148	*Rhodococcus*	*opacus*	PD630	Max.Const.RFP.pAL5000
DMD149	*Rhodococcus*	*opacus*	PD630	Max.Const.mCherry.pAL5000
DMD150	*Rhodococcus*	*opacus*	PD630	Max.Const.CFP.pAL5000
DMD151	*Rhodococcus*	*opacus*	PD630	Max.Const.sfGFP.pAL5000
DMD163	*Escherichia*	*coli*	DH10B	pTet.mCherry + pConstitutive.TetR.pAL5000
DMD170	*Rhodococcus*	*opacus*	PD630	pTet.mCherry + pConstitutive.TetR.pAL5000

DMD174	*Rhodococcus*	*opacus*	PD630	pTet.mCherry + pConstitutive (optimized).TetR.pAL5000
WRH079	*Escherichia*	*coli*	DH10B	pLPD06740.GFP+.pAL5000.pMB1.Kanamycin
WRH081	*Escherichia*	*coli*	DH10B	pLPD06568.GFP+.pAL5000.pMB1.Kanamycin
WRH082	*Escherichia*	*coli*	DH10B	pLPD06575.GFP+.pAL5000.pMB1.Kanamycin
WRH083	*Escherichia*	*coli*	DH10B	pLPD06699.GFP+.pAL5000.pMB1.Kanamycin

Supplementary Table A.4. Fitted Hill equation parameters. The modified Hill equation, described below in Supplementary Methods, was fit to each promoter data set by minimizing the root mean square error (RMSE). Fmax and Fmin represent the maximum and minimum fluorescence, respectively. The Hill coefficient (n), half-maximal constant (K), and RMSE are also reported.

Figure	Strain	Fmax	Fmin	n	K	RMSE
1.2A	pBAD-EYFP-pAL5000	22222 au	279 au	0.869	4.387 mM	986.298
1.2B	pAcet-GFP+-pAL5000	821 au	162 au	0.772	0.249 nM	39.364
1.2C	pTet-mCherry-pAL5000	608 au	0 au	1.126	0.175 ng/mL	2.055
1.3	pLPD03031-GFP+-pAL5000; 0.05 g/L N	0.916	0	11.162	7.134 hr	0.081
1.3	pLPD03031-GFP+-pAL5000; 0.1 g/L N	0.942	0	11.022	9.118 hr	0.043
1.3	pLPD03031-GFP+-pAL5000; 0.2 g/L N	0.963	0	17.653	11.309 hr	0.039
1.3	pLPD03031-GFP+-pAL5000; 0.3 g/L N	0.950	0	20.963	12.845 hr	0.028
1.3	pLPD03031-GFP+-pAL5000; 0.4 g/L N	0.965	0	23.140	14.551 hr	0.019
1.3	pLPD03031-GFP+-pAL5000; 0.5 g/L N	0.980	0	24.414	16.549 hr	0.023
1.3	pLPD03031-GFP+-pAL5000; 0.6 g/L N	0.936	0	21.849	19.765 hr	0.032
1.3	pLPD03031-GFP+-pAL5000; 0.7 g/L N	0.938	0	25.948	21.312 hr	0.036
1.3	pLPD03031-GFP+-pAL5000; 0.8 g/L N	0.987	0	17.003	23.867 hr	0.0315

Supplementary Table A.5. Phenol sensor time points (hr) for Figure 1.4. Based on growth curves (see Supplementary Figure A.7), the time point closest to early stationary phase was identified for each growth condition such that fluorescent values could be compared on an equal growth basis. The time points (hr) identified for each strain and growth condition are reported below. Cultures were grown in minimal media A.

Construct	1 g/L Glucose	1 g/L Glucose + 0.1 g/L Phenol	1 g/L Glucose + 0.3 g/L Phenol	1 g/L Glucose + 0.75 g/L Phenol
Empty Vector	18	20	31.5	43
pLPD06568	16	18	23	40
pLPD06575	18	20	31.5	43
pLPD06699	14	14	20	34
pLPD06740	23	31.5	34	61

Supplementary Table A.6. Phenol and catechol degradation operon homology. *R. opacus* PD630 genes were compared to the *R. erythropolis* CCM2505 genome and proteome using NCBI blastn and blastx (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The nucleotide percent identity based on local alignment (blastn) is reported as the number of identical base pairs, in addition to the expect (E) value. The E value represents the number of expected hits when searching a database (the smaller value, the more significant match). Local alignment scores for the promoter sequences are not reported because the nucleotide sequences are too short. Once the *R. erythropolis* CCM2505 homologs were identified, a global alignment between each pair of gene or promoter sequences was performed using NCBI global alignment (Needleman-Wunsch). The nucleotide percent identity based on global alignment is reported as the number of identical basepairs. The amino acid positive percent identity between each protein pair is reported as the number of identical or physically similar amino acids, in addition to the E value.

R. opacus **PD630 Gene or Promoter**	***R. erythropolis*** **CCM2505 Gene or Promoter**	**Nucleotide Local Alignment Percent Identity**	**Nucleotide Local Alignment E value**	**Nucleotide Global Alignment Percent Identity**	**Amino Acid Positive Percent Identity**	**Amino Acid E value**
LPD06566	*catC*	224/284 (79%)	2e-62	224/284 (79%)	83/93 (89%)	2e-55
LPD06567	*catB*	892/1121 (80%)	0.0	892/1122 (80%)	350/373 (93%)	0.0
LPD06568	*catA*	638/844 (76%)	5e-165	640/943 (68%)	222/263 (84%)	2e-141

pLPD06568	*catA promoter*	N/A	N/A	130/220 (59%)	N/A	N/A
LPD06569	*catR*	513/674 (76%)	1e-135	567/774 (73%)	209/254 (82%)	6e-136
LPD06574	*pheR*	544/786 (69%)	3e-80	648/1008 (64%)	227/302 (75%)	1e-139
LPD06575	*pheA2*	412/530 (78%)	2e-118	432/570 (76%)	163/184 (88%)	5e-113
pLPD06575	*pheA2 promoter*	N/A	N/A	61/112 (54%)	N/A	N/A
LPD06576	*pheA1*	1316/1559 (84%)	0.0	1359/1632 (83%)	517/543 (95%)	0.0
LPD06739	*pheR*	486/742 (65%)	2e-43	608/957 (64%)	224/311 (72%)	9e-120
LPD06740	*pheA2*	422/540 (78%)	6e-90	434/578 (75%)	157/186 (84%)	3e-107
pLPD06740	*pheA2 promoter*	N/A	N/A	107/214 (50%)	N/A	N/A
LPD06741	*pheA1*	1308/1564 (84%)	0.0	1340/1632 (82%)	508/540 (94%)	0.0

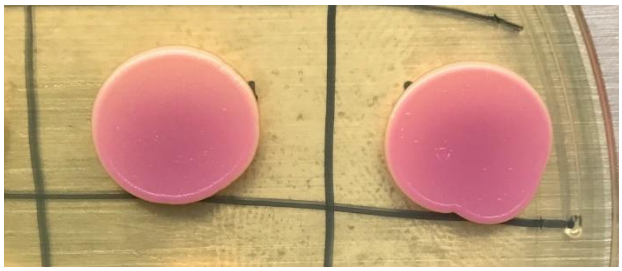
Supplementary Table A.7. Aromatic sensor time points (hr) for Figure 1.5. Based on growth curves (see Supplementary Figure A.7), the time point closest to early stationary phase was identified for each growth condition such that fluorescent values could be compared on an equal growth basis. The time points (hr) identified for each strain and growth condition are reported below. Cultures were induced with 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) in minimal media B, in addition to 1 g/L glucose (GLU).

Strain	**GLU**	**PHE**	**PCA**	**BEN**	**HBA**	**VAN**	**GUA**
Empty Vector	7.5	11.5	7.5	8.5	8.5	10.5	10.5
pLPD06568	7.5	10.5	7.5	8.5	8.5	10.5	9.5
pLPD06575	7.5	11.5	7.5	8.5	8.5	10.5	9.5
pLPD06699	7.5	11.5	7.5	8.5	8.5	10.5	10.5
pLPD06740	7.5	29	8.5	8.5	8.5	10.5	10.5

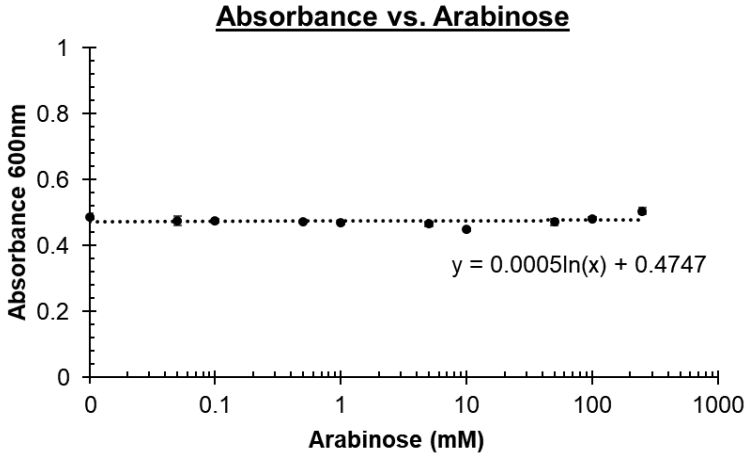
Supplementary Table A.8. Fluorescent reporter excitation and emission wavelengths.

Reporter	Excitation (nm)	Emission (nm)
CFP	435	483
EYFP	485	528
GFP+	488	530
sfGFP	488	530
RFP	535	620
mCherry	535	620

A.2 Supplementary figures

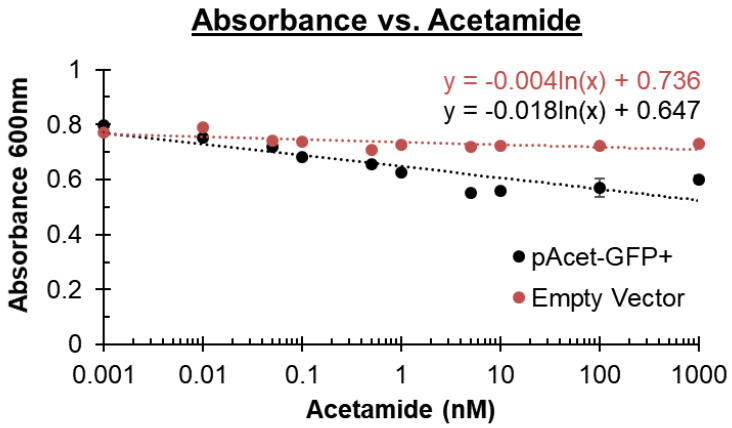


Supplementary Figure A.1. Expression of mCherry in *R. opacus*. Cells constitutively expressing mCherry (DMD149) are shown under white ambient light.

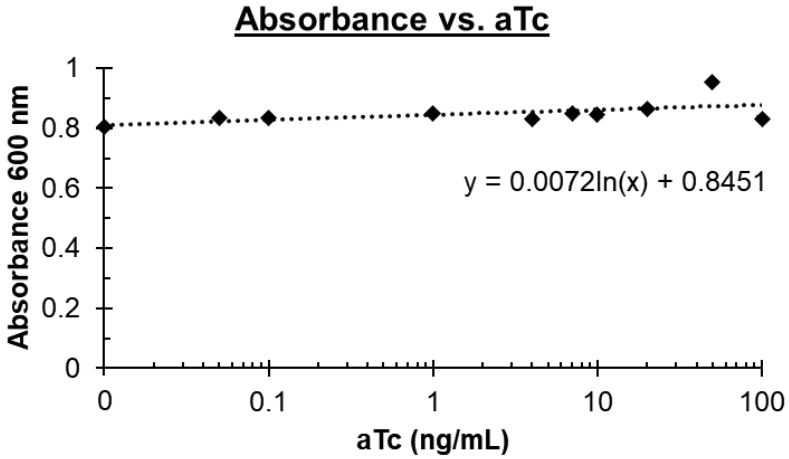


Supplementary Figure A.2. Effect of arabinose on *R. opacus* growth. The absorbance at 600 nm was plotted versus the arabinose inducer concentration (0 to 250 mM). Values represent the average of three replicates, and error bars represent one standard deviation. The dashed line

represents a logarithmic regression, with the equation of the line reported. The addition of arabinose had no significant effect (|logarithmic coefficient| < 0.01) on *R. opacus* growth (strain containing pBAD-EYFP-pAL5000; DMD066). Cells were grown in minimal media A with 1 g/L glucose.

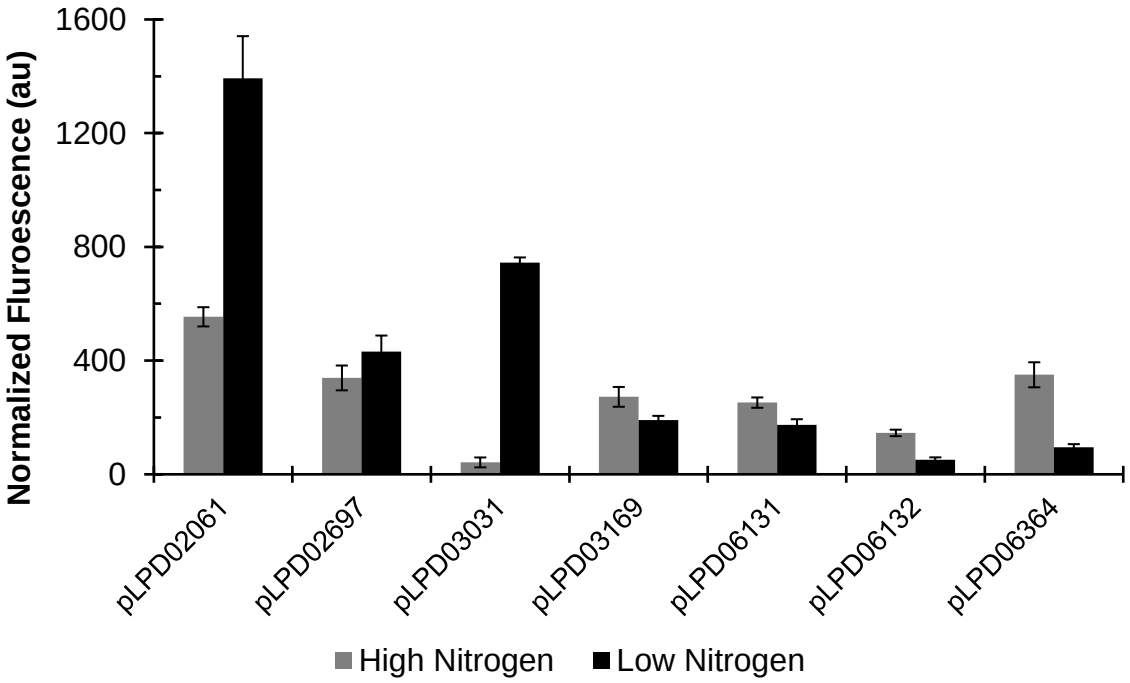


Supplementary Figure A.3. Effect of acetamide on *R. opacus* growth. The absorbance at 600 nm was plotted versus the acetamide inducer concentration (0 to 1000 nM). Values represent the average of three replicates, and error bars represent one standard deviation. The dashed lines represent logarithmic regressions, with the equations of the line reported. The addition of acetamide demonstrated no significant effect (|logarithmic coefficient | < 0.01) on the empty vector strain's (DMD081) growth but led to a reduction in the pAcet-GFP+-pAL5000 strain's (DMD100) growth. The pAcet promoter requires the heterologous expression of four proteins, which in addition to GFP+ adds burdens to the cell. Cells were grown in minimal media A with 2 g/L glucose.

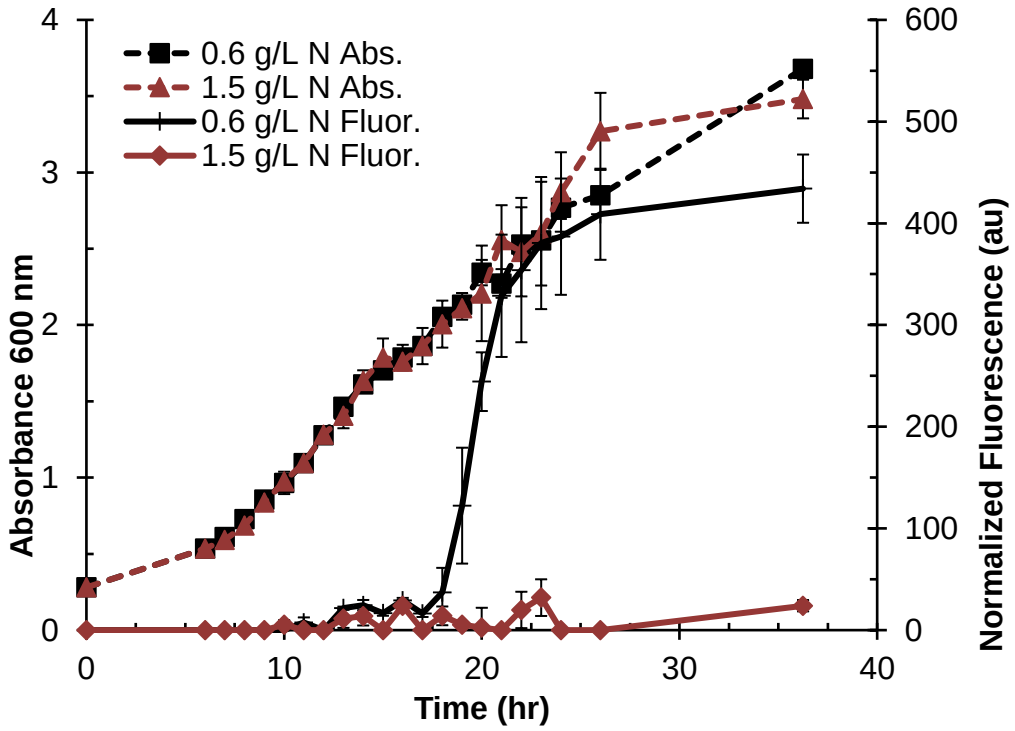


Supplementary Figure A.4. Effect of aTc on *R. opacus* growth. The absorbance at 600 nm was plotted versus the aTc inducer concentration (0 to 100 ng/mL). Values represent the average of

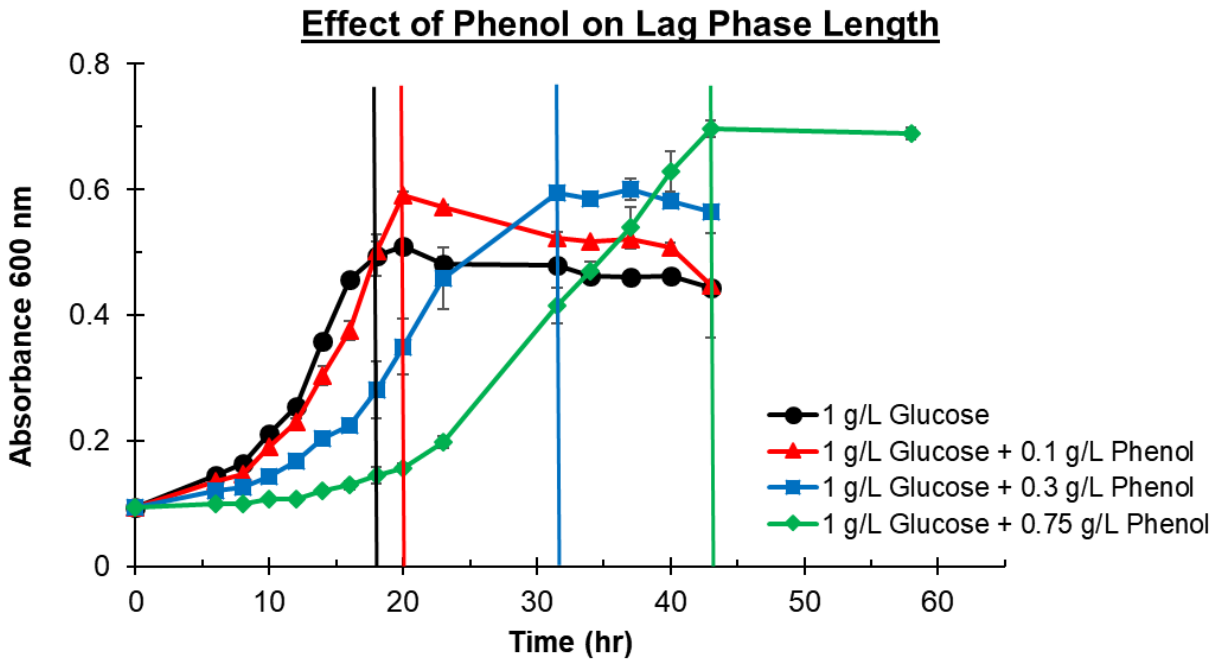
three replicates, and error bars represent one standard deviation. The dashed line represents a logarithmic regression, with the equation of the line reported. The addition of aTc had no significant effect (|logarithmic coefficient| < 0.01) on *R. opacus* growth (strain containing pTet-mCherry-pAL5000; DMD174). Cells were grown in minimal media B with 2 g/L glucose.



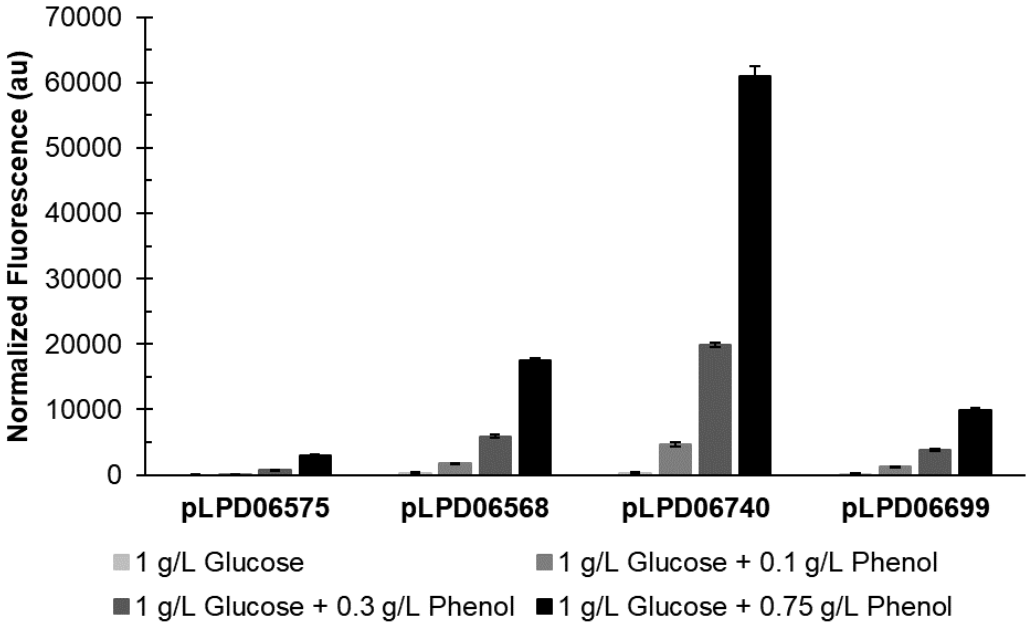
Supplementary Figure A.5. Nitrogen sensor candidate promoters. The promoters of seven genes (LPD02061, LPD02697, LPD03031, LPD03169, LPD06131, LPD06132, and LPD06364) that were found to be highly transcribed under low nitrogen conditions (using previously published RNA-Seq data) were placed in front of GFP+.[4] Cultures were grown in minimal media A with 2 g/L glucose and either 0.05 g/L (low nitrogen) or 1.0 g/L (high nitrogen) ammonium sulfate. Fluorescence was measured after 48 hrs. pLPD03031 was found to be induced ~18-fold in fluorescence between low and high nitrogen conditions. Values represent the average of three replicates, and error bars represent one standard deviation.



Supplementary Figure A.6. Activation of the nitrogen sensor is not dependent on growth phase. The nitrogen sensor strain (containing pLPD3031) was grown in minimal media B with 10 g/L glucose and either 0.6 g/L or 1.5 g/L ammonium sulfate. The absorbance at 600 nm (proxy for growth) was plotted on the primary y-axis (left). The normalized fluorescence was plotted on the secondary y-axis (right). Values represent the average of three replicates, and error bars represent one standard deviation. Lines are drawn to aid eyes and linearly connect data points.



Supplementary Figure A.7. Growth curves of cells expressing GFP+ under the control of pLPD06575. An example of how time points were chosen to compare fluorescence values between different phenol concentrations. *R. opacus* was grown in minimal media A with 1 g/L glucose and either 0, 0.1, 0.3, or 0.75 g/L phenol for Figure 1.4. Increasing concentrations of phenol led to longer lag phases, preventing fluorescent values from being compared on an equal growth basis at a single time point. Fluorescent values were compared at early stationary phase (i.e., when the absorbance at 600 nm, a proxy for cell density, first reaches a plateau). The vertical lines, color coded to match data sets, designate which time point was chosen. This process was performed for each phenol inducible promoters; additional plots are not shown, but time points are reported in Supplementary Table 5. Additionally, this process was applied to Figure 1.5 (cells grown in minimal media B), as each aromatic compound tested (PHE, PCA, BEN, HBA, VAN, and GUA) produced a lag phase of different length (time points listed in Supplementary Table 7). Values represent the average of three replicates, and error bars represent one standard deviation. Lines are drawn to aid eyes and linearly connect data points.



Supplementary Figure A.8. Normalized phenol promoter fluorescence. The upstream regions of LPD06575, LPD06568, LPD06740, and LPD06699 were cloned in front of GFP+. The strains were grown with 1 g/L glucose and either 0, 0.1, 0.3, or 0.75 g/L phenol. This figure represents the fluorescence data shown in Figure 1.4 prior to normalization by the maximum fluorescence value produced by each promoter. Values represent the average of three replicates grown in minimal media A, and error bars represent one standard deviation.

A.3 Supplementary materials and methods

A.3.1 Minimal media recipe A - Used for the fluorescent reporter, pBAD, pAcet, and phenol induction experiments.

Recipe adapted from Schlegel et al., 1961.[11]

Recipe for 1 L of media: Media were filter-sterilized (0.22 μM filter) after all components were mixed and pH was adjusted to 7.0 using 6M HCl.

Chemical/Component	**Amount per 1 L of media**
10x salt solution	100 mL
Hoagland salt solution	1 mL
$CaCl_2$	1 mL of 20 mg/mL stock
$(NH_4)_5[Fe(C_6H_4O_7)_2]$	1 mL of 1.2 mg/mL stock
$NaHCO_3$	0.5 g

10x salt solution:

Chemical	**Amount per 1 L**
Na_2HPO_4	35.7 g
KH_2PO_4	15 g
$MgSO_4*7H_2O$	2 g

Hoagland salt solution:

Chemical	**mg/L**
$(NH_4)_3PO_4$	115.03
H_3BO_3	2.86
$Ca(NO_3)_2$	656.4
$CuSO_4*5H_2O$	0.08
$Na_2EDTA*2H_2O$	33.534
$FeSO_4*7H_2O$	25.02
$MgSO_4$	240.76
$MnCl_2*4H_2O$	1.81
MoO_3	0.016
KNO_3	606.6
$ZnSO_4*7H_2O$	0.22

- Carbon and nitrogen sources were added as filter-sterilized solution.

A.3.2 Minimal media recipe B - Used for the pTet, nitrogen sensor, and aromatic sensor induction experiments.

Recipe adapted from: Chartrain *et al.*, 1998[12].

Recipe for 1 L of media: Media were filter-sterilized (0.22 μM filter) after all components were mixed and pH was adjusted to 7.0 using 6M HCl.

Chemical/Component	**Amount per 1 L of media**
$MgSO_4*7H_2O$	1.0 g
$CaCl_2*2H_2O$	0.015 g (1 mL of 15 g/L stock)
Sodium bicarbonate	0.5 g
Trace Element Solution	1 mL of stock
Stock Solution A	1 mL of stock
Phosphate Buffer (1M)	20 mL of stock

Trace element stock solution:

Chemical	**g/L**
$FeSO_4*7H_2O$	0.5
$ZnSO_4*7H_2O$	0.4
$MnSO_4*H_2O$	0.02
H_3BO_3	0.015
$NiCl_2*6H_2O$	0.01
EDTA	0.25
$CoCl_2*6H_2O$	0.05
$CuCl_2*2H_2O$	0.005

Stock A solution:

Chemical	**g/L**
$NaMoO_4$	2.0
FeNa*EDTA	5.0

Phosphate buffer stock (1 M):

Chemical	**g/L**
K_2HPO_4	113.0
KH_2PO_4	47.0

- Carbon and nitrogen sources were added as filter-sterilized solution.

A.3.3 Electrocompetent cell preparation and transformation protocol for *R. opacus* PD630[13]

2 mL of tryptic soy broth (TSB) in a 50 mL glass tube was inoculated with *R. opacus* PD630 and incubated overnight at 30°C and 250 rpm. Overnight cultures were diluted to an OD_{600} of 0.05 in 100 mL of TSB media, supplemented with 8.5 g/L glycine and 10 g/L sucrose, in a 250 mL baffled shake flask and incubated at 30°C and 250 rpm for ~16 hrs (until OD_{600} reaches ~0.5). Cells were centrifuged in 50 mL conicals (Falcon) at 3500 relative centrifugal force (rcf) and washed twice with sterile deionized water. Cells were re-suspended in 10% glycerol at an OD_{600} of ~10-15. Aliquots of 200 μL were frozen at -80°C. Electrocompetent cells were incubated at 45°C for 5 min, electroporated at 2500 mV with ~500 ng plasmid DNA (Bulldog Bio 2 mm cuvette; time constant 5-6 ms), diluted with 600 μL TSB media, and incubated again at 45°C for 5 min. Transformed cells were then transferred to a 50 mL glass tube and incubated at 30°C and 250 rpm for 3 hrs. Cells were plated on TSB agar plates with kanamycin (50 μg/mL). Colonies appeared after 2-3 days.

A.3.4 Hill equation fitting

The Hill equation was modified, as described in Immethun et al. (2017), to model the promoter transfer function.[3] The model was fit to the experimentally determined fluorescence data such that the RMSE was minimized (Microsoft Excel Solver GRG Nonlinear). Fitted values are listed in Supplementary Table 4.

For pBAD, pAcet, and pTet:

$$F = \frac{(F_{max} - F_{min}) * Inducer^n}{Inducer^n + K^n} + F_{min}$$

$F = Calculated\ fluorescence$
$F_{max} = Maximum\ fluorescence$
$F_{min} = Minimum\ fluorescence$
$K = Half\ maximal\ constant$
$n = Hill\ coefficient$
$Inducer = respective\ promoter\ inducer\ concentration$

For the nitrogen sensor:

$$F = \frac{(F_{max} - F_{min}) * t^n}{t^n + K^n} + F_{min}$$

$t = Time\ (hr)$

Root mean square error (RMSE):

$$\sqrt{\frac{\sum_{i=1}^{n}(F - F_{obs})^2}{n-2}}$$

$F_{obs} = Observed\ fluorescence$
n = Number of data points

A.4 References

1. Kurosawa, K., Wewetzer, S.J. & Sinskey, A.J. Engineering xylose metabolism in triacylglycerol-producing Rhodococcus opacus for lignocellulosic fuel production. Biotechnology for biofuels 6, 134 (2013).

2. Siegl, T., Tokovenko, B., Myronovskyi, M. & Luzhetskyy, A. Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. Metabolic engineering 19, 98-106 (2013).

3. Immethun, C.M. et al. Physical, chemical, and metabolic state sensors expand the synthetic biology toolbox for Synechocystis sp. PCC 6803. Biotechnol. Bioeng. (2017). DOI: 10.1002/bit.26275

4. Yoneda, A. et al. Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630. Nucleic Acids Res 44, 2240-2254 (2016).

5. Rock, J.M. et al. Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. Nature microbiology 2, 16274 (2017).

6. van Kessel, J.C. & Hatfull, G.F. Recombineering in Mycobacterium tuberculosis. Nat. Methods 4, 147-152 (2007).

7. Hoynes-O'Connor, A. & Moon, T.S. Development of Design Rules for Reliable Antisense RNA Behavior in E. coli. ACS synthetic biology 5, 1441-1454 (2016).

8. Lee, Y.J., Hoynes-O'Connor, A., Leong, M.C. & Moon, T.S. Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system. Nucleic Acids Res 44, 2462-2473 (2016).

9. Ehrt, S. et al. Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. Nucleic acids research 33, e21 (2005).

10. Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. & Waldo, G.S. Engineering and characterization of a superfolder green fluorescent protein. Nature biotechnology 24, 79-88 (2006).

11. Schlegel, H.G., Kaltwasser, H. & Gottschalk, G. Ein Submersverfahren zur Kultur wasserstoffoxydierender Bakterien: Wachstumsphysiologische Untersuchungen. Archiv für Mikrobiologie 38, 209-222 (1961).

12. Chartrain, M. et al. Bioconversion of indene to cis (1S,2R) indandiol and trans (1R,2R) indandiol by Rhodococcus species. J. Ferment. Bioeng. 86, 550-558 (1998).

13. Kalscheuer, R., Arenskotter, M. & Steinbuchel, A. Establishment of a gene transfer system for Rhodococcus opacus PD630 based on electroporation and its application for recombinant biosynthesis of poly(3-hydroxyalkanoic acids). Applied microbiology and biotechnology 52, 508-515 (1999).

Appendix B: Construction of genetic logic gates based on the T7 RNA polymerase expression system in *Rhodococcus opacus* PD630

B.1 Supplementary tables

Supplementary Table B.1. List of genetic parts.

Part	Nucleotide Sequence
kanR[1]	atgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggct atgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggg gcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactccaagacgaggca gcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactg aagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcacct tgctcctgccgagaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccg gctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatggaag ccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgtt cgccaggctcaaggcgcggatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgc ttgccgaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtg tggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagagcttggcggcga atgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgccttc tatcgccttcttgacgagttcttctga
hygR[2]	atgagccggatgaccggcaccacctggcggtccgcgatggacggcacgaccgaccggaacgcgc tgctcgccctggcccgcgaactcggccgggtgctcggccggctgcacagggtgccgctgaccgg gaacaccgtgctcacccccattccgaggtcttcccggaactgctgcgggaacgccgcgcggcg accgtcgaggaccaccgcgggtggggctacctctcgccccggctgctggaccgcctggaggact ggctgccggacgtggacacgctgctggccggccgcgaaccccggttcgtccacggcgacctgca cgggaccaacatcttcgtggacctggccgcgaccgaggtcaccgggatcgtcgacttcaccgac gtctatgcgggagactcccgctacagcctggtgcaactgcatctcaacgccttccggggcgacc gcgagatcctggccgcgctgctcgacggggcgcagtggaagcggaccgaggacttcgcccgcga actgctcgccttcaccttcctgcacgacttcgaggtgttcgaggagaccccgctggatctctcc ggcttcaccgatccggaggaactggcgcagttcctctgggggccgccggacaccgcccccggcg cctga
R. opacus origin of replication on pAL5000 (short variant; S)[1]	ttagaacagcggtggattgtcggcttcgttgtgggccttttgagccgcttcctgttctgccgca cgctctttcctcgcccgatagccgagtcgcttaacggtgtccagatgcagcccgaaatgtttgg ccgtttgcggccaagagtggcccctcgtcgtcgtgataggcgcggatgcgttcgcggcgtgcagc ctgctcggcgagccactcgctgcgttcctgcgccacgagccggacgacgtggcgttcggatagt ccggtgattcgagcgccttcggcggcggtcacgcgccgcttttgcggacagtcggctgccggt tgtagccgtcgctgtagccgtcgctcatagcaatgcctccatggctgacgcggactttgcgcgc cgcgcaactgtgctcgccgccgtgcgcgctgctgcgcccttccgcgagatggccgactggcgcg cactgagtgtggcctcgtagaccacgatcccgtccgcccaaatgcgcgacttggttgtgatcca acgccaaatgctgttggcgatggcgcggacctcgctgtccggtagcggtccgggacacacgtcg ttgcacgggaactcggcgtttcgcgcgtggcactcggcatagatcgcgcggccgagtccgtcca cgttccgggtcggcaggtagatccgcatgagggcgggacgataggcccacaacctgacggaatc gaacagtgcgcaattccgccctagcggcgtcggagccgctttgtacgtggtctgctgacgccag cgcggcggtggcatgttcgcgccgagctcggcctcgatgtggctgagtgtgtagagatctgagt ggagccattccgtttcccaggcgatgtggccggggtttttggtcatgaggcctgagtaactgcg gtcgccgtccacggcgcgccgaaggccttcggcgcacgccgccatgtatgcgagcggcttacgc cgcgcgtattcggtgcgtggaacaggggcgttgagtgcccacactgcgtgtgcgtggccgttgg cgcgattgcccacgatcgcgttgggcagcggatgggacccccgggcgctgagcgctcggagcgc tgcgtctggatggtctacgtccacgaccagcaggtttgccagcgctgttgggttcgcctcgatg taccggcggcctagggccgacgcgcggctttggcggtagatcccctcgagcagatcgtcgcttg ccagcggccagtacggcagccagagctgctcaaattcgtcggcgacgtggctca
E. coli origin of replication pMB1[3]	cgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctca agtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccc tcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggg aagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctcc

	aagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatc gtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggat tagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctac actagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttg gtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttttgtttgcaagcagca gattacgcgcagaaaaaaaggatctcaagaagatcctttgatct
pConstitutive from *Streptomyces lividans* TK24 + RBS[4]	tgtgcgggctctaacacgtcctagtatggtaggatgagcaacatttcgacgccgagagattcgc cgcccgaaatgagcacgatccgcatgcttaattaagaaggagatatacat
GFP+[2, 5]	atggctagcaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtg atgttaatgggcacaaattttctgtcagtggagagggtgaaggtgatgctacatacggaaagct taccctaaatttatttgcactactggaaaactacctgttccatggccaacacttgtcactact ttgacctatggtgttcaatgcttttcccgttatccggatcatatgaaacggcatgacttttca agagtgccatgcccgaaggttatgtacaggaacgcactatatctttcaaagatgacgggaacta caagacgcgtgctgaagtcaagtttgaaggtgataccttgttaatcgtatcgagttaaaaggt attgattttaaagaagatggaaacattctcggacacaaactcgagtacaactataactcacaca atgtatacatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaa cattgaagatggctccgttcaactagcagaccattatcaacaaaatactccaattggcgatggc cctgtccttttaccagacaaccattacctgtcgacacaatctgccctttcgaaagatcccaacg aaaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacacatggcatgga tgagctctacaaataa
eGFP[6]	atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcg acgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagct gaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccacc ctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttca agtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaacta caagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggc atcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccaca acgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaa catcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggc cccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacg agaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatgga cgagctgtacaagtaa
mCherry[7]	atggtgagcaagggcgaggaggataacatggccatcatcaaggagttcatgcgcttcaaggtgc acatggagggctccgtgaacggccacgagttcgagatcgagggcgagggcgagggccgcccctа cgagggcacccagaccgccaagctgaaggtgaccaagggtggccccctgcccttcgcctgggac atcctgtcccctcagttcatgtacggctccaaggcctacgtgaagcaccccgccgacatccccg actacttgaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcgaggacgg cggcgtggtgaccgtgacccaggactcctccctgcaggacggcgagttcatctacaaggtgaag ctgcgcggcaccaacttcccctccgacggccccgtaatgcagaagaagaccatgggctgggagg cctcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatcaagcagaggctgaa gctgaaggacggcggccactacgacgctgaggtcaagaccacctacaaggccaagaagcccgtg cagctgcccggcgcctacaacgtcaacatcaagttggacatcacctcccacaacgaggactaca ccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggcatggacgagctgta caagtaa
pLPD06575 (pPhenol) + RBS[5]	cgcaaagcggatcacctccgcgctcagcggatcgacgcgccgttcccacggtcgtactttcgtc cgtacctcatcagcgccccggacaggaagggaggctccgacc
pTet promoter + RBS[8]	tctgaccagggaaaatagccctctgacctggggatttgacttccctatcagtgatagagataat ctgggagtccctatcagtgatagagaaggcggtatcgat
R. opacus optimized *tetR* expression cassette[5]	gtgatagattgcgatgcgtgtgctcggttgaacctctctgtcagatcaggaggacttcgcatgt cgcggttggataagtccaaggtcatcaactcggcgctggagctgctgaacgaggtcggtatcga gggactgaccacgcgcaagttggcccaaaagctcggcgtggaacaaccgaccctctactggcac gtgaagaacaagcgcgcgctgctggatgcgttggccatcgagatgctggatcgccatcacacgc acttttgccccctggagggcgagagctggcaggattttctgcggaacaacgcgaagtcgttccg gtgcgccctcctgagccatcgggatggcgccaaagtccacttgggcacccgcccgaccgagaag caatacgagaccctggagaaccagctggccttcctctgtcaacagggtttttcgttggagaacg ccctgtatgccctgtccgccgtgggacacttcaccttgggctgcgtgttggaggatcaggaaca tcaggtggcgaaagaggaacgggagacccccacgacggattcgatgcccccgttgttgcgccag gcgatcgaactgtttgaccaccagggcgccgaacccgccttttttgttcgggctggagctgatca tctgcgggctggaaaagcaactgaagtgcgaaagcggttcgtga
T7 RNA polymerase[6]	atgaacacgattaacatcgctaagaacgacttctctgacatcgaactggctgctatcccgttca acactctggctgaccattacggtgagcgtttagctcgcgaacagttggcccttgagcatgagtc

	ttacgagatgggtgaagcacgcttccgcaagatgtttgagcgtcaacttaaagctggtgaggtt gcggataacgctgccgccaagcctctcatcactaccctactccctaagatgattgcacgcatca acgactggtttgaggaagtgaaagctaagcgcggcaagcgcccgacagccttccagttcctgca agaaatcaagccggaagccgtagcgtacatcaccattaagaccactctggcttgcctaaccagt gctgacaatacaaccgttcaggctgtagcaagcgcaatcggtcgggccattgaggacgaggctc gcttcggtcgtatccgtgaccttgaagctaagcacttcaagaaaaacgttgaggaacaactcaa caagcgcgtagggcacgtctacaagaaagcatttatgcaagttgtcgaggctgacatgctctct aagggtctactcggtggcgaggcgtggtcttcgtggcataaggaagactctattcatgtaggag tacgctgcatcgagatgctcattgagtcaaccggaatggttagcttacaccgccaaaatgctgg cgtagtaggtcaagactctgagactatcgaactcgcacctgaatacgctgaggctatcgcaacc cgtgcaggtgcgctggctggcatctctccgatgttccaaccttgcgtagttcctcctaagccgt ggactggcattactggtggtggctattgggctaacggtcgtcgtcctctggcgctggtgcgtac tcacagtaagaaagcactgatgcgctacgaagacgtttacatgcctgaggtgtacaaagcgatt aacattgcgcaaaacaccgcatggaaaatcaacaagaaagtcctagcggtcgccaacgtaatca ccaagtggaagcattgtccggtcgaggacatccctgcgattgagcgtgaagaactcccgatgaa accggaagacatcgacatgaatcctgaggctctcaccgcgtggaaacgtgctgccgctgctgtg taccgcaaggacaaggctcgcaagtctcgccgtatcagccttgagttcatgcttgagcaagcca ataagtttgctaaccataaggccatctggttcccttacaacatggactggcgcggtcgtgttta cgctgtgtcaatgttcaacccgcaaggtaacgatatgaccaaaggactgcttacgctggcgaaa ggtaaaccaatcggtaaggaaggttactactggctgaaaatccacggtgcaaactgtgcgggtg tcgataaggttccgttccctgagcgcatcaagttcattgaggaaaaccacgagaacatcatggc ttgcgctaagtctccactggagaacacttggtgggctgagcaagattctccgttctgcttcctt gcgttctgctttgagtacgctggggtacagcaccacggcctgagctataactgctcccttccgc tggcgtttgacgggtcttgctctggcatccagcacttctccgcgatgctccgagatgaggtagg tggtcgcgcggttaacttgcttcctagtgaaaccgttcaggacatctacgggattgttgctaag aaagtcaacgagattctacaagcagacgcaatcaatgggaccgataacgaagtagttaccgtga ccgatgagaacactggtgaaatctctgagaaagtcaagctgggcactaaggcactggctggtca atggctggcttacggtgttactcgcagtgtgactaagcgttcagtcatgacgctggcttacggg tccaaagagttcggcttccgtcaacaagtgctggaagataccattcagccagctattgattccg gcaagggtctgatgttcactcagccgaatcaggctgctggatacatggctaagctgatttggga atctgtgagcgtgacggtggtagctgcggttgaagcaatgaactggcttaagtctgctgctaag ctgctggctgctgaggtcaaagataagaagactggagagattcttcgcaagcgttgcgctgtgc attgggtaactcctgatggtttccctgtgtggcaggaatacaagaagcctattcagacgcgctt gaacctgatgttcctcggtcagttccgcttacagcctaccattaacaccaacaaagatagcgag attgatgcacacaaacaggagtctggtatcgctcctaactttgtacacagccaagacggtagcc accttcgtaagactgtagtgtgggcacacgagaagtacggaatcgaatcttttgcactgattca cgactccttcggtaccattccggctgacgctgcgaacctgttcaaagcagtgcgcgaaactatg gttgacacatatgagtcttgtgatgtactggctgatttctacgaccagttcgctgaccagttgc acgagtctcaattggacaaaatgccagcacttccggctaaaggtaacttgaacctccgtgacat cttagagtcggacttcgcgttcgcgtaa
pT7[6]	taatacgactcactataggg
lacI expression cassette[6]	gacaccatcgaatggcgcaaaacctttcgcggtatggcatgatagcgcccggaagagagtcaat tcagggtggtgaatgtgaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttа tcagaccgtttcccgcgtggtgaaccaggccagccacgtttctgcgaaaacgcgggaaaaagtg gaagcggcgatggcggagctgaattacattcccaaccgcgtggcacaacaactggcgggcaaac agtcgttgctgattggcgttgccacctccagtctggccctgcacgcgccgtcgcgaaattgtcgc ggcgattaaatctcgcgccgatcaactgggtgccagcgtggtggtgtcgatggtagaacgaagc ggcgtcgaagcctgtaaagcggcggtgcacaatcttctcgcgcaacgcgtcagtgggctgatca ttaactatccgctggatgaccaggatgccattgctgtggaagctgcctgcactaatgttccggc gttatttcttgatgtctctgaccagacacccatcaacagtattattttctcccatgaagacggt acgcgactgggcgtggagcatctggtcgcattgggtcaccagcaaatcgcgctgttagcgggcc cattaagttctgtctcggcgcgtctgcgtctggctggctggcataaatatctcactcgcaatca aattcagccgatagcggaacgggaaggcgactggagtgccatgtccggttttcaacaaaccatg caaatgctgaatgagggcatcgttcccactgcgatgctggttgccaacgatcagatggcgctgg gcgcaatgcgcgccattaccgagtccgggctgcgcgttggtgcggatatctcggtagtgggata cgacgataccgaagacagctcatgttatatcccgccgttaaccaccatcaaacaggattttcgc ctgctggggcaaaccagcgtggaccgcttgctgcaactctctcagggccaggcggtgaagggca atcagctgttgcccgtctcactggtgaaaagaaaaaccaccctggcgcccaatacgcaaaccgc ctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagc gggcagtgagcgcaacgcaattaatgtaagttagctcactcattaggcaccgggatctcgaccg atgcccttgagagccttcaacccagtcagctccttccggt
lacO[6]	ggaattgtgagcggataacaattcccctctaga
$lacO_{min}$[9]	ttgtgagcggataacaa

RBS-1[4]	ggtaccaagcttattggcactagtcgagcaacggaggtacggac
RBS-2[10]	catttcgacgccgagagattcgccgcccgaaatgagcacgatccgcatgcttaattaagaaggagatatacat
pRS23365 (rRNA)	gtccgatgagataagctggaaaagttgccccggagcggctggcgagtgtgccagaagtgatggtgtgtgcgtgttctttgagaactcaacagtgtgtcgatgaatgtcagtgccaaatgtttttggtactccgcatcacggatgatcaactggcccttgtggtttggttgtgaggtgtggtgtgggtatttgctggctcttctccttcttccgtcggaggggggtcagtgtttgagaatgctagtttgagtttttgattgctagtgatttgactcgatgtctatgactgattagggacctttttggggtttcgaaatctagagtc
pRS28745 (Trp tRNA)	gtttagggtggattgggtttgttccggggctttgaagtacactgagatttctgggattactcagcgctgcgcgctgccctacggtgctgtacggccggcaagggtggcgcgcgtgtcatgtaatcacagagaggtcggtcaagagttggttcgaggctagtttcggaccaggtctcccgaccaa
pRS28770 (Thr tRNA)	cacacagaaccccggtgagcgacagctcaccggggttctgtcgttcgtggacgagacggggcgccccatgtatgcggtgatgccgaccggtcgtcgggtccatcgaatttcgcgtcccagtcgatgtcctggacggccttcttcgcgaccgcctgggcggacttcggcgtgatcctggagctcgatcgcgggtctcgagcatggtcgccggccgaactgtcgggtgcaccggcatgtgaaagccgctgaccaggcaatttggttctggacgcgaactctgtgtaatctcgtcaaggcttcagcgcgcagcgatgtgcgagcgaagca
pRS00085 (Ile tRNA)	acgggatgggcgtgtgggaccggctgaacaacgcgttcacggatatcgtcgcggagtcgagctctgacgggctcgtgacggccgggcaggtcttcggatacgccgccgtcatcggtctcgcgaacatggtgttgttcaccgcattggcgacgatcggagccttcatctacaacctgtgttccgacctcgtcggggcgtcgaagtgacgctcgccgaccgggactgaacgcggaaaacgacgagggctgagctgggaaaagagcgtctcgccggaccggttttggtagttgccggtccggtggggtaatcttccgactcggttcagggaactgcgcgagatgagtttccagacttcag
pRS06620 (Glu tRNA)	gccccccttcgtctagcggcctaggacgccgccctttcaaggcggtagcgcgggttcgaatccccgtaggggggtacgtttgaccggtgtcgagcatgcggtatggttcgacgg
pRS02495 (Ser tRNA)	cttcgactggttcgacagccacctgcgataaagcgaccatcgttgcgctctgaccagccgttttgcctttgtggtggaaaggtgtgtaatctttcgcaccggcggtccgaagggatcggaaccaccccagggaggcgtgccagagcggccgaatgggactcactgctaatgagttgtccctttttacgaggga ccggaggttcaaatcctctcgcctccgcc
pamiM[11]	agaacggggcttgtggccgtccctgtcgtgtggtaaaatgtccacaa
pnhM[11]	acatctacacattgacatccgttccgatgtgatgtaaaaattgtcacg
pcs[11]	ccctcctactgttgtcggcgccccgaacgtggttcactgacatcggtc
pcs/pamiM	ccctcctactgttgtgggcgccccgaacgtggtaaaatgacatcggtc
pCory1[12]	ttgacattaatttgaatctgtgttataatggttc
pCory2[12]	ttgacaaataaggttgtatgtgctataatggacc
pLac$_{RO1}$ + RBS-2	acatctacacattgacatccgttccgatgtgatgtaaaaattgtcacg ggaattgtgagcggataacaattcccctctagacatttcgacgccgagagattcgccgcccgaaatgagcacgatccgcatgcttaattaagaaggagatatacat
pLac$_{RO2}$ + RBS-2	acatctacacattgacatccgttccgatgtgatgtaaaaattgtcacgcatttgtgagcggataacaaagccgcccgaaatgagcacgatccgcatgcttaattaagaaggagatatacat
pLac$_{RO3}$	ccctcctactgttgtgagcggataacaagtggtaaaatgacatcggtc
pLPD06764 (pHBA) + RBS	catagctcgtctttcactcagtgaaagaggggtgttgcgatcacacgtcgggcagcacagtgactggtgtcacgaactacgcgaggagtctgctcat
pLPD00563 (pVAN) + RBS	gacgaaacccttccgacattcggaaaccgagtgtggtggatcacctccagtgccgcacagtggtggagacgacgctgacgcctcaccgagtccggcgtcacgaataacgcagcagcccagcatattccggaggcactcat
pLPD06580 (pBEN) + RBS	ggagaagttcacgccagtcgtgtgaagtcggcccgcgtgcgcacgccgaggccctccgagcctcggcgtacgcatgccgaacacgtcctagtcagtactcggtaaatgcctagtgtgatgcagcgcacaggctgcaagagtggaatccattgccggacccgggcgccctcgcagggcgcaagccccccggagatcttggaggatcccat
pLPD06578 (pGUA) + RBS	ctgcctcactgaagaatcgttccgacctcattgaaaccgtgtcgtcacgcccagtcgaggataactgagaccgcgatcacaagtagtcgcgcagatggtccagagaccgcgcagagtggctagaccaccctcctgcgacggcctacgttcttcctcacccgctcacacaacgaaagttgtttccgagcagttgctttgaatgaggagattcag

Supplementary Table B.2. Summary of plasmids.

Plasmid Name	Functional Insert	Backbone	Antibiotic Resistance	Length (bp)	Source
pTara:500	pBAD-*T7 RNAP*	p15a	Chloramphenicol	8022	Shis et al. 2013[6]
pET28:GFP	pT7(LacO)-*eGFP* + *lacI*	pBR322	Kanamycin	5993	Shis et al. 2013[6]
pRH036	pLPD06575-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4861	DeLorenzo et al. 2017[5]
pDD57	pConstitutive.RBS-2-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4776	DeLorenzo et al. 2018[10]
pDD65	Empty backbone	pAL5000(S) + pMB1	Kanamycin	3659	DeLorenzo et al. 2018[10]
pDD131	pTet-*mCherry* + *tetR*	pAL5000(S) + pMB1	Kanamycin	6225	DeLorenzo et al. 2017[5]
pDD132	ROCI-3 (LPD3763/3764) Integrative vector	Integrative into *R. opacus* + p15a	Hygromycin B	4238	DeLorenzo et al. 2018[10]
pDD159	pConstitutive.RBS-1-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4749	This study; derived from pDD57
pDD166	pRS23365-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4988	This study; derived from pDD57
pDD167	pRS28745-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4751	This study; derived from pDD57
pDD168	pRS28770-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4912	This study; derived from pDD57
pDD169	pRS00085-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4928	This study; derived from pDD57
pDD170	pRS02495-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4791	This study; derived from pDD57
pDD172	pRS06620-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4802	This study; derived from pDD57
pDD181	pT7(LacO)-*eGFP* + *lacI*	pAL5000(S) + pMB1	Kanamycin	6373	This study; derived from pET28:GFP and pDD65
pDD183	pTet-*T7 RNAP*.ROCI3	Integrative into *R. opacus* + p15a	Hygromycin B	6996	This study; derived from pTara:500, pDD131, and pDD132
pDD184	pLPD06575-*T7 RNAP*.ROCI3 or pPhenol-*T7 RNAP*.ROCI3	Integrative into *R. opacus* + p15a	Hygromycin B	5846	This study; derived from pTara:500, pRH036 and pDD132
pDD188	pamiM-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4478	This study; derived from pDD57
pDD189	pnhM-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4479	This study; derived from pDD57
pDD190	pcs-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4479	This study; derived from pDD57
pDD191	pcs/pamiM-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4479	This study; derived from pDD57
pDD192	pCory1-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4465	This study; derived from pDD57
pDD193	pCory2-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4465	This study; derived from pDD57
pDD203	$pLac_{RO1}$-*GFP*+	pAL5000(S) + pMB1	Kanamycin	6003	This study; derived from pDD57
pDD219	pT7-*lacI* + $pLac_{RO1}$-*GFP*+	pAL5000(S) + pMB1	Kanamycin	5926	This study; derived pDD203
pDD229	$pLac_{RO3}$-*GFP*+	pAL5000(S) + pMB1	Kanamycin	5970	This study; derived from pDD57 and pET28:GFP
pDD232	$pLac_{RO2}$-*GFP*+	pAL5000(S) + pMB1	Kanamycin	5970	This study; derived from pDD57

pDD272	pT7(LacO)-*tetR* + *lacI* + pTet-*mCherry*	pAL5000(S) + pMB1	Kanamycin	7027	This study; derived from pDD131 and pET28:GFP
pDD299	pT7-*eGFP*	pAL5000(S) + pMB1	Kanamycin	6373	This study; derived from pDD181
pDD300	pLPD06764-*GFP*+ or pHBA-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4761	This study; derived from pDD57
pDD301	pLPD00563-*GFP*+ or pVAN-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4803	This study; derived from pDD57
pDD302	pLPD06580-*GFP*+ or pBEN-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4872	This study; derived from pDD57
pDD303	pLPD06578-*GFP*+ or pGUA-*GFP*+	pAL5000(S) + pMB1	Kanamycin	4877	This study; derived from pDD57

Supplementary Table B.3. Summary of strains.

Strain Name	Genus	Species	Strain	Plasmid Contained	Relevant Figure
WRH082	*Escherichia*	*coli*	DH10B	pRH036	N/A
DMD056	*Escherichia*	*coli*	DH10B	pDD65	N/A
DMD057	*Escherichia*	*coli*	DH10B	pDD57	N/A
DMD081	*Rhodococcus*	*opacus*	PD630	pDD65	4, 6, S1, S2, and S6
DMD146	*Rhodococcus*	*opacus*	PD630	pDD57	S1 and S2
DMD199	*Escherichia*	*coli*	DH10B	pDD132	N/A
DMD220	*Escherichia*	*coli*	DH10B	pDD131	N/A
DMD234	*Escherichia*	*coli*	DH10B	pDD159	N/A
DMD237	*Rhodococcus*	*opacus*	PD630	pDD159	S1 and S2
DMD241	*Escherichia*	*coli*	DH10B	pDD167	N/A
DMD242	*Escherichia*	*coli*	DH10B	pDD168	N/A
DMD243	*Escherichia*	*coli*	DH10B	pDD169	N/A
DMD246	*Rhodococcus*	*opacus*	PD630	pDD167	S1
DMD247	*Rhodococcus*	*opacus*	PD630	pDD168	S1
DMD248	*Rhodococcus*	*opacus*	PD630	pDD169	S1
DMD252	*Escherichia*	*coli*	DH10B	pDD170	N/A
DMD253	*Escherichia*	*coli*	DH10B	pDD172	N/A
DMD255	*Escherichia*	*coli*	DH10B	pDD166	N/A
DMD259	*Escherichia*	*coli*	DH10B	pTara:500	N/A
DMD260	*Escherichia*	*coli*	DH10B	pET28:GFP	N/A
DMD264	*Escherichia*	*coli*	DH10B	pDD183	N/A
DMD265	*Escherichia*	*coli*	DH10B	pDD184	N/A
DMD268	*Escherichia*	*coli*	DH10B	pDD181	N/A
DMD276	*Rhodococcus*	*opacus*	PD630 ΦpTet-*T7 RNAP*.ROCI3	pDD183	N/A

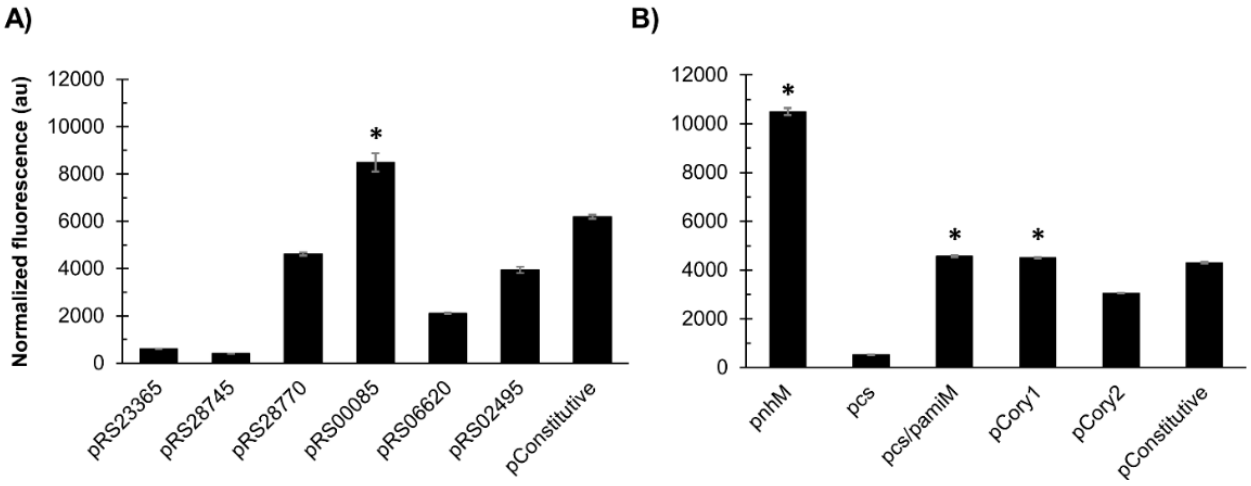
DMD277	*Rhodococcus*	*opacus*	PD630 ΦpLPD06575-*T7 RNAP*.ROCI3	pDD184	N/A
DMD293	*Rhodococcus*	*opacus*	PD630 ΦpTet-*T7 RNAP*.ROCI3	pDD181	2
DMD294	*Rhodococcus*	*opacus*	PD630 ΦpLPD06575-*T7 RNAP*.ROCI3	pDD181	2
DMD287	*Escherichia*	*coli*	DH10B	pDD188	N/A
DMD288	*Escherichia*	*coli*	DH10B	pDD189	N/A
DMD289	*Escherichia*	*coli*	DH10B	pDD190	N/A
DMD290	*Escherichia*	*coli*	DH10B	pDD191	N/A
DMD291	*Escherichia*	*coli*	DH10B	pDD192	N/A
DMD292	*Escherichia*	*coli*	DH10B	pDD193	N/A
DMD299	*Rhodococcus*	*opacus*	PD630	pDD189	S1
DMD300	*Rhodococcus*	*opacus*	PD630	pDD190	S1
DMD301	*Rhodococcus*	*opacus*	PD630	pDD191	S1
DMD302	*Rhodococcus*	*opacus*	PD630	pDD192	S1
DMD303	*Rhodococcus*	*opacus*	PD630	pDD193	S1
DMD306	*Rhodococcus*	*opacus*	PD630 ΦpTet-*T7 RNAP*.ROCI3	pDD65	1 and 2
DMD307	*Rhodococcus*	*opacus*	PD630 ΦpLPD06575-*T7 RNAP*.ROCI3	pDD65	1, 2, 3, 5, S3, and S5
DMD309	*Escherichia*	*coli*	DH10B	pDD203	N/A
DMD319	*Rhodococcus*	*opacus*	PD630	pDD203	4
DMD326	*Escherichia*	*coli*	DH10B	pDD219	N/A
DMD356	*Escherichia*	*coli*	DH10B	pDD229	N/A
DMD359	*Escherichia*	*coli*	DH10B	pDD232	N/A
DMD362	*Rhodococcus*	*opacus*	PD630	pDD229	4
DMD363	*Rhodococcus*	*opacus*	PD630	pDD232	4 and S6
DMD364	*Rhodococcus*	*opacus*	PD630 ΦpLPD06575-*T7 RNAP*.ROCI3	pDD219	5 and S5
DMD375	*Escherichia*	*coli*	DH10B	pDD272	N/A
DMD377	*Rhodococcus*	*opacus*	PD630 ΦpLPD06575-*T7 RNAP*.ROCI3	pDD272	3 and S3
DMD492	*Rhodococcus*	*opacus*	PD630	pDD166	S1
DMD493	*Rhodococcus*	*opacus*	PD630	pDD170	S1
DMD494	*Rhodococcus*	*opacus*	PD630	pDD172	S1
DMD495	*Escherichia*	*coli*	DH10B	pDD299	N/A
DMD496	*Escherichia*	*coli*	DH10B	pDD300	N/A
DMD497	*Escherichia*	*coli*	DH10B	pDD301	N/A
DMD498	*Escherichia*	*coli*	DH10B	pDD302	N/A

DMD499	*Escherichia*	*coli*	DH10B	pDD303	N/A
DMD500	*Rhodococcus*	*opacus*	PD630 ΦpTet-*T7 RNAP*.ROCI3	pDD299	1
DMD501	*Rhodococcus*	*opacus*	PD630 ΦpLPD06575-*T7 RNAP*.ROCI3	pDD299	1
DMD502	*Rhodococcus*	*opacus*	PD630	pDD300	6 and S6
DMD503	*Rhodococcus*	*opacus*	PD630	pDD301	6 and S6
DMD504	*Rhodococcus*	*opacus*	PD630	pDD302	6 and S6
DMD505	*Rhodococcus*	*opacus*	PD630	pDD303	6 and S6

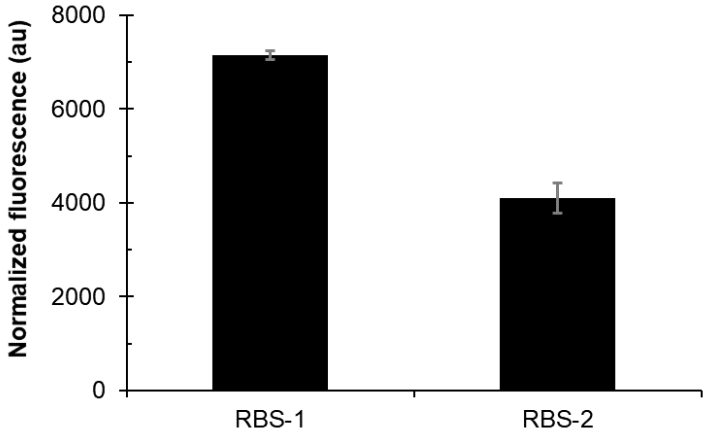
Supplementary Table B.4. Fitted Hill equation parameters for Figures 3.1, 3.4, 3.6, and Supplementary Figure B.5. The modified Hill equation, described below in Supplementary Methods, was fit to each set of normalized fluorescence (au) data by minimizing the root mean square error (RMSE). F_{max} and F_{min} represent the fitted maximum and minimum normalized fluorescence (au), respectively. The fitted Hill coefficient (n), half-maximal constant (K), and RMSE are also reported. Data fitting was done to provide a guide to the eye, rather than to obtain biologically meaningful parameters (especially for Figure 6 in which sensor output saturation was not reached, giving unrealistic parameter values).

Figure	**Strain**	**F_{max}**	**F_{min}**	**n**	**K**	**RMSE**
3.1B	pPhenol-*T7 RNAP* + pT7-*eGFP*	15334 au	827 au	1.49	0.07 g/L	802
3.1C	pTet-*T7 RNAP* + pT7-*eGFP*	4025 au	880 au	2.69	0.03 ng/mL	116
3.4A	pLac$_{RO1}$.*GFP*+	1247 au	64 au	1.92	0.05 mM	23.2
3.4B	pLac$_{RO2}$.*GFP*+	2143 au	161 au	1.65	0.02 mM	45.3
3.4C	pLac$_{RO3}$.*GFP*+	933 au	80 au	1.55	0.03 mM	13.3
3.6A	pHBA-*GFP*+	380 au	26 au	0.57	2.45 g/L	2.9
3.6B	pVAN-*GFP*+	326 au	9 au	1.69	0.26 g/L	9.4
3.6C	pBEN-*GFP*+	956 au	27 au	1.45	2.80 g/L	15.1
3.6D	pGUA-*GFP*+	3950 au	7 au	1.24	1.49 g/L	66.7
B.S5	pPhenol-*T7 RNAP* + pT7-*lacI* + pLac$_{RO1}$.*GFP*+	776 au	0 au	1.63	0.10 g/L	21.7

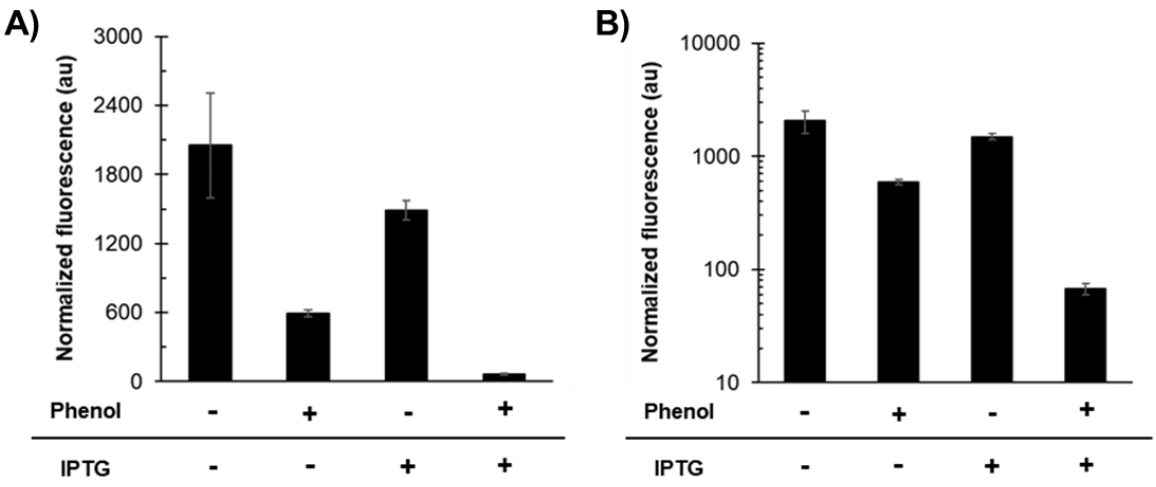
B.2 Supplementary figures



Supplementary Figure B.1. Screen for strong constitutive promoters in *R. opacus*. **A)** Six tRNA and rRNA promoters from the *R. opacus* genome and a previously published constitutive promoter (pConstitutive) were placed upstream of RBS-1 and *GFP+*.[10] **B)** Two native promoters from the *Rhodococcus ruber* genome (pnhM and pcs), a novel hybrid of two *R. ruber* promoters (pcs/pamiM), two synthetic promoters designed for *C. glutamicum* (pCory1 and pCory2), and pConstitutive were placed upstream of RBS-2 and *GFP+*.[11, 12] Values are averages of three replicates, and error bars represent one standard deviation. A star (*) indicates that expression from pConstitutive is significantly lower than expression from the compared promoter ($p < 0.05$, unpaired t-test with equal variance). See Supplementary Table 1 for genetic part sequences. The strengths of RBS-1 and RBS-2 are compared in Supplementary Figure B.2.



Supplementary Figure B.2. Comparison of RBS sequences in *R. opacus*. Two RBS sequences from literature (see Supplementary Table B.1 for sequences) were placed downstream of pConstitutive and upstream of *GFP+*. Bars are averages of three replicates and error bars represent one standard deviation.



Supplementary Figure B.3. NAND logic gate in *R. opacus* using 0.5 g/L phenol. Normalized fluorescence of the NAND genetic circuit in response to phenol and IPTG inducers, as appropriate, at concentrations of 0.5 g/L and 1 mM, respectively. aTc was not provided during the main experimental culture (see Methods B for details). **A)** Linear y-axis. **B)** Log y-axis. An initial OD_{600} of 0.1 was used, and cultures were grown in minimal media B with 10 g/L glucose as a carbon source. The decrease in normalized fluorescence between uninduced [0 0] (the highest ON) and phenol- and IPTG-induced [1 1] (OFF) was 31-fold. The decrease in normalized fluorescence between when either phenol [1 0] or IPTG [0 1] was individually provided and when both were provided [1 1] was 9- and 22-fold, respectively. Values are averages of three replicates, and error bars represent one standard deviation.

A)

LacO GGAATTGTGAGCGGATAACAATTCCCCTCTAGA

pnhM TGTAAAAATTGTCACG CATTTCGACGCCGAGAGATTCGC

$pLac_{RO1}$ TGTAAAAATTGTCACG GGAATTGTGAGCGGATAACAATTCCCCTCTAGA CATTTCGACGCCGAGAGATTCGC

B)

$LacO_{Min}$ TTGTGAGCGGATAACAA

pnhM TGTAAAAATTGTCACG CATTTCGACGCCGAGAGATTCGC

$pLac_{RO2}$ TGTAAAAATTGTCACG CATT**TGTGA**GC**G**GA**T**A**ACAAA**GC

C)

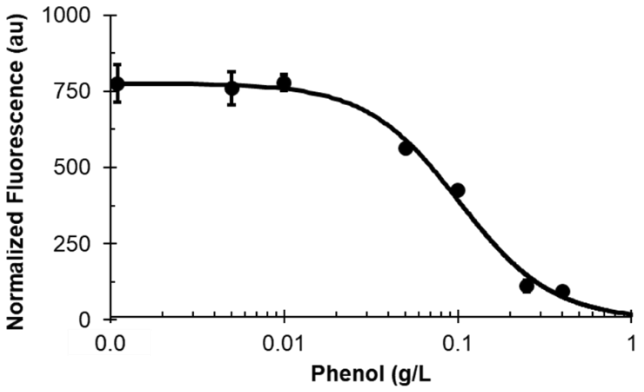
$LacO_{Min}$ TTGTGAGCGGATAACAA

pamiM/pcs ACTGTTGTGGGCGCCCCGAACGTGGTAAAATGAC

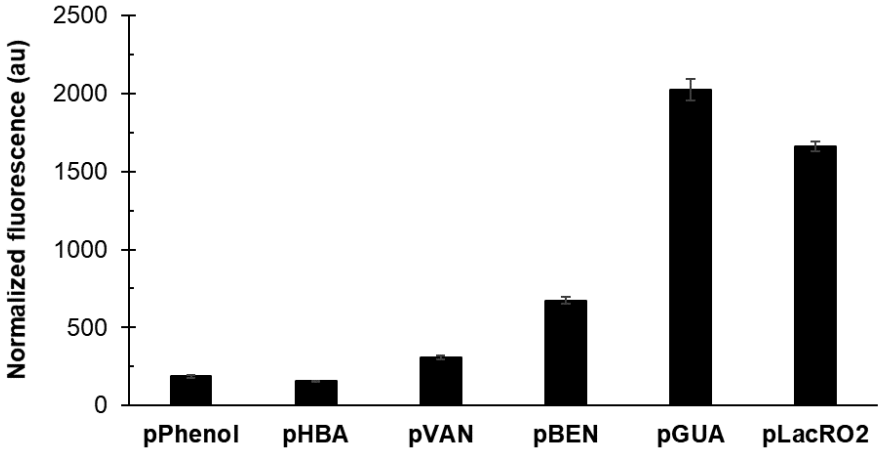
$pLac_{RO3}$ ACTGTTGTG**A**GCG**GATAAC**A**A**GTGGTAAAATGAC

Supplementary Figure B.4. Bolded nucleotides (nt) represent changes from the original constitutive promoter or RBS sequence. Underlined nucleotides represent the core lacO

recognition site, referred to as $LacO_{Min}$. Blue nucleotides represent -10 promoter sites. Red nucleotides represent -35 promoter sites. **A)** To create $pLac_{RO1}$, an insulated 33 nt lacO site was placed between pnhM and RBS-2 as shown in the figure. **B)** To create Lac_{RO2}, a 17 nt minimal lacO site ($LacO_{Min}$) was overlaid with RBS-2 by changing several nucleotides as shown in the figure. **C)** To create $pLac_{RO3}$, $LacO_{Min}$ was overlaid with the -35 site (and its immediate downstream region) of pcs/pamiM by changing several nucleotides as shown in the figure. Only the relevant sections of each promoter or RBS are shown due to space limitations. See Supplementary Table B.1 for complete part sequences.



Supplementary Figure B.5. Normalized fluorescence of the IMPLY circuit in response to phenol in the absence of IPTG. Values are averages of three replicates at 0, 0.005, 0.01, 0.05, 0.1, 0.25, and 0.4 g/L phenol, and error bars represent one standard deviation. The solid black line represents a fitted curve (see Supplementary Methods B and Supplementary Table B.4).



Supplementary Figure B.6. Comparison of inducible systems using GFP+. The normalized fluorescence of pPhenol, pHBA, pVAN, pBEN, pGUA, and $pLac_{RO2}$ was measured when cultures were grown on 1 g/L glucose and with inducer concentration of 0.6 g/L PHE, 1 g/L HBA, 1.5 g/L VAN, 5 g/L BEN, 1.5 g/L GUA, or 1 mM IPTG as appropriate. Values are averages of three replicates, and error bars represent one standard deviation.

B.3 Hill equation fitting

The Hill equation was adapted from DeLorenzo et al. (2017), to fit a line to the normalized fluorescence data.[5] The model was fit to the experimentally collected data, such that the root mean square error (RMSE) was minimized. Fitted values are listed in Supplementary Table B.4.

For Figures 3.1, 3.4, and 3.6:

$$F = \frac{(F_{max} - F_{min}) * Inducer^n}{Inducer^n + K^n} + F_{min}$$

$F = Calculated\ normalized\ fluorescence$
$F_{max} = Maximum\ normalized\ fluorescence$
$F_{min} = Minimum\ normalized\ fluorescence$
$K = Half\ maximal\ constant$
$n = Hill\ coefficient$
Inducer = phenol (g/L), aTc (ng/mL), IPTG (mM), 4-hydroxybenzoic acid (g/L), vanillic acid (g/L), sodium benzoate (g/L), or guaiacol (g/L)

Root mean square error (RMSE):

$$\sqrt{\frac{\sum_{i=1}^{n}(F - F_{obs})^2}{n - 2}}$$

$F_{obs} = Observed\ normalized\ fluorescence$
$n = Number\ of\ data\ points$

For Supplementary Figure B.5:

$$F = \frac{(F_{max} - F_{min}) * K^n}{Inducer^n + K^n} + F_{min}$$

$F = Calculated\ normalized\ fluorescence$
$F_{max} = Maximum\ normalized\ fluorescence$
$F_{min} = Minimum\ normalized\ fluorescence$
$K = Half\ maximal\ constant$
$n = Hill\ coefficient$
$Inducer = phenol\ (g/L)$

Root mean square error (RMSE):

$$\sqrt{\frac{\sum_{i=1}^{n}(F - F_{obs})^2}{n - 2}}$$

$F_{obs} = Observed\ normalized\ fluorescence$
$n = Number\ of\ data\ points$

B.4 References

1. Xiong, X.; Wang, X.; Chen, S., Engineering of a xylose metabolic pathway in Rhodococcus strains. Appl Environ Microbiol 2012, 78 (16), 5483-91.

2. Ehrt, S.; Guo, X. V.; Hickey, C. M.; Ryou, M.; Monteleone, M.; Riley, L. W.; Schnappinger, D., Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. Nucleic Acids Res 2005, 33 (2), e21.

3. Kurosawa, K.; Wewetzer, S. J.; Sinskey, A. J., Engineering xylose metabolism in triacylglycerol-producing Rhodococcus opacus for lignocellulosic fuel production. Biotechnology for Biofuels 2013, 6 (1), 134.

4. Siegl, T.; Tokovenko, B.; Myronovskyi, M.; Luzhetskyy, A., Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. Metab. Eng. 2013, 19, 98-106.

5. DeLorenzo, D. M.; Henson, W. R.; Moon, T. S., Development of Chemical and Metabolite Sensors for Rhodococcus opacus PD630. ACS synthetic biology 2017, 6 (10), 1973-1978.

6. Shis, D. L.; Bennett, M. R., Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. Proc Natl Acad Sci U S A 2013, 110 (13), 5028-33.

7. Lee, Y. J.; Hoynes-O'Connor, A.; Leong, M. C.; Moon, T. S., Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system. Nucleic Acids Res 2016, 44 (5), 2462-73.

8. Rock, J. M.; Hopkins, F. F.; Chavez, A.; Diallo, M.; Chase, M. R.; Gerrick, E. R.; Pritchard, J. R.; Church, G. M.; Rubin, E. J.; Sassetti, C. M.; Schnappinger, D.; Fortune, S. M., Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. Nat Microbiol 2017, 2, 16274.

9. Bahl, C. P.; Wu, R.; Stawinsky, J.; Narang, S. A., Minimal length of the lactose operator sequence for the specific recognition by the lactose repressor. Proc Natl Acad Sci U S A 1977, 74 (3), 966-70.

10. DeLorenzo, D. M.; Rottinghaus, A. G.; Henson, W. R.; Moon, T. S., Molecular Toolkit for Gene Expression Control and Genome Modification in Rhodococcus opacus PD630. ACS synthetic biology 2018, 7 (2), 727-738.

11. Jiao, S.; Yu, H.; Shen, Z., Core element characterization of Rhodococcus promoters and development of a promoter-RBS mini-pool with different activity levels for efficient gene expression. N Biotechnol 2018, 44, 41-49.

12. Rytter, J. V.; Helmark, S.; Chen, J.; Lezyk, M. J.; Solem, C.; Jensen, P. R., Synthetic promoter libraries for Corynebacterium glutamicum. Appl. Microbiol. Biotechnol. 2014, 98 (6), 2617-23.

Appendix C: Supporting information for molecular toolkit for gene expression control and genome modification in *Rhodococcus opacus* PD630

C.1 Supplementary tables

Supplementary Table C.1. Constitutive promoter library. Promoters were generated by saturation mutagenesis of -35 (4 of 6 base pairs) and -10 (6 base pairs) sites of a strong constitutive promoter from *Streptomyces lividans* TK24, as annotated in Siegl et al. (2013), driving GFP+.[1] Promoters are listed in the order of weakest to strongest based on normalized fluorescence. The mutagenized sites are italicized and underlined. Mutations differing from the original sequence are in bold. A gap in the sequence signifies deletion. See Figure 4.1 for normalized fluorescence produced by each promoter.

Promoter Name	Promoter Nucleotide Sequence
1	tgtgcg***ca****c****a***ctaacacgtcctagtatgg***c****a****atca***gagcaa
2	tgtgcg***g****aag***ctaacacgtcctagtatgg***atatg****t***gagcaa
3	tgtgcg***caag***ctaacacgtcctagtatgg***a****a****taga***gagcaa
4	tgtgcg***tt****c****a***ctaacacgtcctagtatgg***cgt****g****ga***gagcaa
5	tgtgcg***atgc***ctaacacgtcctagtatgg***a****ag****atc***gagcaa
6	tgtgcg*ggct*ctaacacgtcctagtatgg*t****ga****g****tg***gagcaa
7	tgtgcg***tag****t***ctaacacgtcctagtatgg***gtttcg***gagcaa
8	tgtgcg*ggct*ctaacacgtcctagtatgg***ggtatg***gagcaa
9	tgtgcg***attg***ctaacacgtcctagtatgg*ta****tccc***gagcaa
10	tgtgcg*gg****gg***ctaacacgtcctagtatgg*taggat*gagcaa
11	tgtgcg***gtta***ctaacacgtcctagtatgg*taggat*gagcaa
12	tgtgcg***acaa***ctaacacgtcctagtatgg*t****ctaca***gagcaa
13	tgtgcg***cctg***ctaacacgtcctagtatgg*taggat*gagcaa
14	tgtgcg*ggct*ctaacacgtcctagtatgg*ta****t****gat*gagcaa
15	tgtgcg*g****cc****c****c***taacacgtcctagtatgg*t****tct****a****a***gagcaa
16	tgtgcg*ggct*ctaacacgtcctagtatgg***agac****at***gagcaa
17	tgtgcg***c****gc****c***taacacgtcctagtatgg*taggat*gagcaa
18	tgtgcg***ct****c****c***taacacgtcctagtatgg*taggat*gagcaa
19	tgtgcg*ggct*ctaacacgtcctagtatgg*ta****c****ga****a***gagcaa

20	tgtgcg***c****g****gg***ctaacacgtcctagtatgg*taggat*gagcaa
21	tgtgcg***atgg***ctaacacgtcctagtatgg*taggat*gagcaa
22	tgtgcg***cctg***ctaacacgtcctagtatgg*taggat*gagcaa
23	tgtgcg*g****atc***ctaacacgtcctagtatgg*taggat*gagcaa
24	tgtgcg***atga***ctaacacgtcctagtatgg*taggat*gagcaa
Original or pConstitutive (Siegl et al.)[1]	tgtgcg*ggct*ctaacacgtcctagtatgg*taggat*gagcaa
25	tgtgcg*ggct*ctaacacgtcctagtatgg*t****g****gg* gagcaa

Supplementary Table C.2. List of genetic parts.

Part	Nucleotide Sequence
BbaJ23104 + UTR	ttgacagctagctcagtcctaggtattgtgctagccttctctgctgttaaggtttgcttagaag cggataacaatttcacacatactagagaaagaggagaaatactag
***R. opacus/E. coli* origin of replication on pNG2[2]**	atggtaaatctgcgcagacagccctgtgcagctgaaacgcggttacgtatagcttgccatatgt ctagccatacgtaaccgcaggtaaaaggcatatttttcgcgtgtcatggctagtaaataacacc ggtgtcatttagagtcagggaaagacaatgaaaaacgaagaaagccaccgggcggcaacccgat gactttcgcttatcacccagcacacacctgggagaaatcacggtcatgagtttacagactcatg cgcagaatgcgcacactaaaacacctacccgcgtcgagcgcgaccgtggtggactggacaacac cccagcatctgccagtgaccgcgacctttttacgcgatcatctaggccgcgatgtactccacggt tcagtcacacgagactttaaaaaggcctatcgacgcaacgctgacggcacgaactcgccgcgta tgtatcgcttcgagactgatgctttaggacggtgcgagtacgccatgctcaccaccaagcagta cgccgccgtcctggtcgtagacgttgaccaagtaggtaccgcaggcggtgaccccgcagactta aacccgtacgtccgcgacgtggtgcgctcactgattactcatagcgtcgggccagcctgggtgg gtattaacccaactaacggcaaagcccagttcatatggcttattgaccctgtctacgctgaccg taacggtaaatctgcgcagatgaagcttcttgcagcaaccacgcgtgtgctgggtgagctttta gaccatgacccgcacttttcccaccgctttagccgcaacccgttctacacaggcaaagccccta ccgcttatcgttggtataggcagcacaaccgggtgatgcgccttggagacttgataaagcaggt aagggatatggcaggacacgaccagttcaaccccaccccacgccagcaattcagctctggccgc gaacttatcaacgcggtcaagacccgccgtgaagaagcccaagcattcaaagcactcgcccagg acgtagacgcggaaatcgccggtggtctcgaccagtatgacccggaacttatcgacggtgtgcg tgtgctctggattgtccaaggaaccgcagcacgcgacgaaacagcctttagacatgcgcttaag actggccaccgcttgcgccagcaaggccaacgcctgacagacgcagcaatcatcgacgcctatg agcacgcctacaacgtcgcacacacccacggcggtgcaggccgcgacaacgagatgccacccat gcgcgaccgccaaaccatggcaaggcgcgtgcgcgggtatgtcgcccaatccaagagcgagacc tacagcggctctaacgcaccaggtaaagccaccagcagcgagcggaaagccttggccacgatgg gacgcagaggcggacaaaaagccgcacaacgctggaaaacagaccccgagggcaaatatgcgca agcacaaaggtcgaagcttgaaaagacgcaccgtaagaaaaaggctcaaggacgatctacgaag tcccgtattagccaaatggtgaacgatcagtatttccagacagggacagttcccacgtgggctg aaataggggcagaggtaggagtctctcgcgccacggttgctaggcatgtcgcggagctaaagaa gagcggtgactatccggacgtttaaggggtctcataccgtaagcaatatacggttcccctgccg ttaggcagttagataaaacctcacttgaagaaaaccttgaggggcagggcagcttatatgcttc aaagcatgacttcctctgttctcctagacctcgcaaccctccgccataacctcacc
Kanamycin resistance gene cassette[3]	tcagaagaactcgtcaagaaggcgatagaaggcgatgcgctgcgaatcgggagcggcgataccg taaagcacgaggaagcggtcagcccattcgccgccaagctcttcagcaatatcacgggtagcca acgctatgtcctgatagcggtccgccacacccagccggccacagtcgatgaatccagaaaagcg gccatttttccaccatgatattcggcaagcaggcatcgccatgggtcacgacgagatcctcgccg tcgggcatccgcgccttgagcctggcgaacagttcggctggcgcgagcccctgatgctcttcgt ccagatcatcctgatcgacaagaccggcttccatccgagtacgtgctcgctcgatgcgatgttt cgcttggtggtcgaatgggcaggtagccggatcaagcgtatgcagccgccgcattgcatcagcc atgatggatactttctcggcaggagcaaggtgagatgacaggagatcctgccccggcacttcgc ccaatagcagccagtcccttcccgcttcagtgacaacgtcgagcacagctgcgcaaggaacgcc cgtcgtggccagccacgatagccgcgctgcctcgtcttggagttcattcagggcaccggacagg tcggtcttgacaaaaagaaccgggcgcccctgcgctgacagccggaacacggcggcatcagagc agccgattgtctgttgtgcccagtcatagccgaatagcctctccacccaagcggccggagaacc tgcgtgcaatccatcttgttcaatcatgcgaaacgatcctcatcctgtctcttgatcagatctt

	gatcccctgcgccatcagatccttggcggcaagaaagccatccagtttactttgcagggcttcc caaccttaccagagggcgccccagctggcaattccggttcgcttgctgtccataaaaccgccca gtctagctatcgccatgtaagcccactgcaagctacctg
Gentamicin resistance gene cassette[2]	ctgcccaaggttgccgggtgacgcacaccgtggaaacggatgaaggcacgaacccagttgacat aagcctgttcggttcgtaaactgtaatgcaagtagcgtatgcgctcacgcaactggtccagaac cttgaccgaacgcagcggtggtaacggcgcagtggcggttttcatggcttgttatgactgtttt tttgtacagtctatgcctcgggcatccaagcagcaagcgcgttacgccgtgggtcgatgtttga tgttatggagcagcaacgatgttacgcagcagggcagtcgccctaaaacaaagttaggtggctc aagtatgggcatcattcgcacatgtaggctcggccctgaccaagtcaaatccatgcgggctgct cttgatcttttcggtcgtgagttcggagacgtagccacctactcccaacatcagccggactccg attacctcgggaacttgctccgtagtaagacattcatcgcgcttgctgccttcgaccaagaagc ggttgttggcgctctcgcggcttacgttctgcccaagtttgagcagccgcgtagtgagatctat atctatgatctcgcagtctccggcgagcaccggaggcagggcattgccaccgcgctcatcaatc tcctcaagcatgaggccaacgcgcttggtgcttatgtgatctacgtgcaagcagattacggtga cgatcccgcagtggctctctatacaaagttgggcatacgggaagaagtgatgcactttgatatc gacccaagtaccgccacctaacaattcgttcaagccgagatcggcttcccggccgcggcttggt ttcatcagccatccgcttgccctcatctgttacgccggcggtagccggccagcctcgcagagca ggattcccgttgagcaccgccaggtgcgaataagggacagtgaagaaggaacacccgctcgcgg gtgggcctacttcacctatcctgcccggctgacgccgttggatacaccaaggaaagtctacacg aacccttggcaaaatcctgtatatcgtgcgaaaaaggatggatataccgaaaaaatcgctata atgaccccgaagcagggttatgcagcggaaaagcgct
Hygromycin B resistance gene cassette[4]	atgagccggatgaccggcaccacctggcggtccgcgatggacggcacgaccgaccggaacgcgc tgctcgccctggcccgcgaactcggccgggtgctcggccggctgcacagggtgccgctgaccgg gaacaccgtgctcacccccattccgaggtcttcccggaactgctgcgggaacgccgcgcggcg accgtcgaggaccaccgcgggtggggctacctctcgccccggctgctggaccgcctggaggact ggctgccggacgtggacacgctgctggccggccgcgaaccccggttcgtccacggcgacctgca cgggaccaacatcttcgtggacctggccgcgaccgaggtcaccgggatcgtcgacttcaccgac gtctatgcgggagactcccgctacagcctggtgcaactgcatctcaacgccttccggggcgacc gcgagatcctggccgcgctgctcgacggggcgcagtggaagcggaccgaggacttcgcccgcga actgctcgccttcaccttcctgcacgacttcgaggtgttcgaggagaccccgctggatctctcc ggcttcaccgatccggaggaactggcgcagttcctctgggggccgccggacaccgcccccggcg cctgacgccccgggccgcccggcgccgcccccggcccccggcggccgcaaaataaaaaaggggga cctctagggtccccaa
Chloramphenicol resistance gene cassette (with a codon optimized gene for *R. opacus*)	tgtgcgggctctaacacgtcctagtatggtaggatgagcaacatttcgacgccgagagattcgc cgcccgaaatgagcacgatccgcatgcttaattaagaaggagatatacatatggagaaaaagat cacggggtacaccaccgtggacatcagccagtggcaccgtaaggaacatttcgaagcgttccag tccgtcgcgcaatgtacgtacaatcagaccgtccaactcgatattacggcgtttctgaagaccg tgaagaagaataagcacaaattttaccccgcctttattcatatcctggcgcggctgatgaacgc ccatccagaatttcgcatggccatgaaagatggcgagttggtcatctgggatagcgtccatcca tgttatacggtgttccacgagcagaccgaaacctttcgagcttgtggagcgagtatcacgatg atttccggcagttcttgcacatttatagccaagatgtggcctgctatggtgagaatctggccta ctttccgaaaggtttcattgagaacatgtttttgtgtccgccaatccgtgggtctcgttcacg agcttcgacctcaacgtcgcgaatatggataacttttttgcgccggtctttaccatggggaaat attatacgcagggcgataaggtgctgatgccactcgcgatccaagtgcatcatgccgtctgcga tggcttccacgtcggtcggatgttgaacgaactccaacaatattgcgatgagtggcaaggtggt gcgtaagctatgccaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgt tgtttgtcggtgaacgctctctgag
***R. opacus* origin of replication on pAL5000 (short variant; S)[3]**	ttagaacagcggtggattgtcggcttcgttgtgggccttttgagccgcttcctgttctgccgca cgctctttcctcgcccgatagccgagtcgcttaacggtgtccagatgcagcccgaaatgtttgg ccgtttgcggccaagagtggcccctcgtcgtcgtgataggcgcggatgcgttcgcggcgtgcagc ctgctcggcgagccactcgctgcgttcctgcgccacgagccggacgacgtggcgttcggatagt ccggtgattcgagcgccttcggcggcggtcacgcgccgcttttgcggacagtcggctgccggt tgtagccgtcgctgtagccgtcgctcatagcaatgcctccatggctgacgcggactttgcgcgc cgcgcaactgtgctcgccgccgtgcgcgctgctgcgcccttccgcgagatggccgactggcgcg cactgagtgtggcctcgtagaccacgatcccgtccgcccaaatgcgcgacttggttgtgatcca acgccaaatgctgttggcgatggcgcggacctcgctgtccggtagcggtccgggacacacgtcg ttgcacgggaactcggcgtttcgcgcgtggcactcggcatagatcgcgcggccgagtccgtcca cgttccgggtcggcaggtagatccgcatgagggcgggacgataggcccacaacctgacggaatc gaacagtgcgcaattccgccctagcggcgtcggagccgctttgtacgtggtctgctgacgccag cgcggcggtggcatgttcgcgccgagctcggcctcgatgtggctgagtgtgtagagatctgagt ggagccattccgtttcccaggcgatgtggccggggtttttggtcatgaggcctgagtaactgcg gtcgccgtccacggcgcgccgaaggccttcggcgcacgccgccatgtatgcgagcggcttacgc cgcgcgtattcggtgcgtggaacaggggcgttgagtgcccacactgcgtgtgcgtggccgttgg

	cgcgattgcccacgatcgcgttgggcagcggatgggacccccgggcgctgagcgctcggagcgc tgcgtctggatggtctacgtccacgaccagcaggtttgccagcgctgttgggttcgcctcgatg taccggcggcctagggccgacgcgcggctttggcggtagatcccctcgagcagatcgtcgcttg ccagcggccagtacggcagccagagctgctcaaattcgtcggcgacgtggctca
***R. opacus* origin of replication on pAL5000 (long variant; L)[4]**	atcgagccgagaacgttatcgaagttggtcatgtgtaatcccctcgtttgaactttggattaag cgtagatacacccttggacaagccagttggattcggagacaagcaaattcagccttaaaaaggg cgaggcctgcggtggtggaacaccgcagggcctctaaccgctcgacgcgctgcaccaaccagcc cgcgaacggctggcagccagcgtaaggcgcggctcatcgggcggcgttcgccacgatgtcctgc acttcgagccaagcctcgaacacctgctggtgtgcacgactcacccggttgttgacaccgcgcg cggccgtgcgggctcggtggggcggctgtgtcgcccttgccagcgtgagtagcgcgtacctcac ctcgcccaacaggtcgcacacagccgattcgtacgccataaagccaggtgagcccaccagctcc gtaagttcgggcgctgtgtggctcgtacccgcgcattcaggcggcagggggtctaacgggtcta aggcggcgtgtacggccgccacagcggctctcagcggcccggaaacgtcctcgaaacgacgcat gtgttcctcctggttggtacaggtggttgggggtgctcggctgtcgctggtgttccaccaccag ggctcgacgggagagcgggggagtgtgcagttgtggggtggcccctcagcgaaatatctgactt ggagctcgtgtcggaccatacaccggtgattaatcgtggtctactaccaagcgtgagccacgtc gccgacgaatttgagcagctctggctgccgtactggccgctggcaagcgacgatctgctcgagg ggatctaccgccaaagccgcgcgtcggccctaggccgccggtacatcgaggcgaacccaacagc gctggcaaacctgctggtcgtggacgtagaccatccagacgcagcgctccgagcgctcagcgcc cgggggtcccatccgctgcccaacgcgatcgtgggcaatcgcgccaacggccacgcacacgcag tgtgggcactcaacgcccctgttccacgcaccgaatacgcgcggcgtaagccgctcgcatacat ggcggcgtgcgccgaaggccttcggcgcgccgtcgacggcgaccgcagttactcaggcctcatg accaaaaaccccggccacatcgcctgggaaacggaatggctccactcagatctctacacactca gccacatcgaggccgagctcggcgcgaacatgccaccgccgcgctggcgtcagcagaccacgta caaagcggctccgacgccgctagggcggaattgcgcactgttcgattccgtcaggttgtgggcc tatcgtcccgccctcatgcggatctacctgccgacccggaacgtggacggactcggccgcgcga tctatgccgagtgccacgcgcgaaacgccgaattcccgtgcaacgacgtgtgtcccggaccgct accggacagcgaggtccgcgccatcgccaacagcatttggcgttggatcacaaccaagtcgcgc atttgggcggacgggatcgtggtctacgaggccacactcagtgcgcgccagtcggccatctcgc ggaagggcgcagcagcgcgcacggcggcgagcacagttgcgcggcgcgcaaagtccgcgtcagc catggaggcattgctatgagcgacggctacagcgacggctacagcgacggctacaaccggcagc cgactgtccgcaaaaagcggcgcgtgaccgccgccgaaggcgctcgaatcaccggactatccga acgccacgtcgtccggctcgtggcgcaggaacgcagcgagtggctcgccgagcaggctgcacgc cgcgaacgcatccgcgcctatcacgacgacgagggccactcttggccgcaaacggccaaacatt tcgggctgcatctggacaccgttaagcgactcggctatcgggcgaggaaagagcgtgcggcaga acaggaagcggctcaaaaggcccacaacgaagccgacaatccaccgctgttctaacgcaattgg ggagcgggtgtcgcgggggttccgtgggggttccgttgcaacgggtcggacaggtaaaagtcc tggtagacgctagttttctggtttgggccatgcctgtctcgttgcgtgtttcgttgcgcccgtt ttgaataccagccagacgagacggggttctacgaatcttggtcgataccaagccatttccgctg aatatcggggagctcaccgccagaatcggtggttgtggtgatgtacgtggcgaactccgttgta gtgcctgtggtggcatccgtggccactctcgttgcacggttcgttgtgccgttacaggccccgt tgacagctcaccgaacgtagttaaaacatgctggtcaaactaggtttaccaacgatacgagtca gctcatctagggccagttctaggcgttgttcgttgcgcggttcgttgcgcatgtttcgtgtggt tgctagatggctccgcaaccacacgcttcgaggttgagtgcttccagcacgggcgcgatccaga agaacttcgtcgtgcgactgtcctcgtt
***R. opacus* origin of replication on pB264[5]**	cacacagctcttcagaggaacgtcgggcggtcgagcgcggcgtcgtcgcgtgtggagttgcgcg agcgggtggtcgaggtcgagaaggtggtgcatcggccggccacgacggccattgctcgtttgaa caggtctcacagaccatcagggtgcaacgcacccgcacggcagggctgtcccaaggggacgggg ctggtagtggacgtcatcgcagctgatactcatcgctcgattcctgccagggcgtgatggacag tgccgaccgagcatcccagctcattggcgatagctctcatggacatgccagtaccgcgtagctc ccgtatcttctcgcgccgttgctgtgcccgggacaggtaggcctctcgcggctctgatgtccac cggatgatgctggccgtcgatacgcctgtgcgttcggcgagctcccgcactgttatcccgttgc gaggcagacgctcagtcatgtgtggcctcctcaattgccagtcgcacctgatccagtttggtgc tgcgtgcacgccgggcctggccgctggcgacgccgcctttgcggccacgggcggattggatggt cgagagtgtggcttcgtagacggcgataccgtctttccagatgcgggatcgcgtggtgatccat cggtgaatgctggcggcgatggcgcgggcttcagagggtgggagttcctcggtgtagctggcgt tgagggcggtgacttcgcggtggagggcgagttcgaggccggcgctgtcatgagtgggcaggta gttgcgcatgaggccggggcgggtggcccaggtgcgagcggcgtggaagattgcgcagttgcgg ccgagtcctatgggggcgtcacggtgggccttggtgtgctgccagcgcggggagggcatgtgtg tgccgagttcggcctcgagctcggcgaggctgcgcggttcggtgtggatccagtgcgtgtccca gcctgagtgtgtggggttcttggtcatcaggcctgagtagccgatgtccccttcgacggcccgg cgtagcccttcggtaacggcggcggcgtacgcgaggggcttccgtcgggcgtactcggtgcggg tgaacggctcgcgcagccaccagtgcaggtgcgcatgcccgttgctgctgttttcgatcacggc

	gttcggcatagggtgattgcctgccgctgacagtgcgcgcagggcggagtcggggtggtcgacg tcgacgacgagcagattgctcagtgtctgcgggttggcctcgatgtagcgccgctcaagtgcgg cagggcggcgcattcgatagacccccctcaaggaagtcgtttgttgccaaaggccagaacggtag ccacatttgctcccagtccccaccttcccgcaccccgtgtgtgtccatgctggtgaccgtcgca tgctctcgaacgaaacctgggcatttccctcggtgtgttcaaagttaatcgtgaagccaagtca gatgtttcggtctttgaagagcacacac
***E. coli* origin of replication pMB1[2]**	cgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctca agtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccc tcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcggg aagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctcc aagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatc gtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggat tagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctac actagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttg gtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagca gattacgcgcagaaaaaaggatctcaagaagatcctttgatct
***E. coli* origin of replication pBR322[6]**	agatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaa accaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggta actggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccacc acttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgc tgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcg cagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccg aactgagatacctacagcgtgagcattgagaaagcgccacgcttcccgaagggagaaaggcgga caggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaac gcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgat gctcgtcaggggggcggagcctatggaaaaacgccagcaacgcg
Che9c60 and Che9c61 genes (polycistronic)[6]	atgagtgtgcccacacaggacggaatgcaccggttcgtcgacgaggacgtctaccacgctgacc ggggctcgctgtcggtatccggcgcgaagctgctgttgccgccgtcgtgtcccgcgaaattccg ctgggagatggacaacacccggaagccgaaaaaggtctgggacttcggacatgtcgcgcacaaa ctggtgctcggcaagggtgccgagttcgagatcctcgaccccgaggtgcacgggctgaaggcgg acggtacgccgtcggagaagccgaccgcgacgggcatgtggcgcaaggccgaggctgaggctcg caaacagggcaaggtgccgattcacgtcgacctgttcacgaaggcgtacgacatggccgaaaag gtgcgtcagcacccgacagccggcccgatcttcgccaatcctgacggcgaggccgaggtcgcgc tgtactacaccgaccccgagaccggcgtgcggctgcgtggccggatcgactggctcactgacga tatcgatgattacaagacgtcgatgaccgcgaacccggccgagctgaaaaccaagttctacaag ctcggctatttcatgcaggcggcctggtacatcgatctactggtcgccctcgggctcgccgaga acccgcgattccggttcatcacgcaggagaaagaaccgccctacgtcgtgactccgatccagta cgacgacgaggcgatcgaagaggggcggcgccgcaaccgccaggcgatccggctctacgccgac tgcatggaatcgggcaagtggcctgactacagcgacgacgtggtcacgatcagcctgccctcgt gggggctgccgcgaccgcagaccgtcggcgacgtcgtcaccgacagctatatctacgacaccga cccgctcgaagaggccgacccgattgaaggggattacatctatggctgaaaatgctgtcaccaa gcaggattcgcccaaggcacccgagacgatctcgcaagtgctgcaggtgctcgtgccgcagctg gcgcgtgcagtgcccaagggtatggaccccgaccgcatcgcgcggatcgtgcaaaccgagatcc gcaagtcgcgcaacgcgaaagctgctggcatcgctaagcagtccctcgacgactgcacgcaaga gtcatttgccggtgcgctgctgacctcggccgcgctcggtctcgagcccggtgtcaacggcgag tgctacctcgtgccctaccgcgacacccggcgcggcgtggtcgagtgccagctgatcatcggct accagggcatcgtcaaactgttctggcagcacccgcgcgcctcccggatcgacgcgcagtgggt cggcgcgaacgacgaattccattacaccatgggcctcaatccgacgctgaaacacgtgaaggcc aagggtgatcggggtaatccggtctacttctacgcgatcgtcgaggtgaccggcgctgagccgc tgtgggacgtgttcaccgccgacgagatcagggaattgcgtcgcggcaaggtcggatcctcggg cgacatcaaggacccgcagcgctggatggagcggaagaccgcgctcaaacaggtgctgaagctg gcaccgaagacgacgcggctcgacgcggcgatccgcgccgacgatcgcccgggcaccgacctgt cacagtcgcaggcgctcgcgctgccgtcgaccgtcaagccgacggccgactacatcgacggcga gatcgccgagccgcacgaggtcgatacgccaccgaagtcgtcgcgcgcacaacgcgcgcagcgc gccaccgcgccggcgcccgacgtgcagatggccaatcccgatcagctgaagcgcctcggcgaga tccagaaggccgaaaagtacaacgacgccgattggttcaagttcctcgccgattcggccggcgt caaagccacgcgcgccgccgacctcacgttcgacgaggcgaaggctgtcatcgacatgttcgac gggcccaacgcatga
pConstitutive from *Streptomyces lividans* TK24 + UTR[1]	tgtgcgggctctaacacgtcctagtatggtaggatgagcaacatttcgacgccgagagattcgc cgcccgaaatgagcacgatccgcatgcttaattaagaaggagatatacat

Gene encoding EYFP[7, 8]	atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcg acgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagct gaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccacc ttcggctacggcctgcaatgcttcgcccgctaccccgaccacatgaagctgcacgacttcttca agtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaacta caagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggc atcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccaca acgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaa catcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccatcggcgacggc cccgtgctgctgcccgacaaccactacctgagctaccagtccgccctgagcaaagaccccaacg agaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatgga cgagctgtacaaggctgcgaattga
Gene encoding GFP+[4, 8]	atggctagcaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtg atgttaatgggcacaaattttctgtcagtggagagggtgaaggtgatgctacatacggaaagct tacccttaaatttatttgcactactggaaaactacctgttccatggccaacacttgtcactact ttgacctatggtgttcaatgcttttcccgttatccggatcatatgaaacggcatgacttttca agagtgccatgcccgaaggttatgtacaggaacgcactatatctttcaaagatgacgggaacta caagacgcgtgctgaagtcaagtttgaaggtgataccttgttaatcgtatcgagttaaaaggt attgattttaaagaagatggaaacattctcggacacaaactcgagtacaactataactcacaca atgtatacatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaa cattgaagatggctccgttcaactagcagaccattatcaacaaaatactccaattggcgatggc cctgtccttttaccagacaaccattacctgtcgacacaatctgccctttcgaaagatcccaacg aaaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacacatggcatgga tgagctctacaaataa
***araC* cassette + pBAD promoter + UTR**[7, 8]	ttatgacaacttgacggctacatcattcacttttttcttcacaaccggcacggaactcgctcggg ctggccccggtgcatttttttaaatacccgcgagaaatagagttgatcgtcaaaaccaacattgc gaccgacggtggcgataggcatccgggtggtgctcaaaagcagcttcgcctggctgatacgttg gtcctcgcgccagcttaagacgctaatccctaactgctggcggaaaagatgtgacagacgcgac ggcgacaagcaaacatgctgtgcgacgctggcgatatcaaaattgctgtctgccaggtgatcgc tgatgtactgacaagcctcgcgtacccgattatccatcggtggatggagcgactcgttaatcgc ttccatgcgccgcagtaacaattgctcaagcagatttatcgccagcagctccgaatagcgccct tccccttgcccggcgttaatgatttgcccaaacaggtcgctgaaatgcggctggtgcgcttcat ccgggcgaaagaaccccgtattggcaaatattgacggccagttaagccattcatgccagtaggc gcgcggacgaaagtaaacccactggtgataccattcgcgagcctccggatgacgaccgtagtga tgaatctctcctggcgggaacagcaaaatatcacccggtcggcaaacaaattctcgtccctgat ttttcaccaccccctgaccgcgaatggtgagattgagaatataacctttcattcccagcggtcg gtcgataaaaaaatcgagataaccgttggcctcaatcggcgttaaacccgccaccagatgggca ttaaacgagtatcccggcagcaggggatcattttgcgcttcagccatacttttcatactcccgc cattcagagaagaaaccaattgtccatattgcatcagacattgccgtcactgcgtcttttactg gctcttctcgctaaccaaaccggtaaccccgcttattaaaagcattctgtaacaaagcgggacc aaagccatgacaaaaacgcgtaacaaaagtgtctataatcacggcagaaaagtccacattgatt atttgcacggcgtcacactttgctatgccatagcatttttatccataagattagcggatcctac ctgacgcttttttatcgcaactctctactgtttctccatacagcggataaagtagcaaagagaag gaggttagga
Gene encoding *S. thermophilus* dCas9[9]	atgtcggacctggtgctgggcctggccatcggcatcggctcggtgggcgtgggcatcctgaaca aggtgaccggcgagatcatccacaagaactcgcgcatcttcccggccgcccaggccgagaacaa cctggtgcgccgcaccaaccgccagggccgccgcctggcccgccgcaagaagcaccgccgcgtg cgcctgaaccgcctgttcgaggagtcgggcctgatcaccgacttcaccaagatctcgatcaacc tgaacccgtaccagctgcgcgtgaagggcctgaccgacgagctgtcgaacgaggagctgttcat cgccctgaagaacatggtgaagcaccgcggcatctcgtacctggacgacgcctcggacgacggc aactcgtcggtgggcgactacgcccagatcgtgaaggagaactcgaagcagctggagaccaaga ccccgggccagatccagctggagcgctaccagacctacggccagctgcgcggcgacttcaccgt ggagaaggacggcaagaagcaccgcctgatcaacgtgttcccgacctcggcctaccgctcggag gccctgcgcatcctgcagacccagcaggagttcaacccgcagatcaccgacgagttcatcaacc gctacctggagatcctgaccggcaagcgcaagtactaccacggcccgggcaacgagaagtcgcg caccgactacggccgctaccgcacctcgggcgagaccctggacaacatcttcggcatcctgatc ggcaagtgcaccttctacccggacgagttccgcgccgccaaggcctcgtacaccgcccaggagt tcaacctgctgaacgacctgaacaacctgaccgtgccgaccgagaccaagaagctgtcgaagga gcagaagaaccagatcatcaactacgtgaagaacgagaaggccatgggcccggccaagctgttc aagtacatcgccaagctgctgtcgtgcgacgtggccgacatcaagggctaccgcatcgacaagt cgggcaaggccgagatccacaccttcgaggcctaccgcaagatgaagaccctggagaccctgga catcgagcagatggaccgcgagaccctggacaagctggcctacgtgctgaccctgaacaccgag cgcgagggcatccaggaggccctggagcacgagttcgccgacggctcgttctcgcagaagcagg

	tggacgagctggtgcagttccgcaaggccaactcgtcgatcttcggcaagggctggcacaactt ctcggtgaagctgatgatggagctgatcccggagctgtacgagacctcggaggagcagatgacc atcctgacccgcctgggcaagcagaagaccacctcgtcgtcgaacaagaccaagtacatcgacg agaagctgctgaccgaggagatctacaacccggtggtggccaagtcggtgcgccaggccatcaa gatcgtgaacgccgccatcaaggagtacggcgacttcgacaacatcgtgatcgagatggcccgc gagaccaacgaggacgacgagaagaaggccatccagaagatccagaaggccaacaaggacgaga aggacgccgccatgctgaaggccgccaaccagtacaacggcaaggccgagctgccgcactcggt gttccacggccacaagcagctggccaccaagatccgcctgtggcaccagcagggcgagcgctgc ctgtacaccggcaagaccatctcgatccacgacctgatcaacaactcgaaccagttcgaggtgg acgccatcctgccgctgtcgatcaccttcgacgactcgctggccaacaaggtgctggtgtacgc caccgccaaccaggagaagggccagcgcaccccgtaccaggccctggactcgatggacgacgcc tggtcgttccgcgagctgaaggccttcgtgcgcgagtcgaagaccctgtcgaacaagaagaagg agtacctgctgaccgaggaggacatctcgaagttcgacgtgcgcaagaagttcatcgagcgcaa cctggtggacacccgctacgcctcgcgcgtggtgctgaacgccctgcaggagcacttccgcgcc cacaagatcgacaccaaggtgtcggtggtgcgcggccagttcacctcgcagctgcgccgccact ggggcatcgagaagacccgcgacacctaccaccaccacgccgtggacgccctgatcatcgccgc ctcgtcgcagctgaacctgtggaagaagcagaagaacaccctggtgtcgtactcggaggaccag ctgctggacatcgagaccggcgagctgatctcggacgacgagtacaaggagtcggtgttcaagg ccccgtaccagcacttcgtggacaccctgaagtcgaaggagttcgaggactcgatcctgttctc gtaccaggtggactcgaagttcaaccgcaagatctcggacgccaccatctacgccacccgccag gccaaggtgggcaaggacaaggccgacgagacctacgtgctgggcaagatcaaggacatctaca cccaggacggctacgacgccttcatgaagatctacaagaaggacaagtcgaagttcctgatgta ccgccacgacccgcagaccttcgagaaggtgatcgagccgatcctggagaactacccgaacaag cagatcaacgacaagggcaaggaggtgccgtgcaacccgttcctgaagtacaaggaggagcacg gctacatccgcaagtactcgaagaagggcaacggcccggagatcaagtcgctgaagtactacga ctcgaagctgggcaaccacatcgacatcaccccgaaggactcgaacaacaaggtggtgctgcag tcggtgtcgccgtggcgcgccgacgtgtacttcaacaagaccaccggcaagtacgagatcctgg gcctgaagtacgccgacctgcagttcgacaagggcaccggcacctacaagatctcgcaggagaa gtacaacgacatcaagaagaaggagggcgtggactcggactcggagttcaagttcaccctgtac aagaacgacctgctgctggtgaaggacaccgagaccaaggagcagcagctgttccgcttcctgt cgcgcaccatgccgaagcagaagcactacgtggagctgaagccgtacgacaagcagaagttcga gggcggcgaggccctgatcaaggtgctgggcaacgtggccaactcgggccagtgcaagaagggc ctgggcaagtcgaacatctcgatctacaaggtgcgcaccgacgtgctgggcaaccagcacatca tcaagaacgagggcgacaagccgaagctggacttctaa
sgRNA cassette	tgtgcgggctctaacacgtcctagtatggtaggatgagcaacaNNNNNNNNNNNNNNNNNNNNg tttttgtactcgaaagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcatttt atggcagggtgttttttttttgtcgacttggggaccctagaggtccccttttttatttttttgg atgaagagcattatcaatggctcttcacaggaagcttaaaaaaaaaaagccccgcgattgcgggg ctttttttttt
***E. coli* origin of replication p15a[10]**	cctagggatatattccgcttcctcgctcactgactcgctacgctcggtcgttcgactgcggcga gcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatacttaacaggg aagtgagagggccgcggcaaagccgtttttccataggctccgccccctgacaagcatcacgaa atctgacgctcaaatcagtggtggcgaaacccgacaggactataaagataccaggcgtttcccc ctggcggctccctcgtgcgctctcctgttcctgcctttcggtttaccggtgtcattccgctgtt atggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagct ggactgtatgcacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcgtctt gagtccaacccggaaagacatgcaaaagcaccactggcagcagccactggtaattgatttagag gagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcg ctcctccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgcc ctgcaaggcggttttttcgttttcagagcaagagattacgcgcagaccaaaacgatctcaagaa gatcatcttattaa
Tandem rrnB T1 and T7Te terminator for EYFP[10]	accaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtt tgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcgttta ta
Tandem rrnB T2 and TsynB terminator for dCas9$_{Sth}$[9]	agaaggccatcctgacggatggccttttttctagaaaaaaaaaaaagcgccgcaactgcggcgctt tttttttt

Supplementary Table C.3. qPCR primers used for plasmid copy number analysis. The "expected value of amplicon" or E-value refers to the likelihood that the amplicon sequence will match or align with the examined genome, excluding the target sequence. As the E-value approaches zero, the likelihood of an off-target match decreases, leading to a higher degree of specificity. E-values were determined using NCBI's blastn. PCR reaction efficiency was calculated using the slope of a dilution series of known PCR amplicon quantity conducted in duplicate (see Materials and Methods C; Supplementary Figure C.1).

Primer Name	Target	Nucleotide Sequence (5' to 3')	Expected Value of Amplicon	Efficiency %	Amplicon Length (nt)
PD630P1F	Plasmid 1	F: CGTACAACAGCTACCAGTTCACG	9e-86	97	164
PD630P1R		R: GCGGTGTTACTGAACGTGTCC			
PD630P2F	Plasmid 2	F: GACGATGAATCTGCTGTCCTCC	2e-72	95	140
PD630P2R		R: AGGTTCGGGTGAGCTTTCAGG			
PD630P3F	Plasmid 3	F: GCCATCAGTCCTCGTCTTCC	3e-101	92	192
PD630P3R		R: ACTCCTGGGTGTACCTGCTG			
PD630P4F	Plasmid 4	F: GGTTCGTTCTTCCCATACAGGG	3e-96	90	183
PD630P4R		R: GGTGCTCAATCATTTGGTGGTCG			
PD630P5F	Plasmid 5	F: TGGGACATGGCGATGCTACC	5e-109	96	206
PD630P5R		R: CGTTTGCGTTGGTCGTTCGG			
PD630P6F	Plasmid 6	F: CCACCCACACACTTGAGACG	9e-81	96	155
PD630P6R		R: GCGAGTCACGAGGTTAAGGC			
PD630P7F	Plasmid 7	F: GCTCGGTCCCACTGACATATTCG	2e-113	95	214
PD630P7R		R: CGGCAGCGGAATTTAGTCGG			
PD630P8F	Plasmid 8	F: AAGCACAGTTTGTCGTGTATGTCG	1e-74	96	144
PD630P8R		R: GTATGGGTACTGTCAAGATGGTGC			
PD630P9F	Plasmid 9	F: CACCGCATCACAACCGTTTTCC	6e-119	93	224
PD630P9R		R: CCATCACCCGTGTCCTGATCG			
PD630PCF	Chromosome	F: GAAAGCGTTCGAGCACGTCG	2e-77	88	149
PD630PCR		R: CCAGGTCAAACATATTCTCGGAGC			
DKanF	pAL5000 (S) and pB264	F: GCTCTTCGTCCAGATCATCC	No significant similarity	90	175
DKanR		R: TCATCTCACCTTGCTCCTGC			
DGentF	pNG2	F: CTATGATCTCGCAGTCTCCG	No significant similarity	91	128
DGentR		R: CGTCACCGTAATCTGCTTGC			
DHygF	pAL5000 (L)	F: GTCCACGAAGATGTTGGTCC	No significant similarity	89	188
DHygR		R: AGGTCTTCCCGGAACTGCTG			

Supplementary Table C.4. Summary of candidate *R. opacus* chromosomal integration sites (ROCI-#) and endogenous plasmid integration sites (ROP#I-#). Neutral integration sites were identified as discussed in Figure 2.4. The replaced location and LPD gene annotation (NCBI database Refseq NZ_CP003949.1) are reported. Among >200 bp of each site, the replaced region with the integration cassette is shown in the "Location (nucleotide)" column. Annotations were confirmed by NCBI blastp to examine if genes had homologs with differing annotations in related species. "Up" refers to the gene upstream of the integration site. "Down" refers to the gene downstream of the integration site.

Integration Site Name	Location (nucleotide)	Location (LPD annotation)	Gene Annotations
ROCI-1	4,222,959 to 4,223,163	Between LPD04028 & LPD04029	Up: Hypothetical protein Down: Putative thiamine biosynthesis lipoprotein ApbE
ROCI-2	4,896,351 to 4,896,498	Between LPD04690 & LPD04691	Up: Hypothetical protein Down: Hypothetical protein
ROCI-3	3,978,043 to 3,978,140	Between LPD03763 & LPD03764	Up: Hypothetical protein Down: Hypothetical protein
ROP8I-1	147,710 to 147,711	Between LPD16184 & LPD16185	Up: Hypothetical protein Down: Hypothetical protein

Supplementary Table C.5. Explanation of gel lane labels in Supplementary Figure C.5. The gel lane labels correspond to the following. **A**: cell lysis control; **B**: internal integration cassette control; **C**: genome/integration cassette upstream junction; and **D**: integration cassette/genome downstream junction (see Supplementary Figure C.5A for a schematic of primer locations). The numbers associated with lanes C and D correspond to the primer pairs for individual integration sites.

Strain	Gel Lane Labels (A-D)	Nucleotide Sequences (5' to 3')	Length (bp)
All	A	F: GACGAGATCAGAGTCCGACAACTCG	1402
		R: CCGAAAATGTTACGCAGAGTGCG	
All	B	F: GAAGAAGTCGTGCAGCTTCATGTGG	568
		R: TTCTAATCCGCATATGATCAATTCAAGG	
ROCI-1	C1	F: TCTGCTGTAGATCGTGATGGTCATGG	1051
		R: CTGTACAAGGCTGCGAATTGAGACC	
ROCI-2	C2	F: CTGCTGCTGATACTCGAGCACATCC	1138
		R: CTGTACAAGGCTGCGAATTGAGACC	
ROCI-3	C3	F: CCTAAGTCGATCGACGATCACCAGG	779
		R: CTGTACAAGGCTGCGAATTGAGACC	
ROP8I-1	C4	F: CAACCGCACTAATCTAAGCGAAACG	1160
		R: CTGTACAAGGCTGCGAATTGAGACC	
ROCI-1	D1	F: CGAATCAATACGGTCGAGAAGTAACAGG	1078
		R: AATTCGAAGGACCAGCAGATCAATTCC	
ROCI-2	D2	F: CGAATCAATACGGTCGAGAAGTAACAGG	788
		R: TCGATGCTGATGGTGATGGTCTGAC	
ROCI-3	D3	F: CGAATCAATACGGTCGAGAAGTAACAGG	637
		R: GCGAACTCCGCTTCGAATACTGC	
ROP8I-1	D4	F: CGAATCAATACGGTCGAGAAGTAACAGG	644
		R: AGCCAAGGAACAGCACATCACCAC	

Supplementary Table C.6. Growth rates of neutral site integration strains. An integration cassette containing constitutively expressed EYFP and HygR genes was integrated into one of three chromosomal candidate neutral sites (ROCI 1, 2, and 3) and one endogenous plasmid candidate neutral site (ROP8I-1). The growth rates (hr^{-1}) were calculated by taking a linear regression of natural log transformed absorbance (600 nm) across at least four 1-hour interval time points. The growth rates of ROCI-2, ROCI-3, and ROP8I-1 were indistinguishable from the wild type (WT) growth rate, as determined by a Students t-test (two tail, equal variance). The growth rate for ROCI-1 exhibited an 18% reduction ($p = 0.001$). Values represent the average of three replicates with one standard deviation.

Strain	Growth rate (hr^{-1})	Significant difference compared to WT?
Wild type (WT)	0.20 ± 0.01	N/A
ROCI-1	0.16 ± 0.01	Yes ($p = 0.001$)
ROCI-2	0.18 ± 0.01	No
ROCI-3	0.19 ± 0.01	No
ROP8I-1	0.19 ± 0.01	No

Supplementary Table C.7. CRISPRi sgRNA sequences. sgRNA PAM sites and target sequences were chosen and designed as described by Rock *et al.*[9]

sgRNA	PAM	Target sequence
A	NNAGAAG	ATGTGGTCGGGGTAGCGGGC
B	NNAGCAT	CCGCTTCTAAGCAAACCTTA
C	NNAGCAG	GCTTCTAAGCAAACCTTAACAGC

Supplementary Table C.8. Plasmids used in this work.

Plasmid name	Functional insert	Backbone	Antibiotic resistance	Length (bp)	Source
pUV15tetORm	pTet.GFP+	pAL5000 (L)	Hygromycin B	8026	Ehrt et al. 2005[4]
pXYLA	pTac.xylA	pAL5000 (S)	Kanamycin	5823	Xiong et al. 2012[3]
pAL358	LacZ	pNG2	Gentamicin	4344	Kurosawa et al. 2013[2]
pCI078	BbaJ23104-EYFP	pBR322	Kanamycin	4636	Immethun et al. 2015[10]
pDD26/pJV53	pAcetamide. Che9c60.Che9c61	pAL5000	Kanamycin	8812	Van Kessel & Hatfull 2007[6]
pDD42	ROCI-1 (LPD04028/04029) EYFP integrative vector	Integrative	Hygromycin B	4881	This study; derived from pCI078 and pUV15tetORm

pDD43	pConstitutive.GFP+ -35 saturation mutagenesis library	pAL5000 (S)	Kanamycin	4776	This study; derived from pDD57
pDD44	pConstitutive.GFP+ -10 saturation mutagenesis library	pAL5000 (S)	Kanamycin	4776	This study; derived from pDD57
pDD45	pConstitutive.GFP+ -35 & -10 simultaneous saturation mutagenesis library	pAL5000 (S)	Kanamycin	4776	This study; derived from pDD57
pDD56	pBAD.EYFP	pAL5000 (S)	Kanamycin	6013	DeLorenzo et al. 2017[8]
pDD57	pConstitutive.GFP+	pAL5000 (S)	Kanamycin	4776	This study; derived from pDD65 and pUV15tetORm
pDD65	Empty backbone	pAL5000 (S)	Kanamycin	3659	This study; derived from pXYLA
pDD67	Empty backbone	pNG2	Gentamicin	6209	This study; derived from pAL358
pDD82	Empty backbone	pB264	Kanamycin	3611	This study; derived from pDD65
pDD99	Empty backbone	pAL5000 (L)	Hygromycin B	6057	This study; derived from pUV15tetORm
pDD107	ROP8I-1 (LPD16184/16185) EYFP integrative vector	Integrative	Hygromycin B	4547	This study; derived from pCI078 and pUV15tetORm
pDD108	ROCI-2 (LPD04690/04691) EYFP integrative vector	Integrative	Hygromycin B	4290	This study; derived from pCI078 and pUV15tetORm
PLJR965	pTet.dCas9$_{Sth}$	Mycobacteria integrative	Kanamycin	8631	Rock et al. 2017[9]
pDD112	Chloramphenicol cassette (optimized)	pNG2	Chloramphenicol	5021	This study; derived from pDD67
pDD119	pAcetamide. Che9c60.Che9c61	pB264	Kanamycin	8585	This study; derived from pJV53 and pDD82
pDD120	pConstitutive. Che9c60.Che9c61	pB264	Kanamycin	6247	This study; derived from pDD57 and pDD119
pDD132	ROCI-3 (LPD03763/03764) EYFP integrative vector	Integrative	Hygromycin B	4238	This study; derived from pCI078 and pUV15tetORm
pDD143	pBAD.dCas9$_{Sth}$. pCon.sgRNA.A	pAL5000 (S)	Gentamicin	9551	This study; derived from

					PLJR965 and pDD56
pDD145	pBAD.dCas9$_{Sth}$. pCon.sgRNA.B	pAL5000 (S)	Gentamicin	9551	This study; derived from PLJR965 and pDD56
pDD147	pBAD.dCas9$_{Sth}$. pCon.sgRNA.C	pAL5000 (S)	Gentamicin	9554	This study; derived from PLJR965 and pDD56

Supplementary Table C.9. Strains used in this work.

Strain name	Genus	Species	Strain	Plasmid contained	Corresponding figure
DMD012	*Escherichia*	*coli*	DH10B	pCI078	N/A
DMD022	*Rhodococcus*	*opacus*	PD630	Wild type	3 and 5
DMD026	*Escherichia*	*coli*	DH10B	pDD67	N/A
DMD052	*Escherichia*	*coli*	DH10B	pDD42	N/A
DMD056	*Escherichia*	*coli*	DH10B	pDD65	N/A
DMD057	*Escherichia*	*coli*	DH10B	pDD56	N/A
DMD061	*Escherichia*	*coli*	DH10B	pDD26/pJV53	N/A
DMD080	*Rhodococcus*	*opacus*	PD630	pDD26/pJV53	N/A
DMD081	*Rhodococcus*	*opacus*	PD630	pDD65	2
DMD093	*Escherichia*	*coli*	DH10B	pDD82	N/A
DMD095	*Escherichia*	*coli*	DH10B	pUV15tetORm	N/A
DMD096	*Escherichia*	*coli*	DH10B	pAL358	N/A
DMD097	*Escherichia*	*coli*	DH10B	pDD57	N/A
DMD098	*Escherichia*	*coli*	DH10B	pDD119	N/A
DMD101	*Rhodococcus*	*opacus*	PD630	pDD67	2
DMD102	*Rhodococcus*	*opacus*	PD630	pDD119	N/A
DMD132	*Escherichia*	*coli*	DH10B	pDD99	N/A
DMD133	*Rhodococcus*	*opacus*	PD630	pDD99	2 and 3
DMD143	*Escherichia*	*coli*	DH10B	pDD108	N/A
DMD144	*Escherichia*	*coli*	DH10B	pDD107	N/A
DMD146	*Rhodococcus*	*opacus*	PD630	pDD57	1
DMD153	*Escherichia*	*coli*	DH10B	PLJR965	N/A
DMD155	*Escherichia*	*coli*	DH10B	pDD120	N/A
DMD157	*Rhodococcus*	*opacus*	PD630	pDD120	4

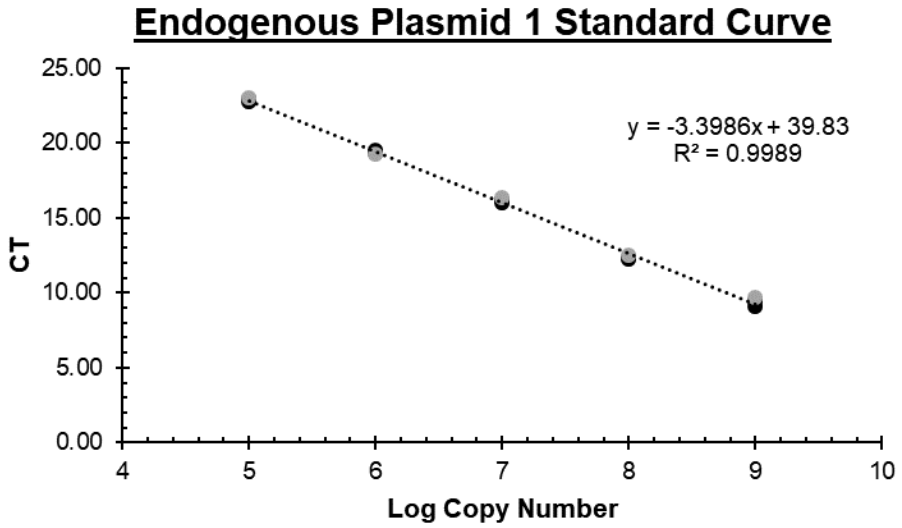
DMD160	*Escherichia*	*coli*	DH10B	pDD112	N/A
DMD161	*Rhodococcus*	*opacus*	PD630	pDD112	3
DMD175	*Rhodococcus*	*opacus*	PD630	pDD82	2
DMD179	*Rhodococcus*	*opacus*	PD630	Integrated pDD42	5
DMD180	*Rhodococcus*	*opacus*	PD630	Integrated pDD107	5
DMD181	*Rhodococcus*	*opacus*	PD630	Integrated pDD108	5 and 6
DMD199	*Escherichia*	*coli*	DH10B	pDD132	N/A
DMD207	*Rhodococcus*	*opacus*	PD630	Integrated pDD132	5
DMD221	*Escherichia*	*coli*	DH10B	pDD143	N/A
DMD223	*Escherichia*	*coli*	DH10B	pDD145	N/A
DMD225	*Escherichia*	*coli*	DH10B	pDD147	N/A
DMD227	*Rhodococcus*	*opacus*	PD630	pDD143 + integrated pDD108	6
DMD228	*Rhodococcus*	*opacus*	PD630	pDD145 + integrated pDD108	6
DMD229	*Rhodococcus*	*opacus*	PD630	pDD147 + integrated pDD108	6
DMD233	*Escherichia*	*coli*	DH10B	pXYLA	N/A
EPC1	*Escherichia*	*coli*	DH10B	pDD43	N/A
EPC2	*Escherichia*	*coli*	DH10B	pDD44	N/A
EPC3	*Escherichia*	*coli*	DH10B	pDD45	N/A
RPC1 through RPC25	*Rhodococcus*	*opacus*	PD630	Selected constitutive promoter-GFP+ constructs	1

Supplementary Table C.10. Fitted Hill equation parameters for Figure 2.6. The modified Hill equation, described below in C.3 Supplementary Methods, was fit to each set of normalized fluorescence (au) or percent repression data by minimizing the root mean square error (RMSE). F_{max} and F_{min} represent the fitted maximum and minimum normalized fluorescence (au), respectively. PR_{max} and PR_{min} represent the fitted maximum and minimum percent repression, respectively. The fitted Hill coefficient (n), half-maximal constant (K), and RMSE are also reported.

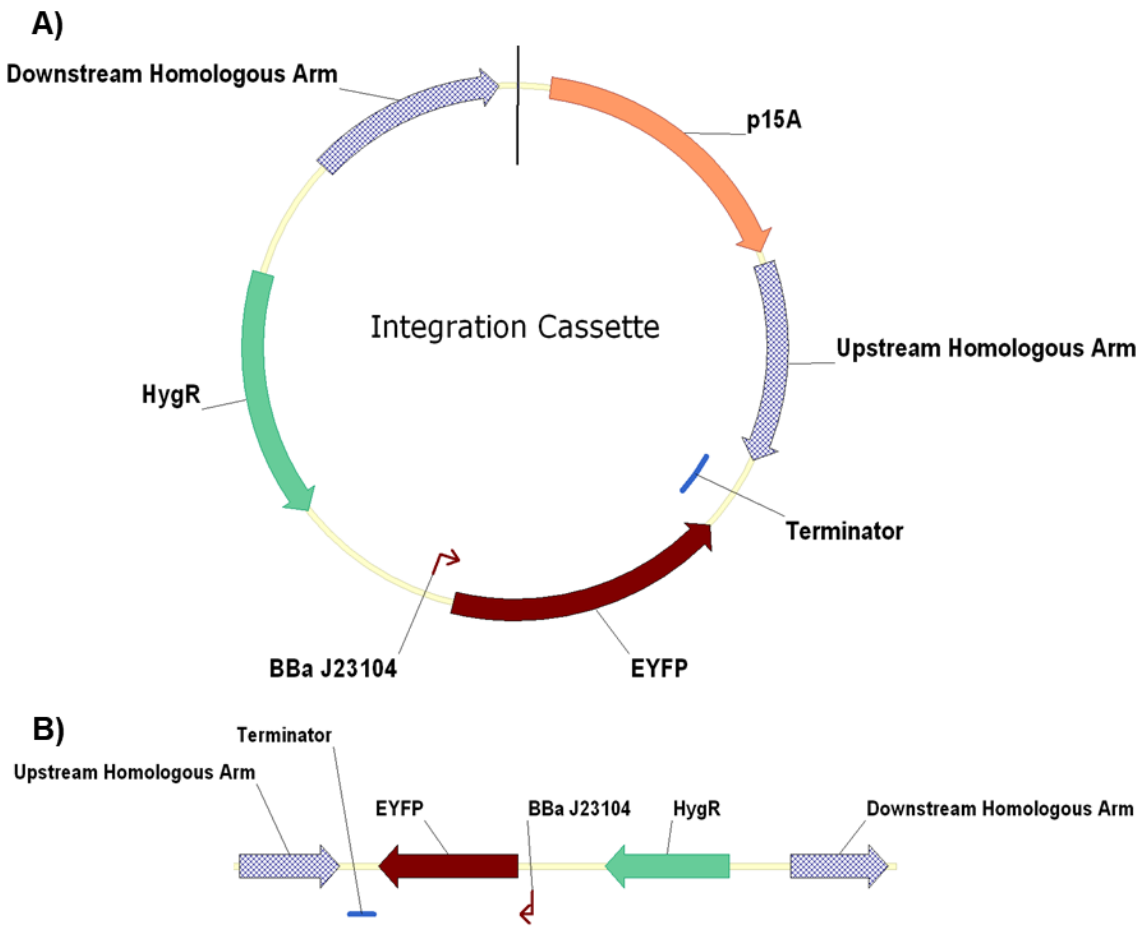
Figure	**Strain**	**F_{max} or PR_{max}**	**F_{min} or PR_{min}**	**n**	**K**	**RMSE**
2.6A	pBAD.dCas9$_{Sth}$.pCon.sgRNA.A	988 (au)	564 (au)	1.163	2.460	7.237
2.6A	pBAD.dCas9$_{Sth}$.pCon.sgRNA.B	1301 (au)	772 (au)	1.658	2.895	11.778
2.6A	pBAD.dCas9$_{Sth}$.pCon.sgRNA.C	957 (au)	624 (au)	1.413	6.358	13.962
2.6B	pBAD.dCas9$_{Sth}$.pCon.sgRNA.A	58.4	27.1	1.163	2.460	0.534

2.6B	pBAD.dCas9$_{Sth}$.pCon.sgRNA.B	43.1	4.1	1.658	2.895	0.869
2.6B	pBAD.dCas9$_{Sth}$.pCon.sgRNA.C	54.0	29.4	1.413	6.358	1.030

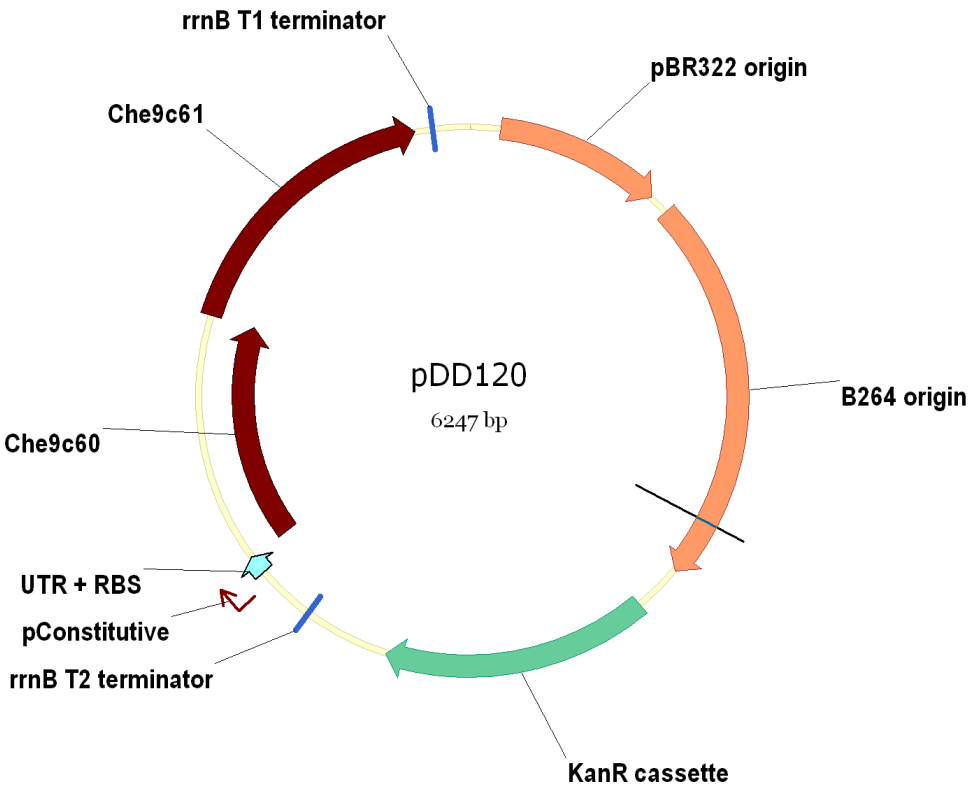
C.2 Supplementary figures



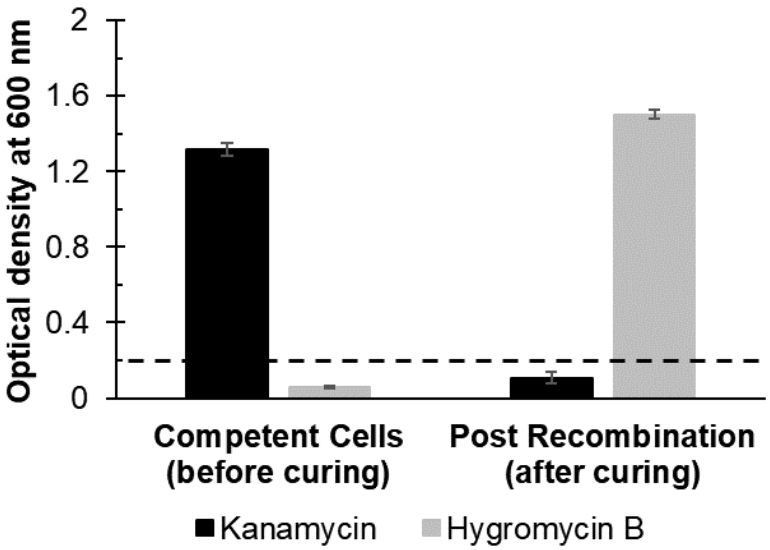
Supplementary Figure C.1. Representative qPCR standard curve. Data for endogenous plasmid 1 is shown as an example. The calculated PCR efficiency was 97% (see Materials and Methods for the corresponding equation). Experiment was conducted in duplicate (black and grey points). The dashed line represents a linear regression of the combined data set. The efficiency values for all primer pairs and targets are listed in Supplementary Table C.3.



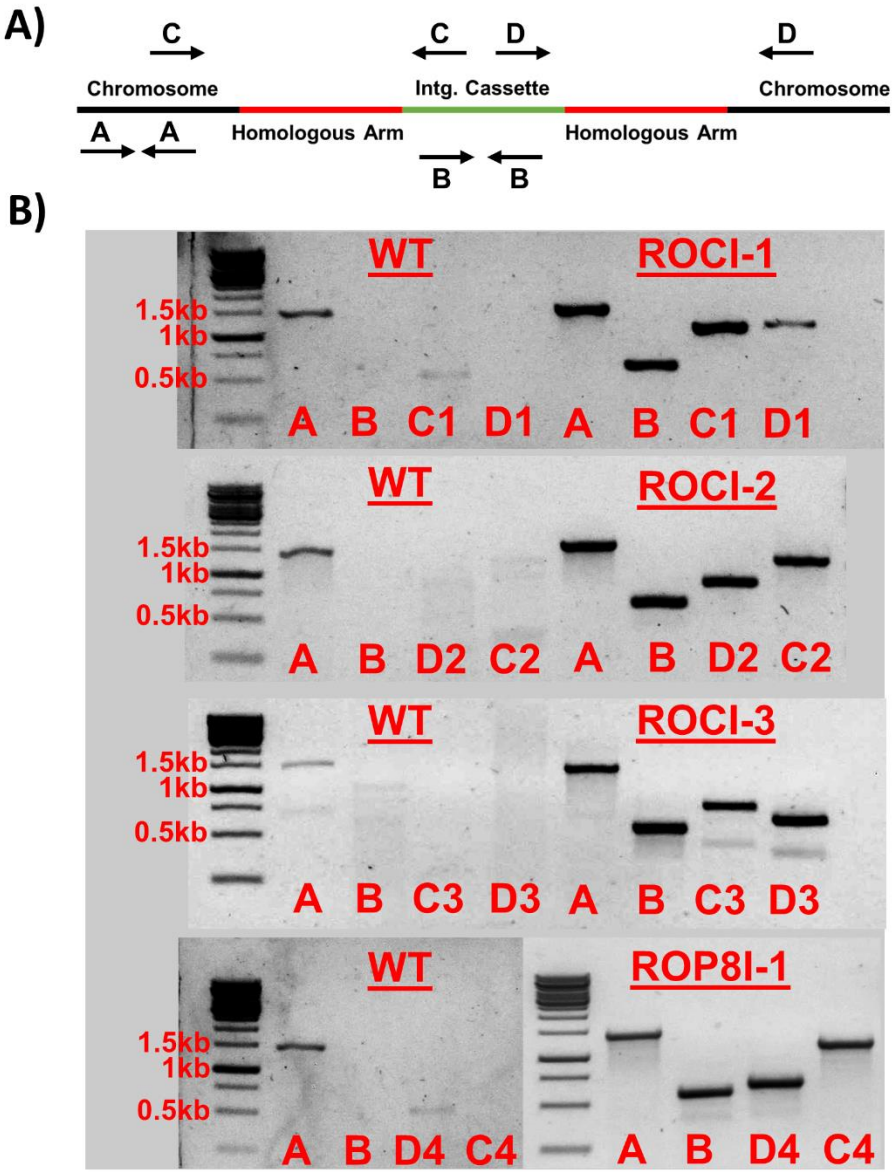
Supplementary Figure C.2. Representative integration cassette. **A)** A plasmid containing *E. coli* origin of replication (p15A), upstream and downstream homologous arms of ~500 base pairs, a constitutively expressed EYFP gene, and a HygR gene. **B)** The integration vector was linearized via PCR to better facilitate double homologous recombination.[6] All parts sequences are listed in Supplementary Table C.2.



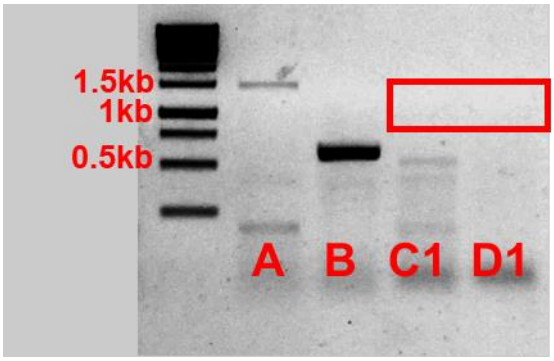
Supplementary Figure C.3. Vector map of the curable recombinase plasmid (pDD120). The bacteriophage recombinases Che9c60 and Che9c61 were constitutively expressed on the pB264 backbone. All parts sequences are listed in Supplementary Table C.2. The plasmid can be cured by removal of selection pressure and a round of colony purification.



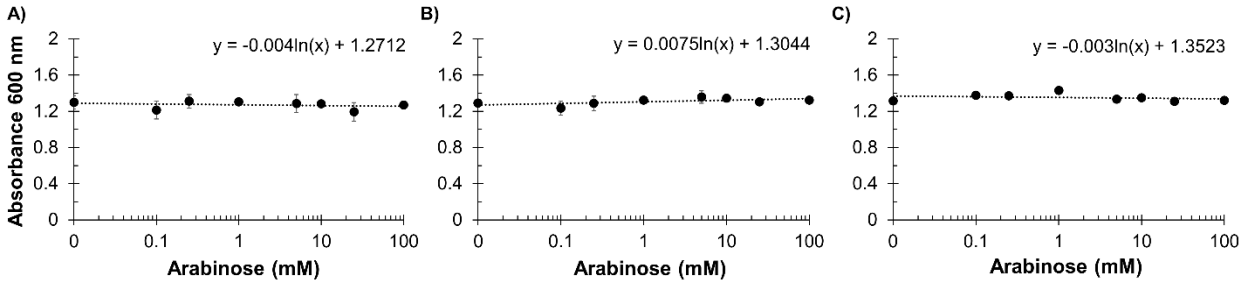
Supplementary Figure C.4. Curing of pDD120 during recombination experiments. The electro-competent cells containing the recombinases Che9c60 and Che9c61 on the pB264 backbone (pDD120) were electroporated with a linearized DNA integration cassette (see Materials and Methods). The cells were plated on a TSB agar plate containing only hygromycin B (selection for the integration cassette). Thus, there was no antibiotic selection for the recombinase vector (kanamycin), which led to curing of the plasmid after a round of colony purification. The electro-competent cell strain and HygR integrated cell strain were both grown in minimal media B with either kanamycin or hygromycin B as designated. The dashed line represents the initial OD_{600} of 0.2. Bars represent the average of three replicates and error bars represent one standard deviation.



Supplementary Figure C.5. Confirmation of a successful integration event. **A)** A schematic of primer binding locations for primer sets A through D. Primer set A binds at a location external to the integration site and was used to confirm successful cell lysis and that an adequate amount of DNA was present in the PCR reaction. Primer set B binds within the core region of the integration cassette and determines if the cassette was integrated, although it does not confirm whether integration at the correct site occurred. Primer set C spans the upstream junction of the integration site. Primer set D spans the downstream junction of the integration site. Primer sets A and B are universal for all tested strains, while sets C and D are for specific integration sites as designated by number (1-4). All expected band sizes are listed in Supplementary Table C.5. **B)** Images of electrophoresis gels confirming successful integration. Gel image colors were inversed for enhanced readability. All images include that of a wild type strain (WT) control. See Supplementary Table C.5 for full explanation of gel lane labels and correct PCR band sizes.



Supplementary Figure C.6. Illegitimate homologous recombination. The ROCI-1 integration cassette was transformed into wild type cells (no Che9c recombinases present). The cassette successfully recombines with the genome, but not in the correct location. The integration cassette is present (B), but it failed to integrate at site ROCI-1 (C1 and D1 bands which should be inside the red box are missing [see Supplementary Table C.5 for band sizes]). Gel image colors were inversed for enhanced readability.



Supplementary Figure C.7. CRISPRi culture absorbance at 600 nm. The cell density of each culture was estimated using the absorbance at 600 nm. The absorbance data for sgRNAs A, B, and C across all arabinose induction concentrations are reported. The positive control had a comparable absorbance at 600 nm of 1.19 ± 0.01. Values represent the average of three replicates grown in minimal media B, and error bars represent one standard deviation. The dashed line represents a logarithmic regression and the equation of each respective line is reported. Induction of $dCas9_{Sth}$ with arabinose had no impact on growth of any of the strains (|logarithmic coefficient| < 0.01).

C.3 Supplementary methods

C.3.1 Propagated standard deviation calculation for plasmid copy number

The propagated standard deviation for each plasmid copy number ratio was calculated according to the following equation to account for the respective standard deviations of both the chromosome and plasmid copy numbers, which were measured in triplicate.

$$Propagated\ standard\ deviation\ of\ copy\ number\ ratio = \frac{\mu_P}{\mu_C} * \sqrt{\left(\frac{\sigma_P}{\mu_P}\right)^2 + \left(\frac{\sigma_C}{\mu_C}\right)^2}$$

$\mu_P = mean\ of\ absolute\ plasmid\ copy\ number$
$\mu_C = mean\ of\ absolute\ chromosome\ copy\ number$
$\sigma_P = standard\ deviation\ of\ absolute\ plasmid\ copy\ number$
$\sigma_C = standard\ deviation\ of\ absolute\ chromosome\ copy\ number$

C.3.2 Propagated standard deviation for percent repression

The propagated standard deviation for percent repression was calculated according to the following equation to account for the respective standard deviations of both the CRISPRi and positive control (EYFP only) culture fluorescence values, which were measured in triplicate.

$$Propagated\ standard\ deviation\ of\ percent\ repression = 100 * \frac{\mu_I}{\mu_U} * \sqrt{\left(\frac{\sigma_I}{\mu_I}\right)^2 + \left(\frac{\sigma_U}{\mu_U}\right)^2}$$

$\mu_I = mean\ of\ CRISPRi\ culture\ normalized\ fluorescence$
$\mu_U = mean\ of\ positive\ control\ normalized\ fluorescence$
$\sigma_I = standard\ deviation\ of\ CRISPRi\ culture\ normalized\ fluorescence$
$\sigma_U = standard\ deviation\ of\ positive\ control\ culture\ normalized\ fluorescence$

C.3.3 Hill equation fitting

The Hill equation was adapted from DeLorenzo et al. (2017), to fit a line to both the normalized fluorescence and percent repression data.[8] The model was fit to the experimentally collected data such that the RMSE was minimized (Microsoft Excel Solver GRG Nonlinear). Fitted values are listed in Supplementary Table 10.

For CRISPRi normalized fluorescence plots:

$$F = \frac{(F_{max} - F_{min}) * K^n}{Inducer^n + K^n} + F_{min}$$

$F = Calculated\ fluorescence$
$F_{max} = Maximum\ normalized\ fluorescence$
$F_{min} = Minimum\ normalized\ fluorescence$
$K = Half\ maximal\ constant$
$n = Hill\ coefficient$
$Inducer = Arabinose\ inducer\ concentration\ (mM)$

Root mean square error (RMSE):

$$\sqrt{\frac{\sum_{i=1}^{n}(F - F_{obs})^2}{n - 2}}$$

$F_{obs} = Observed\ fluorescence$
$n = Number\ of\ data\ points$

For CRISPRi percent repression plots:

$$PR = \frac{(PR_{max} - PR_{min}) * Inducer^n}{Inducer^n + K^n} + PR_{min}$$

$PR = Calculated\ percent\ repression$
$PR_{max} = Maximum\ percent\ repression$
$PR_{min} = Minimum\ percent\ repression$
$K = Half\ maximal\ constant$
$n = Hill\ coefficient$
$Inducer = Arabinose\ inducer\ concentration\ (mM)$

Root mean square error (RMSE):

$$\sqrt{\frac{\sum_{i=1}^{n}(PR - PR_{obs})^2}{n-2}}$$

$PR_{obs} = Observed\ fluorescence$
$n = Number\ of\ data\ points$

C.4 References

1. Siegl, T., Tokovenko, B., Myronovskyi, M., and Luzhetskyy, A. (2013) Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes, Metabolic engineering 19, 98-106.

2. Kurosawa, K., Wewetzer, S. J., and Sinskey, A. J. (2013) Engineering xylose metabolism in triacylglycerol-producing Rhodococcus opacus for lignocellulosic fuel production, Biotechnology for biofuels 6, 134.

3. Xiong, X., Wang, X., and Chen, S. (2012) Engineering of a xylose metabolic pathway in Rhodococcus strains, Applied and environmental microbiology 78, 5483-5491.

4. Ehrt, S., Guo, X. V., Hickey, C. M., Ryou, M., Monteleone, M., Riley, L. W., and Schnappinger, D. (2005) Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor, Nucleic acids research 33, e21.

5. Lessard, P. A., O'Brien, X. M., Currie, D. H., and Sinskey, A. J. (2004) pB264, a small, mobilizable, temperature sensitive plasmid from Rhodococcus, BMC microbiology 4, 15.

6. van Kessel, J. C., and Hatfull, G. F. (2007) Recombineering in Mycobacterium tuberculosis, Nat. Methods 4, 147-152.

7. Immethun, C. M., DeLorenzo, D. M., Focht, C. M., Gupta, D., Johnson, C. B., and Moon, T. S. (2017) Physical, chemical, and metabolic state sensors expand the synthetic biology toolbox for Synechocystis sp. PCC 6803, Biotechnol. Bioeng. 114, 1561-1569.

8. DeLorenzo, D. M., Henson, W. R., and Moon, T. S. (2017) Development of Chemical and Metabolite Sensors for Rhodococcus opacus PD630, ACS synthetic biology 6, 1973-1978.

9. Rock, J. M., Hopkins, F. F., Chavez, A., Diallo, M., Chase, M. R., Gerrick, E. R., Pritchard, J. R., Church, G. M., Rubin, E. J., Sassetti, C. M., Schnappinger, D., and Fortune, S. M. (2017) Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform, Nature microbiology 2, 16274.

10. Immethun, C. M., Ng, K. M., DeLorenzo, D. M., Waldron-Feinstein, B., Lee, Y. C., and Moon, T. S. (2015) Oxygen-responsive genetic circuits constructed in synechocystis sp. PCC 6803, Biotechnol. Bioeng. 113, 433-442.

Appendix D: Supplementary materials for selection of stable reference genes for RT-qPCR in *Rhodococcus opacus* PD630

D.1 Supplementary tables

Supplementary Table D.1. Candidate reference gene (RG) primer sequences. The forward and reverse oligonucleotide sequences used for RT-qPCR for each respective RG candidate. Also listed are the predicted melting temperature (T_M) for each respective oligonucleotide and the annealing temperature (T_A) for each oligonucleotide pair as predicted by the ThermoFisher's Tm calculator (https://www.thermofisher.com/us/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html). See "RT-qPCR primer design criteria" for how oligonucleotide sequences were selected.

Reference gene (RG)	Primer sequence	T_M	T_A
RG1	F: CCCGCCGAGTCCGTGTTGTTCTTG R: CGACAGCCGAGTGCGACAACCTCATC	65 °C 67 °C	60 °C
RG2	F: CCTGACCTGTCGGCACGAATGAGC R: CCTTCACGGCATCCCCAACGC	65 °C 65 °C	60 °C
RG3	F: ATTCCGTGAGTAGTGGCGAGCGAAAGC R: CCACAACCCCACGAATGCAACACCTG	66 °C 65 °C	60 °C
RG4	F: GCTGTCGGCTGAGGTCGCCATC R: CTGCTGCGCCATCTTCACCATGTCG	66 °C 65 °C	60 °C
RG5	F: CTCACCTTCCGACGTGACGCTGTC R: GCGACGATTGTGGCGGCATCACTC	65 °C 66 °C	60 °C
RG6	F: GCCCACGCACCTCGTCGTCG R: GACCTTGACGCCCATCTCGGTGTAGG	67 °C 66 °C	61 °C
RG7	F: GCGTCCGTGGTGAACTCCAACTCC R: TCCAGATCCTCTCCGAGCCGAAGAAC	65 °C 65 °C	60 °C
RG8	F: CTCGGAGGAAGGTGGGGACGACG R: CCTCACGGTATCGCAGCCCTCTG	65 °C 65 °C	60 °C
RG9	F: TGGATAAGCGGCAGCGACCACTTGG R: GGAGTCGGGTGTGGTGAAGGAAGC	66 °C 65 °C	60 °C
RG10	F: CGTCGTCGGTGGGTGTTGCATGTC R: TCAGGATGCCGTAGCACCTCGACTG	65 °C 65 °C	60 °C

Supplementary Table D.2. Ranking of candidate reference genes (RGs) by CT standard deviation. The ten candidate RGs were ranked by the standard deviation of their C_T values as calculated by Bestkeeper. Analysis was performed on the pooled data set containing all four growth conditions. A standard deviation greater than 1 is considered unstable. The minimum and maximum C_T values for each RG are also provided.

Rank	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
	RG4	RG6	RG2	RG10	RG9	RG1	RG5	RG3	RG8	RG7
min [C_T]	22.93	25.47	25.86	28.34	28.32	23.44	24.56	8.74	10.48	24.93
max [C_T]	27.41	29.01	28.20	30.61	30.33	25.23	25.96	10.16	11.74	25.77
std dev [± C_T]	1.37	1.13	0.67	0.61	0.46	0.40	0.37	0.33	0.29	0.21

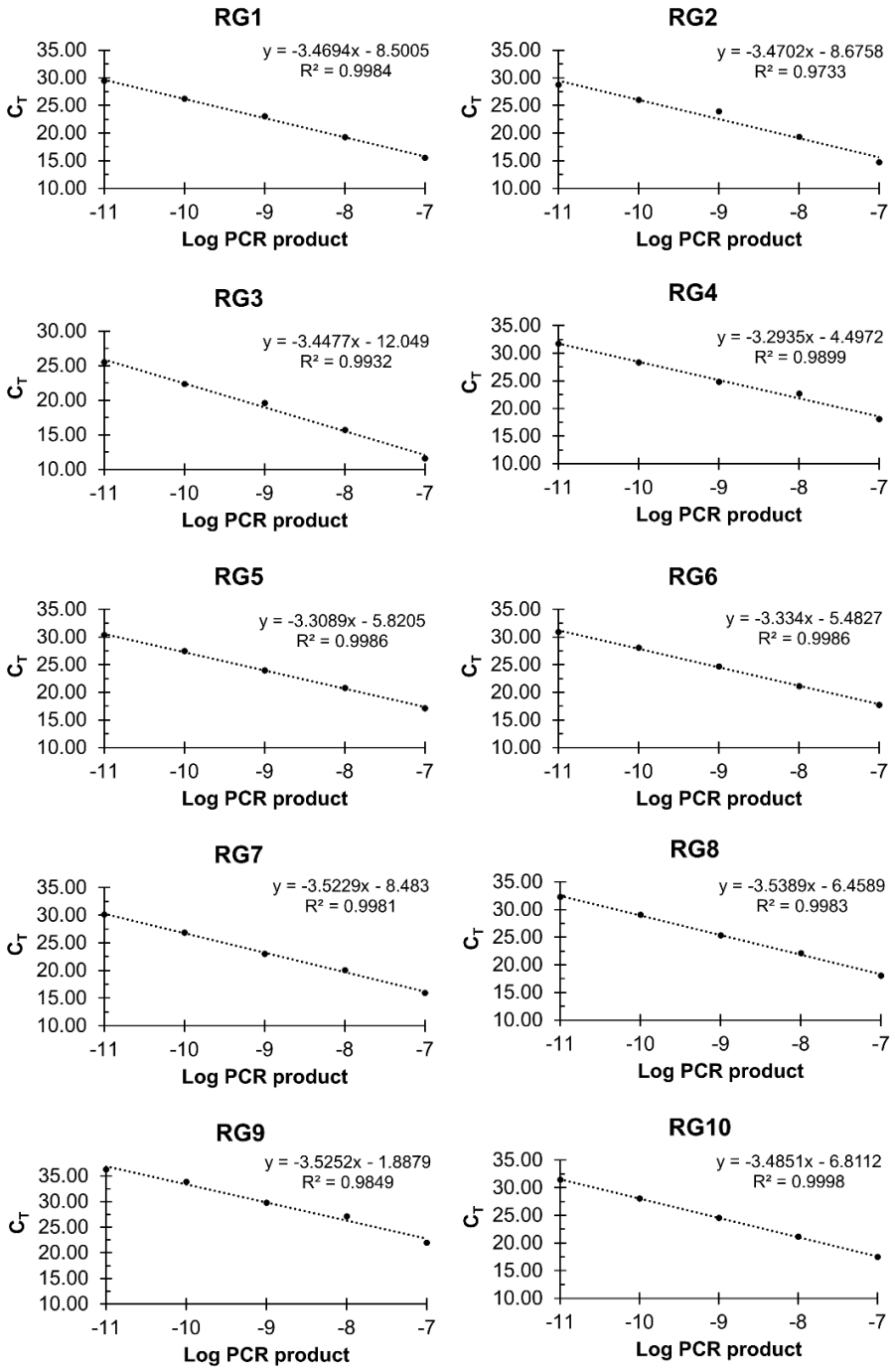
Supplementary Table D.3. Bestkeeper ranking with significance value. The ten candidate RGs were ranked by their Bestkeeper r-value calculated on the pooled data set containing the results from all four growth conditions in biological triplicates. Three technical replicates for each biological replicate were averaged prior to input into Bestkeeper. Bolded values represent a p-value < 0.05.

Rank	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
Gene	RG10	RG1	RG5	RG2	RG4	RG9	RG6	RG8	RG7	RG3
r-value	-0.192	-0.046	-0.006	0.326	**0.609**	**0.634**	**0.731**	**0.814**	**0.847**	**0.895**

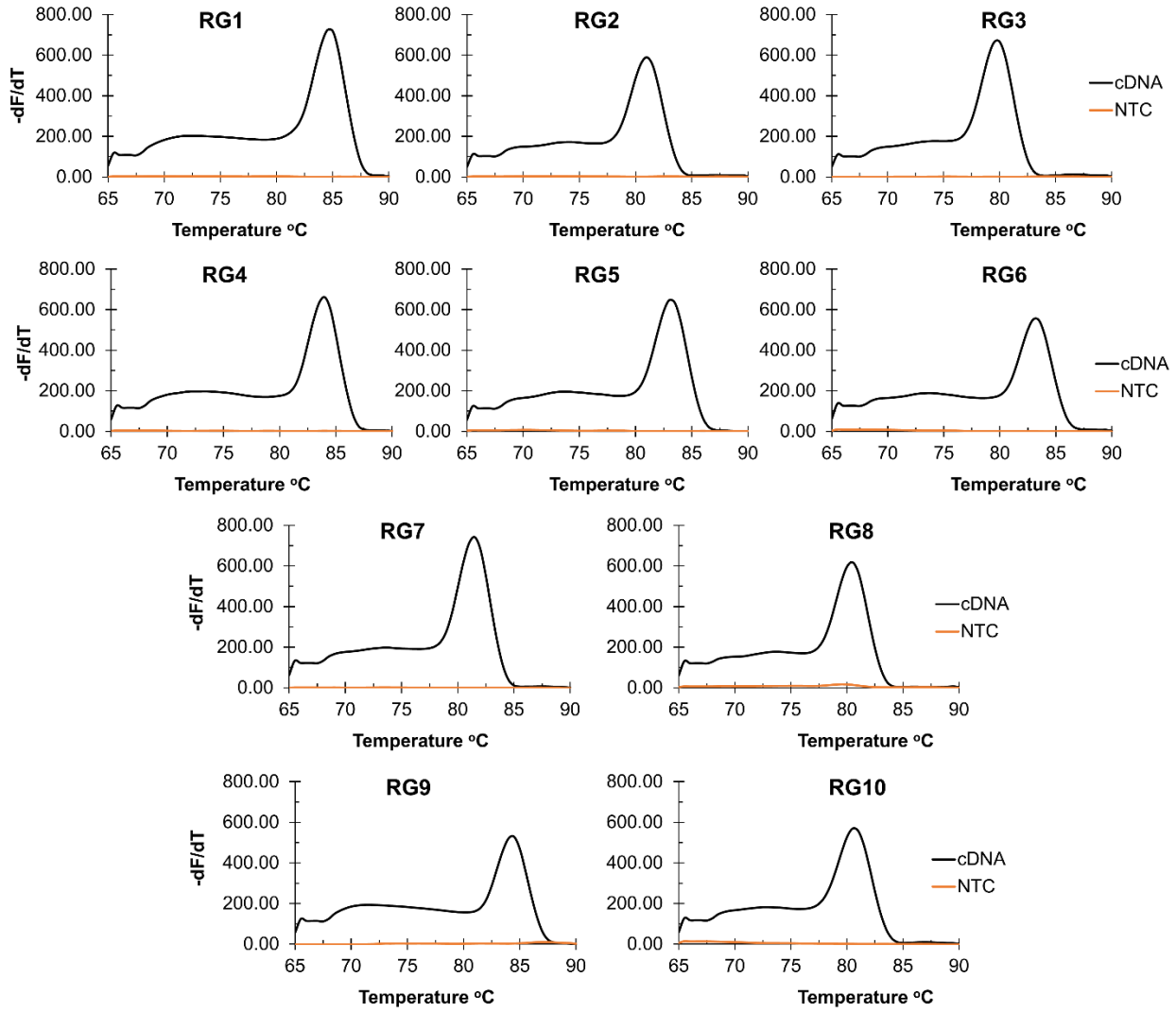
Supplementary Table D.4. Bestkeeper ranking with significance value with rRNA candidates removed. The eight candidate RGs were ranked by their Bestkeeper r-value calculated on the pooled data set containing the results from all four growth conditions in biological triplicates. Three technical replicates for each biological replicate were averaged prior to input into Bestkeeper. Bolded values represent a p-value < 0.05.

Rank	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
Gene	RG10	RG1	RG5	RG2	RG4	RG9	RG7	RG6
r-value	-0.122	-0.043	-0.001	0.418	0.532	**0.624**	**0.716**	**0.845**

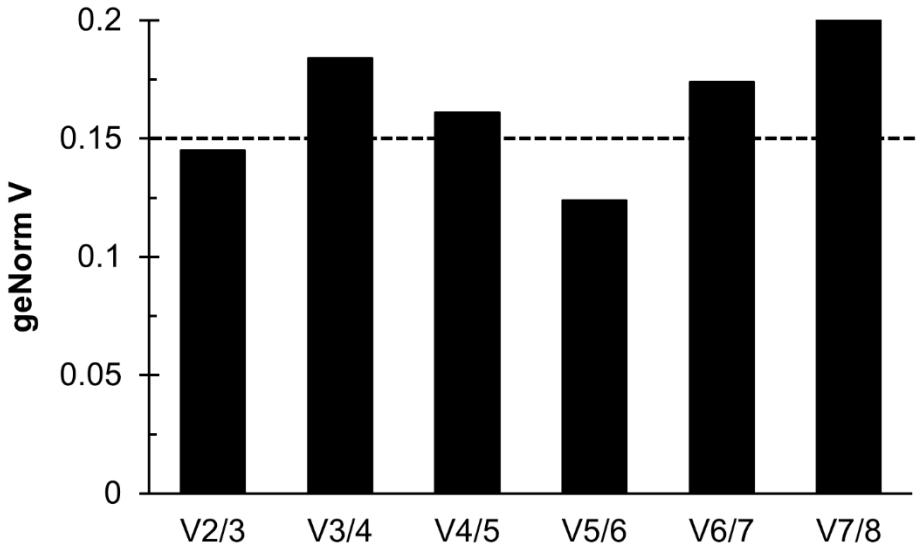
D.2 Supplementary figures



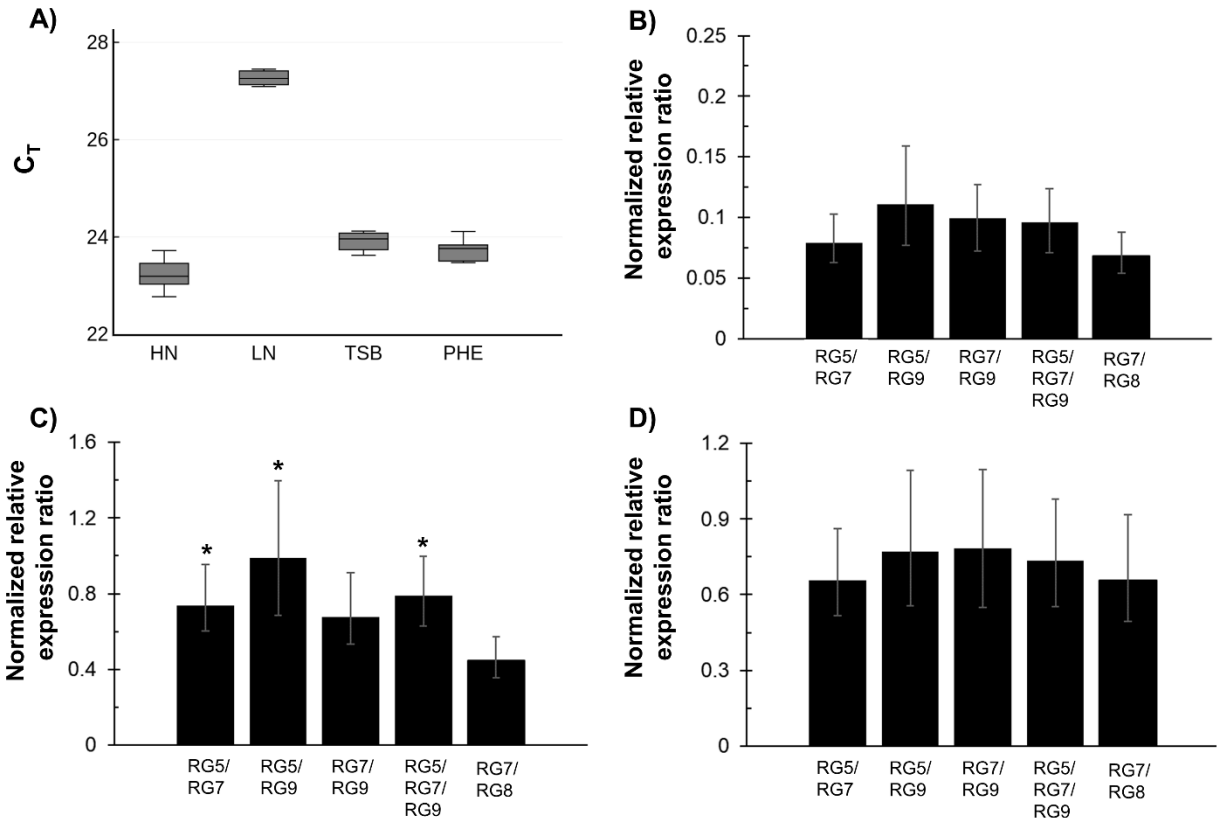
Supplementary Figure D.1. Standard curves for RT-qPCR primers. A five-round 10-fold serial dilution was performed on purified PCR product for each RT-qPCR primer pair. Each data point represents the average of duplicate values. A linear regression (dashed line) was performed for each data set, and the equation of the line and R^2 value are listed on each respective plot.



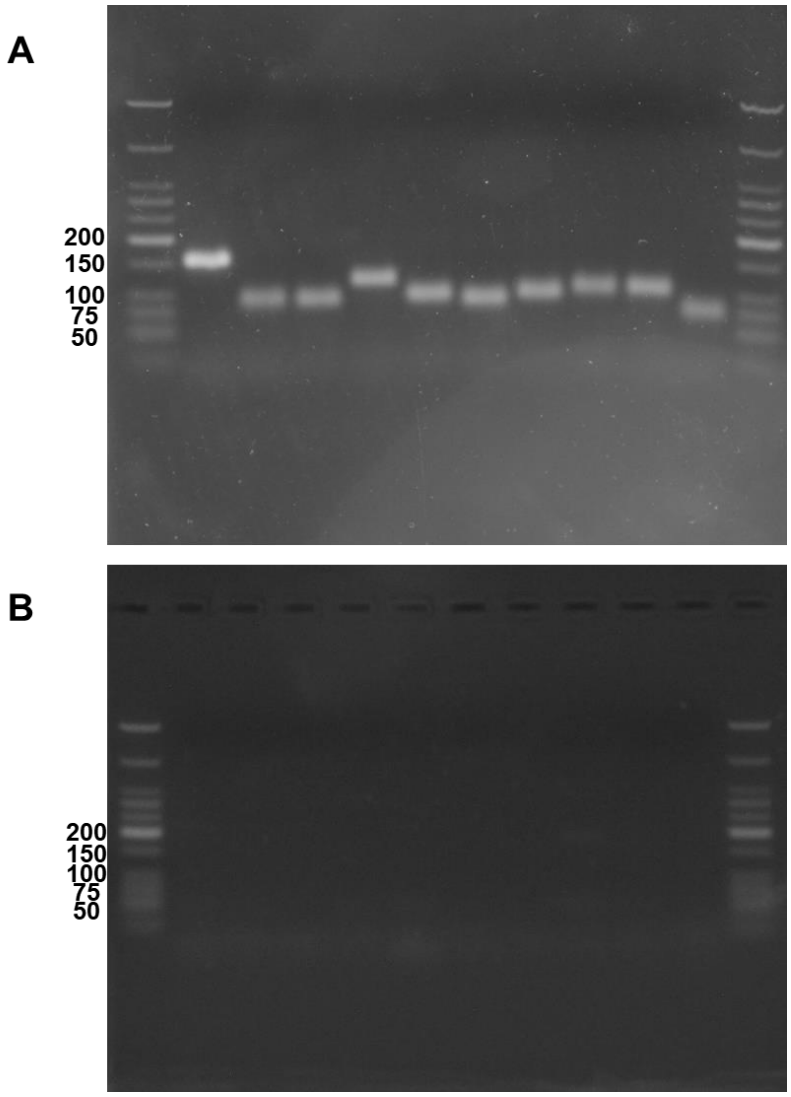
Supplementary Figure D.2. Melt curves for candidate reference gene amplicons. A melting curve analysis was performed on samples using the BioRad CFX96 after RT-qPCR to confirm a single amplicon was produced. The cDNA condition (black line) had cDNA added to the PCR reaction, while the NTC (no template control; orange line) condition had H_2O added instead.



Supplementary Figure D.3. Minimum number of reference genes (RGs) with rRNA removed. The pair-wise variation Vn/Vn+1, where n represents the number of RGs used in the normalization factor (NF), was calculated by geNorm to determine the minimum number of RGs required for normalization. A geNorm V value below 0.15 signifies that no additional benefit is gained from increasing the number of reference genes from n to n+1. The dashed line represents a V value of 0.15.



Supplementary Figure D.4. Effect of reference gene choice on RT-qPCR normalization with rRNA candidates removed. **A)** Box plots of averaged PD630_RS15810 expression data (C_T) for all four growth conditions (HN, LN, TSB, and PHE). Each gray box represents the first through third quartiles, the solid black line represents the median, and the whiskers represent the minimum and maximum values. **B-D)** The normalized relative expression ratio of PD630_LPD05540 going from HN to either LN (B), TSB (C), or PHE (D). The expression data was normalized with either RG5/RG7, RG5/RG9, RG7/RG9, RG5/RG7/RG9, or RG7/RG8 using REST 2009. Error bars represent the 95% confidence interval (CI) as calculated by REST 2009. Stars indicate that a 95% CI range falls outside of the 95% CI range of the RG7/RG8 normalized ratio.



Supplementary Figure D.5. The original version of electrophoresis gel images. **A)** The original non-edited, full length electrophoresis gel image showing single amplicon bands after RT-qPCR for each reference gene candidate (RG 1 to RG 10; left to right). **B)** The original non-edited, full length electrophoresis gel image confirming no bands for the negative template controls (NTCs) after RT-qPCR for each gene candidate (RG 1 to RG 10; left to right). The sizes for the nucleotide ladder are indicated to left of bands (50 to 200 bp). The size of each amplicon is denoted in Table 5.1.

D.3 RT-qPCR primer design criteria

Primer annealing temperature = ~60 ^{o}C when using 125 nM primer (ThermoFisher Tm calculator)

Amplicon length = 70-160 bp

Try to have only one G/C at 3’ end. No more than 2 G/Cs

Two or fewer consecutive G/C pairs in predicted hairpin structure (IDT OligoAnalyzer 3.1; https://www.idtdna.com/calc/analyzer); avoid binding on 3’ end

Primer hairpin = $\Delta G > -2$ kcal/mol

Primer homodimer = $\Delta G > -7$ kcal/mol

Primer heterodimer = $\Delta G > -7$ kcal/mol

Appendix E: Summary of all published graduate work

2019

- Chatterjee, A.*, **DeLorenzo, D.M.***, Moon, T.S. Bioconversion of renewable feedstocks by *Rhodococcus opacus*. *Current Opinion in Biotechnology (2019)*. *** = co-first authorship**
- **DeLorenzo, D.M.** and Moon, T.S. Construction of genetic logic gates based on the T7 RNA polymerase expression system in *Rhodococcus opacus* PD630. *ACS Synthetic Biology* (2019). 8(8), 1921-1930.
- Anthony, W.*, Carr, R.R.*, **DeLorenzo, D.M.**, Campbell, T., Shang, Z., Foston, M., Moon, T.S., Dantas, G. Development of *Rhodococcus opacus* as a chassis for lignin valorization and bioproduction of high-value compounds. *Biotechnology for Biofuels (*2019*)*. 12(192).

2018

- Henson, W.R.*, Campbell, T.*, **DeLorenzo, D.M.*,** Gao, Y., Berla, B., Kim, S.J., Foston, M., Moon, T.S., Dantas, G. Multi-omic elucidation of aromatic catabolism in adaptively evolved *Rhodococcus opacus*. *Metabolic Engineering* (2018). 49, 69-83. *** = co-first authorship**
- **DeLorenzo, D.M.** and Moon, T.S. Selection of stable reference genes for RT-qPCR in *Rhodococcus opacus* PD630. *Scientific Reports* (2018). 8, 6019.
- **DeLorenzo, D.M.,** Rottinghaus, A.G., Henson, W.R., Moon, T.S. Molecular toolkit for gene expression control and genome modification in *Rhodococcus opacus* PD630. *ACS Synthetic Biology* (2018). 7(2), 727−738.

2017

- **DeLorenzo, D.M.,** Henson, W.R., Moon, T.S. Development of Chemical and Metabolite Sensors for *Rhodococcus opacus* PD630. *ACS Synthetic Biology* (2017). 6(10), 1973–1978
- Immethun, C.M., **DeLorenzo, D.M.,** Focht, C.M., Gupta, D., Johnson, C.B., Moon, T.S. Physical, Chemical, and Metabolic State Sensors Expand the Synthetic Biology Toolbox for *Synechocystis* sp. PCC 6803. *Biotechnology and Bioengineering* (2017). 114(7):1561-1569.
- Wan, N. *, **DeLorenzo, D.M.***, Hel, L., Youl, L.*, Immethun, C.M., Wang, G., Baidoo, E.E.K., Hollinshead, W., Keasling, J.D., Moon, T.S., Tang, Y.J. Cyanobacterial carbon metabolism: fluxome plasticity and oxygen dependence. *Biotechnology and Bioengineering* (2017). 114(7):1593-1602 *** = co-first authorship**

2016

- Immethun, C.M., Ng, K., **DeLorenzo, D.M.,** Waldron-Feinstein, B., Lee, J., Moon, T.S. Oxygen-Responsive Genetic Circuits Constructed in *Synechocystis* sp. PCC 6803. *Biotechnology and Bioengineering (*2016). 113(2):433-42.