2017

Bioprospecting For Genes That Confer Biofuel Tolerance To Escherichia Coli Using A Genomic Library Approach

Timothy Tomko
University of Vermont

BIOPROSPECTING FOR GENES THAT CONFER BIOFUEL TOLERANCE TO *ESCHERICHIA COLI* USING A GENOMIC LIBRARY APPROACH

A Dissertation Presented

by

Timothy Tomko

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
Specializing in Mechanical Engineering

October, 2017

Defense Date: August 03, 2017
Thesis Examination Committee:

Mary J. Dunlop, Ph.D., Advisor
Matthew Wargo, Ph.D., Chairperson
Jason Bates, Ph.D.
Rachael Oldinski, Ph.D.
Cynthia J. Forehand, Ph.D., Dean of the Graduate College

ABSTRACT

Microorganisms are capable of producing advanced biofuels that can be used as 'drop-in' alternatives to conventional liquid fuels. However, vital physiological processes and membrane properties are often disrupted by the presence of biofuel and limit the production yields. In order to make microbial biofuels a competitive fuel source, finding mechanisms for improving resistance to the toxic effects of biofuel production is vital. This investigation aims to identify resistance mechanisms from microorganisms that have evolved to withstand hydrocarbon-rich environments, such as those that thrive near natural oil seeps and in oil-polluted waters.

First, using genomic DNA from *Marinobacter aquaeolei*, we constructed a transgenic library that we expressed in *Escherichia coli*. We exposed cells to inhibitory levels of pinene, a monoterpene that can serve as a jet fuel precursor with chemical properties similar to existing tactical fuels. Using a sequential strategy of a fosmid library followed by a plasmid library, we were able to isolate a region of DNA from the *M. aquaeolei* genome that conferred pinene tolerance when expressed in *E. coli*. We determined that a single gene, *yceI*, was responsible for the tolerance improvements. Overexpression of this gene placed no additional burden on the host. We also tested tolerance to other monoterpenes and showed that *yceI* selectively improves tolerance.

Additionally, we used genomic DNA from *Pseudomonas putida* KT2440, which has innate solvent-tolerance properties, to create transgenic libraries in an *E. coli* host. We exposed cells containing the library to pinene, selecting for genes that improved tolerance. Importantly, we found that expressing the sigma factor RpoD from *P. putida* greatly expanded the diversity of tolerance genes recovered. With low expression of $rpoD_{P.\ putida}$, we isolated a single pinene tolerance gene; with increased expression of the sigma factor our selection experiments returned multiple distinct tolerance mechanisms, including some that have been previously documented and also new mechanisms. Interestingly, high levels of $rpoD_{P.\ putida}$ induction resulted in decreased diversity. We found that the tolerance levels provided by some genes are highly sensitive to the level of induction of $rpoD_{P.\ putida}$, while others provide tolerance across a wide range of $rpoD_{P.\ putida}$ levels. This method for unlocking diversity in tolerance screening using heterologous sigma factor expression was applicable to both plasmid and fosmid-based transgenic libraries. These results suggest that by controlling the expression of appropriate heterologous sigma factors, we can greatly increase the searchable genomic space within transgenic libraries.

This dissertation describes a method of effectively screening genomic DNA from multiple organisms for genes to mitigate biofuel stress and shows how tolerance genes can improve bacterial growth in the presence of toxic biofuel compounds. These identified genes can be targeted in future studies as candidates for use in biofuel production strains to increase biofuel yields.

CITATIONS

Material from this dissertation has been published in the following form:

Tomko, T. A., & Dunlop, M. J.. (2015). Engineering improved bio-jet fuel tolerance in Escherichia coli using a transgenic library from the hydrocarbon-degrader Marinobacter aquaeolei. *Biotechnology for Biofuels*, *8*(1), 165.

Tomko, T. A., & Dunlop, M. J.. (2017). Expression of heterologous sigma factor expands the searchable space for biofuel tolerance mechanisms. *ACS Synthetic Biology*.

ACKNOWLEDGEMENTS

I would like to acknowledge the many faculty members, staff, fellow students and family members who helped make my thesis possible. In particular I would like to thank Dr. Mary Dunlop for her guidance, support, and creative solutions as well as for helping me develop as a research scientist and engineer. I would also like to especially thank Dr. Matthew Wargo for his help over the past several months as I finished the remainder of my work using his facilities. I would like to thank my final two committee members, Dr. Jason Bates and Dr. Rachael Oldinski for their guidance and constructive criticism during my time at The University of Vermont. I also want to extend my appreciation to the members of the Dunlop Lab group for their support and advice, including Jesse Fenno who assisted me with running experiments near the end of the project. Finally, I would like to especially thank my family for their constant encouragement and my fiancée, Erica, for her unwavering support and confidence in me.

.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1. Microbes that Thrive in Harsh Environments

Microorganisms are capable of producing advanced biofuels that can be used as 'drop-in' alternatives to conventional petroleum-based liquid fuels (Fischer, Klein-Marcuschamer et al. 2008, Fortman, Chhabra et al. 2008, Lee, Chou et al. 2008). However, many of these fuels are toxic to cells, introducing an undesirable trade-off between cell survival and biofuel production. The primary problem is that hydrocarbons can disrupt key physiological processes and interfere with membrane permeability and function (Nicolaou, Gaida et al. 2010, Dunlop 2011). This inhibits cell growth and leads to reduced biofuel production. Finding ways to combat biofuel toxicity may increase yields of microbial biofuels (Tomas, Welker et al. 2003).

Many microorganisms survive naturally in hydrocarbon-rich environments (Rojas, Duque et al. 2001, Camilli, Reddy et al. 2010). We asked whether the genomes of these microbes could be a source of biofuel tolerance mechanisms when expressed heterologously in a biofuel production host. Microbes that tolerate hydrocarbons have been isolated near natural oil seepages and around oil spills. For example, a hydrocarbon degrading microbe, *Marinobacter aquaeolei* VT8, was isolated at the head of an offshore oil well in Vietnam (Huu, Denner et al. 1999). A previous study showed that this organism harbors two efflux pumps that can serve to improve biofuel tolerance (Dunlop, Dossani et al. 2011). That study also identified an efflux pump from *Alcanivorax borkumensis*, another hydrocarbon-degrader that thrives in oil-polluted waters. In

addition to the ability to metabolize hydrocarbons, *A. borkumensis* possess multiple tolerance mechanisms, including the production of biosurfactants, efflux pumps, and niche-specific stress responses (Schneiker, dos Santos et al. 2006). Studies of the microbial communities in the Gulf of Mexico after the Deepwater Horizon oil spill showed a high proportion of γ-Proteobacteria harboring hydrocarbon-degrading genes (Hazen, Dubinsky et al. 2010). Given the abundance of naturally hydrocarbon-tolerant microorganisms, we hypothesized that the genomes of these organisms may serve as an untapped reservoir of tolerance genes.

1.2. Bio-jet Fuels as Alternatives to Conventional Fuels

Conventional hydrocarbon based fuels are formed through the process known as catagenesis over the course of hundreds of thousands of years (Schobert 2013). Through this process, shown in Figure 1, hydrogen is either lost or gained to form molecules with varying chemical properties.

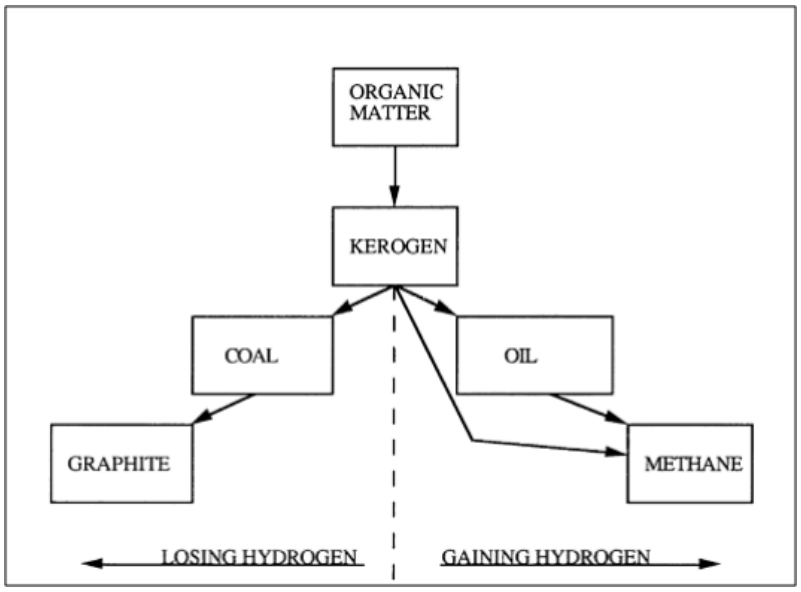


Figure 1. Illustration of catagenesis, the process by which hydrocarbons gain/lose hydrogen (Schobert 2013).

An increasing hydrogen to carbon (H:C) ratio means that the molecules becomes saturated with hydrogen and will continue this process until reaching the final state of methane, CH_4. Conversely, a decreasing H:C ratio means that hydrocarbon becomes more dense until reaching the final state of pure carbon which is known as graphite. Since the process of catagenesis takes place over a very long time period, intermediates with varying levels of H:C ratios are found. These intermediates include hydrocarbons like coal, oil or wax.

Crude oil is the starting material for many types of liquid fuels, however it must be refined via distillation before it can be used (Schobert 2013). The process of

distillation (Figure 2) separates the crude oil into several “cuts” based on their boiling points.

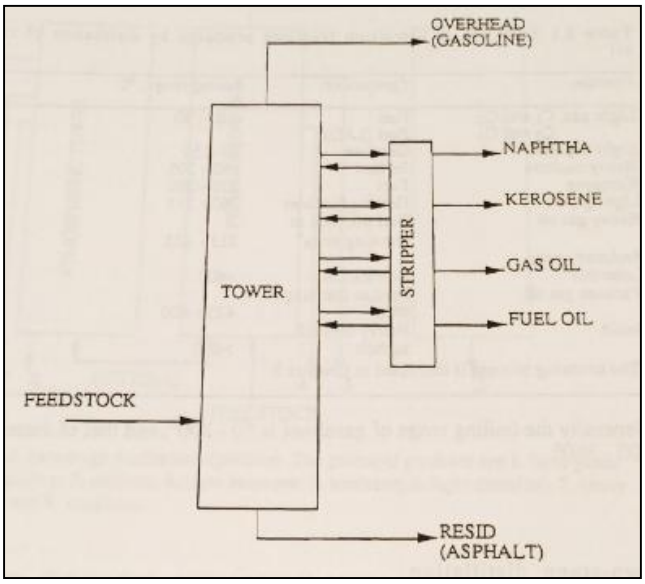


Figure 2. Illustration of a distillation tower separating hydrocarbons into a variety of “cuts” based on their boiling points (Schobert 2013).

Higher boiling point hydrocarbons like resid and fuel oil have very long and complex hydrocarbon chains. Gas oil (diesel fuel) and kerosene (jet fuel) have slightly lower boiling points and therefore come out of the distillation column above the fuel oil and resid. Finally, the lightest products, naphatha and overhead gasoline are all used downstream to make conventional gasoline products.

Biofuels are alternatives to conventional fossil fuels. Instead of allowing catagenesis to take place naturally, biofuels can be produced using microbes with sugar

as the feedstock. Specifically, monoterpenes like pinene, limonene, terpinene and terpinolene can be useful as replacements to conventional jet fuel.

In this study, we focused specifically on improving tolerance to bio-jet fuels. Gasoline and diesel fuel can be supplemented with the use of battery power in terrestrial vehicles, however this is not a feasible option for aircraft that have much more severe weight and size limitations. Aviation fuel must have a low enough freezing point that the fuel does not gel at low temperatures associated with typical flight altitudes. Also, the fuel must burn cleanly so as to not produce large amounts of soot that could potentially damage the turbine (Schobert 2013). These considerations make bio-jet fuels a promising, but challenging, class of biofuels to produce. Monoterpenes ($C_{10}H_{16}$) such as pinene, limonene, terpinene, and terpinolene are composed of two isoprene units (Schobert 2013) and have been shown to be excellent candidates for replacements to commercial Jet-A/A-1 fuels (Harvey, Wright et al. 2009, Brennan, Turner et al. 2012, Chuck and Donnelly 2014, Zhang, Hui et al. 2016). Importantly, several of these bio-jet fuels have been produced by engineered microbial hosts. For example, in 2014, Sarria, *et al*. reported a pinene production pathway in *E. coli* (Sarria, Wong et al. 2014, Tashiro, Kiyota et al. 2016). Other microbial hosts such as *Corynebacterium glutamicum* have also been engineered to produce pinene (Kang, Eom et al. 2014). Limonene production pathways have also been engineered in *E. coli* (Dunlop, Dossani et al. 2011, Alonso-Gutierrez, Chan et al. 2013, Alonso-Gutierrez, Kim et al. 2015). Recently, there has also been work done using cell free systems to produce monoterpenes. This has the potential to alleviate some of the toxicity problems encountered by the previously mentioned

production methods (Korman, Opgenorth et al. 2017). For general strategies on improving chemical pathway design and optimization see the recent review from Chubukov et al. (Chubukov, Mukhopadhyay et al. 2016).

1.3. Challenges with and Solutions to Increased Biofuel Production

A major challenge in the production of advanced biofuels is that toxicity limits the concentration of biofuel a cell can withstand. As mentioned previously, vital physiological processes and membrane properties are often disrupted by the presence of these hydrocarbons; for reviews on mechanisms of biofuel toxicity and engineering approaches to mitigating toxicity see (Ramos, Duque et al. 2002, Nicolaou, Gaida et al. 2010, Dunlop 2011, Peabody, Winkler et al. 2014, Mukhopadhyay 2015, Johnson, Gibbons et al. 2016). Previous tolerance engineering efforts have identified several ways that biofuel tolerance can be increased. For instance, export pumps, including efflux pumps and ABC transporters, can be used export harmful compounds from the cell (Fischer, Klein-Marcuschamer et al. 2008, Dunlop, Dossani et al. 2011, Doshi, Nguyen et al. 2013, Foo and Leong 2013, Boyarskiy, López et al. 2016). Natively, efflux pumps can act to protect cells in solvent rich environments (Ramos, Duque et al. 1998). Pumps can also improve biofuel tolerance when expressed heterologously in a production host. For instance, the TtgB, MexF, and Pp_3456 pumps from *Pseudomonas putida* KT2440 improve tolerance to pinene when expressed in *E. coli* (Dunlop, Dossani et al. 2011).

However, there are many examples where overexpression of efflux pumps can be detrimental (Dunlop, Keasling et al. 2010, Wood and Cluzel 2012, Harrison and Dunlop 2014), suggesting there are optimal expression levels for these pumps that should not be exceeded (Turner and Dunlop 2014). Additionally, several studies have shown that heat shock proteins can play an important role in improving solvent tolerance; recent examples include (Fiocco, Capozzi et al. 2007, Alsaker, Paredes et al. 2010, Reyes, Almario et al. 2011). Additionally, membrane modifications such the shift from *cis* to *trans* unsaturated fatty acids have been successful in preventing hydrocarbons from disrupting cell growth (Holtwick, Meinhardt et al. 1997, Junker and Ramos 1999). General stress response proteins can also mitigate biofuel toxicity and have appeared in screens for tolerance genes (Weber, Polen et al. 2005, Nicolaou, Gaida et al. 2010)

Genomic library approaches have been successful at selecting for genes that improve tolerance. In these selection experiments, cells harboring libraries are exposed to a stressor, such as biofuel, and those cells within the library that contain beneficial tolerance genes will display a competitive growth advantage. Both native and transgenic DNA have been used to create genomic libraries used to identify mechanisms that improve biofuel tolerance (Nicolaou, Gaida et al. 2010, Dunlop 2011). To begin this process, genomic DNA is either digested or sheared into fragment sizes appropriate for subcloning, inserted into the vector of choice, and transformed into the host. Cells containing members of the library are then subjected to a stressor over a period of time and those that survive are screened using microarrays or sequencing. Most experiments

related to biofuel tolerance that have used this approach use autologous libraries (for example, using genomic DNA from *E. coli* to screen in *E. coli*). Woodruff *et al.* used a multi-Scalar Analysis of Library Enrichments (SCALEs) approach to identify nine novel target genes to improve ethanol tolerance in *E. coli* (Woodruff, Pandhal et al. 2013). Other studies involving the overexpression of endogenous genes have shown improvements in isobutanol and ethanol tolerance in *Saccharomyces cerevisiae* (Hong, Lee et al. 2010), and ethanol (Nicolaou, Gaida et al. 2012) and n-butanol (Reyes, Almario et al. 2011) tolerance in *E. coli*. A small number of studies have used transgenic DNA libraries to improve tolerance in processes related to biofuel production. Zingaro *et al.* identified several tolerance mechanisms from *Lactobacillus plantarum,* that when expressed in *E. coli* show improved survival and growth under ethanol stress (Zingaro, Nicolaou et al. 2014). Ruegg *et al.* increased ionic liquid tolerance in an *E. coli* host by screening a fosmid library from *Enterobacter lignolyticus* for tolerance genes (Ruegg, Kim et al. 2014). Simarly, *yceI* was identified from *Marinobacter aquaeolei* as important in improving tolerance to monoterpenes in *E. coli* (Tomko and Dunlop 2015). Genomic libraries represent an efficient way of screening many genes for desirable traits and transgenic libraries extend this approach to a potentially rich resource of tolerance genes.

There are many examples of studies where screening for genes that improve tolerance to exogenous addition of biofuel has proved successful in identifying tolerance genes that also improve biofuel yields. For example, in a recent study of isopentanol tolerance in *E. coli*, six out of eight of the genes that were found to improve tolerance to

exogenous biofuel also increased production titers (Foo, Jensen et al. 2014). An efflux pump that was identified to improve tolerance to limonene also increased yield in a production strain (Dunlop, Dossani et al. 2011). A study performed in *Saccharomyces cerevisiae* showed that improved ethanol tolerance translates to a more efficient glucose to ethanol conversion rate (Alper, Moxley et al. 2006). However, there are counterexamples where improved tolerance does not lead to an increase in biofuel production. For example, Atsumi *et al.* identified mutations that improved isobutanol-tolerance in *E. coli*. However their mutants did not yield higher titers of the biofuel product (Atsumi, Wu et al. 2010). Due to the burden biofuel production pathways place on the cell, it is often more straightforward to screen for tolerance improvements with exogenous biofuel addition and later incorporate the tolerance strategy into a production strain. This approach has proved successful in many previous studies, and in this work we screen for monoterpene tolerance using exogenous addition of the biofuel.

Although heterologous genomic libraries are powerful tools for discovering novel tolerance mechanisms, a potential drawback of this approach is that the transcriptional machinery of the host may not recognize or efficiently transcribe genes controlled by the non-native promoters contained within the library (Warren, Freeman et al. 2008, Uchiyama and Miyazaki 2009). This problem can be fixed to some degree by using plasmid libraries where a strong or inducible promoter is added upstream of the cloning site to increase gene expression. However, there are a number of technical considerations such as the potential for incorrect insert orientation, the potential for many genes on an

insert, and strong terminators within the insert that can interfere with expression. Therefore, it can be preferable to use the native promoters associated with each gene in the library to control gene expression.

Sigma factors work in conjunction with RNA polymerase to initiate transcription. Different sigma factor families are expressed depending on the particular situation within the cell. For example, in many types of bacteria, σ^{70} (RpoD) is the primary sigma factor, responsible for aiding in the transcription of general housekeeping genes (Paget and Helmann 2003). If the promoter of a heterologous gene is compatible with the host, the host's native sigma factor enables RNA polymerase binding to the promoter. However, if the host's sigma factor is not compatible or only weakly compatible with the promoter, a non-native sigma factor can be added to the host to enable gene expression. Recently, this method of improving gene expression in a library was implemented by inserting the general sigma factor *rpoD* from *Lactobacillus plantarum* into *E. coli* to identify genes capable of improving ethanol tolerance (Gaida, Sandoval et al. 2015).

This dissertation aims to identify resistance mechanisms from microorganisms that have evolved to withstand extreme environmental conditions that can be used in engineered biofuel production strains. In Chapter 2 of this dissertation, our goal was to identify novel tolerance genes that allow *E. coli* to survive in high concentrations of bio-jet fuels. We screened the genome of *M. aquaeolei*, using a genomic library approach, for genes that would confer tolerance to pinene in an *E. coli* host. We identified one

tolerance gene that was instrumental in providing pinene tolerance in *E. coli.* Next in Chapter 3 we started working with a *P. putida* library and began including the heterologous sigma factor *rpoD* to increase the library expression within *E. coli*. This approach successfully increased the number of positive tolerance genes we were able to isolate from our library. Finally Chapter 4 outlines some additional work that could be useful for future researchers interested in expanding on our work. The overall goal was to advanced the field of alternative biofuel research by offering tolerance solutions to researchers struggling to produce high enough yields of their toxic compound. Additionally, the techniques applied in this work can be extended to search for many additional genes that respond to varying stressors.

CHAPTER 2: ENGINEERING IMPROVED BIO-JET FUEL TOLERANCE IN *ESCHERICHIA COLI* USING A TRANSGENIC LIBRARY FROM THE HYDROCARBON-DEGRADER *MARINOBACTER AQUAEOLEI*

2.1. Designing a Suitable Fosmid Library for Screening

We wanted to construct a genomic library from a microbe that thrives in hydrocarbon rich environments to screen for tolerance genes in *E. coli*. We selected *Marinobacter aquaeolei* to meet this criterion. *M. aquaeolei* was discovered living near an oil-producing well in Vietnam, meaning that the bacterium has most likely evolved mechanisms to survive in harsh, hydrocarbon rich environments (Huu, Denner et al. 1999).

We elected to use fosmids to increase the *M. aqueolei* genome coverage of our library. Fosmids contain inserts on the order of 40kb and are inserted into cells by using bacteriophages instead of traditional transformation techniques. For a genomic library to have full coverage of the genome, there must be enough unique clones in the library. Due to the large insert size, fosmids require small number of clones to ensure full coverage. Additionally, the large insert size of fosmids makes it more likely to capture multiple genes located near each other that work together to perform a desired function. Using the Clark-Carbon equation below, based on the assumption that recombinant clones follow a Poisson distribution across the genome, library coverage can be calculated with a selected level of certainty that full coverage has been achieved (Clarke and Carbon 1976).

P is the probability of full coverage in the library, *f* is the fraction of the genome contained on one plasmid or fosmid, and *N* is the number of clones.

$$P = 1 - (1 - f)^N$$

For *M. aquaeolei,* the genome is 4.33Mb long and assuming a 4kb insert size and 95% certainty of full coverage, approximately 3,200 clones would be needed to create a plasmid library. Fosmid libraries have a number of advantages over plasmid libraries. Larger insert sizes (~40kb) drastically reduce the number of clones needed to obtain full coverage by a factor of 10 (> 320 clones). Fosmids also offer good transformation efficiencies (>10^9 cfu/mL), which can make it easy to get the necessary clones for full coverage. There are however drawbacks to using fosmids in place of plasmids. With such a high number of genes being carried, there is a higher chance that detrimental heterologous genes may be introduced and effectively negate any potential benefit from positive genes found on the same fosmid. However, work has been done in the past to show that screening fosmids for genes that benefit the host cell can be successful (Ruegg, Kim et al. 2014). With this in mind, we chose fosmids over plasmids to create the *M. aquaeolei* library. Figure 3 illustrates the process of creating the fosmid library.

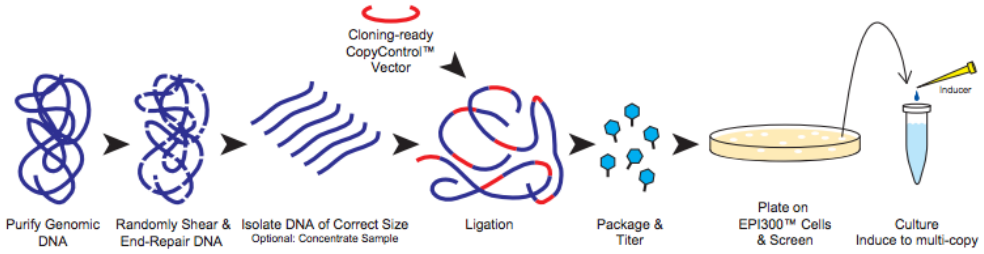


Figure 3: Process of creating the fosmid library (taken from the Epicentre Fosmid Library Production Kit CCFOS110)

To create the library, we used the Epicentre CCFOS110 Fosmid Library Production Kit and were able to generate approximately 1,000 clones. The *M. aquaeolei* genomic DNA used for the creation of the library was obtained from ATCC (#700491). We first ran our genomic DNA on a gel to determine its length. Figure 4 shows that we were starting roughly with genomic DNA fragments on the order of 42 kb. According to the kit protocol, this was an ideal length to use directly as the fosmid insert in the library building process.

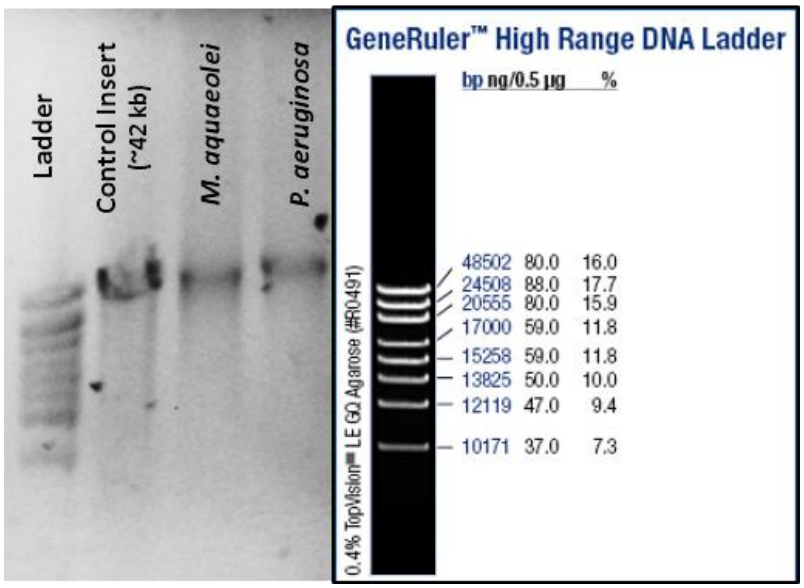


Figure 4. Comparison of genomic DNA samples of versus the human DNA control (~40 kb) provided in the kit and a high range DNA ladder. The genomic DNA for *M. aquaeolei* was centered around ~42 kb meaning that it could be used directly as the insert without gel extraction.

2.2. Testing the *M. aquaeolei* Fosmid Library in Pinene

The *M. aquaeolei* library was then stressed with the bio-jet fuel precursor pinene to determine if any of the *E. coli* harboring the fosmids displayed improved tolerance. The *M. aquaeolei* fosmid library was screened in 0.05% (v/v) pinene over the course of 96 hours, using serial dilutions into fresh media with pinene every 12 hours. We selected this pinene concentration because it severely inhibited growth of the negative control, while cultures with the library survived (Figure 5).

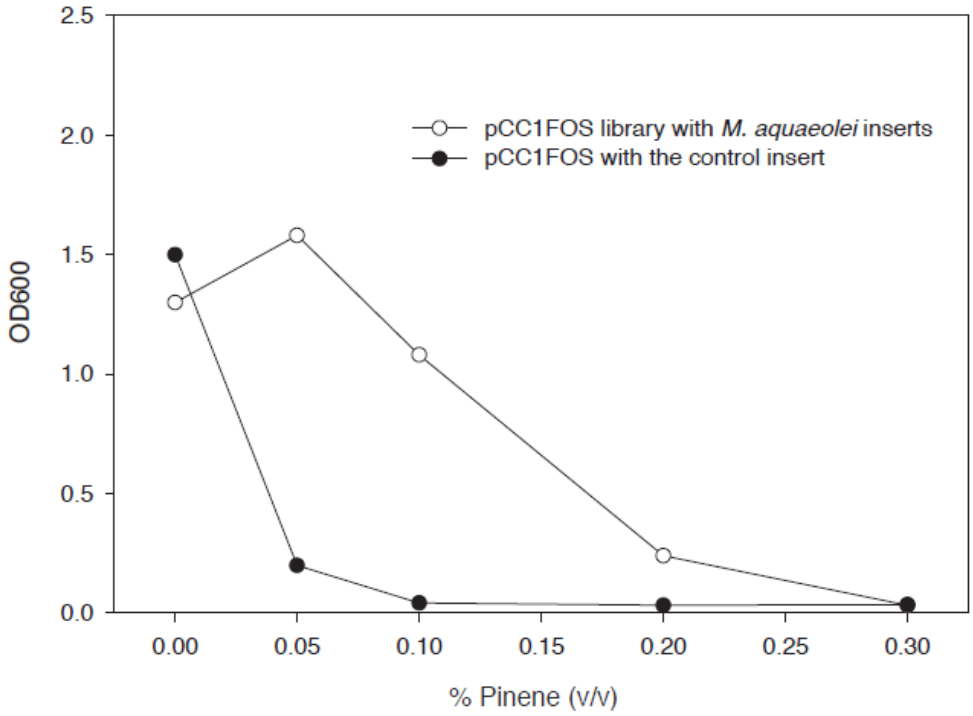


Figure 5. Initial testing of *E. coli* EPI300-TI cells containing the control fosmid and the *M. aquaeolei* fosmid library. 0.05% pinene was selected for subsequent experiments. Measurements of the optical density of the culture were taken at 600nm (OD600).

Based on preliminary tolerance tests, we chose to run the competition experiment (Figure 6) at 0.05% pinene, which inhibited growth of the negative control. The positive control was the fosmid library run with no biofuel added, which controlled for the natural competition between cells to account for any spontaneous mutations that arise. In addition to the fosmid library with no biofuel, we also ran a pBbA5k-rfp negative control with pinene added. The reason for testing the *rfp* control with pinene was to ensure there were no genomic mutations occurring that were causing the improved tolerance. Any

observed tolerance increases would have to be coming from the fosmid and not genomic mutations.

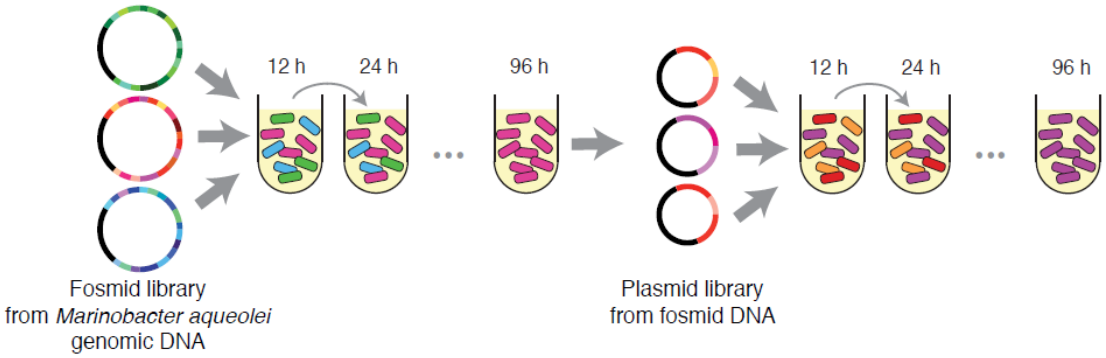


Figure 6. Genomic library approach and sequences isolated in competition experiments. (A) Illustration of the competition experiment used to isolate tolerance genes from *M. aquaeolei*. *E. coli* containing fosmids with *M. aquaeolei* inserts were grown in the presence of 0.05% (v/v) pinene. Every 12 hours, the cultures were diluted into fresh media with pinene. The fosmid that survived the initial competition experiment was then used as a template for a plasmid library and the competition procedure was repeated.

At the 0, 24, 48, and 96-hour time points we plated cells to isolate single colonies, extracted fosmids, and sequenced to check for library convergence. As expected, at 0 hours all three samples that were sequenced were unique since no stress had been applied to the library yet (Table 1).

Time Point	Sample #	Start	End	Length
0 hours (Fosmid)	1	1,930,787	1,970,656	39.9 kb
	2	2,608,132	2,643,666	35.5 kb
	3	3,177,360	3,219,568	42.2 kb
24 hours (Fosmid)	1	2,535,901	2,567,306	31.4 kb
	2	962,181	1,001,405	39.2 kb
	3	1,881,833	1,915,757	33.9 kb
48 hours (Fosmid)	1	1,885,857	1,926,221	40.4 kb
	2	1,885,854	1,926,222	40.4 kb
	3	1,885,857	1,926,222	40.4 kb
96 hours (Fosmid)	1	1,885,855	1,926,241	40.4 kb
	2	1,885,857	1,926,235	40.4 kb
	3	1,885,856	1,926,237	40.4 kb
96 hours (Plasmid)	1	1,885,857	1,888,294	2.6 kb
	2	1,885,857	1,888,294	2.6 kb
	3	1,885,857	1,888,296	2.6 kb

Table 1. Sequencing results from single colonies at various time points in the selection experiments. Fosmids and plasmids were sequenced using primers listed in Methods and the start and stop position on the *Marinobacter aquaeolei* genome were determined from these results. The numbers in the table correspond to the base pair locations in the GenBank sequence for the *Marinobacter aquaeolei* VT8 complete genome (GenBank: CP000514).

After 24 hours, all sequenced samples were unique with no overlap between the inserts. By 48 hours the three sequenced samples were identical and matched those that were subsequently extracted at the 96-hour time point. We note that one of the three fosmids sequenced from the 24-hour time point overlapped the fosmid sequence from 48 and 96 hours, but the start and end points of the sequence differed. We refer to the converged fosmid here as pCC1FOS-96. The insert contained in pCC1FOS-96 had 43 complete genes from *M. aquaeolei*. The next step was to determine which gene or genes were conferring pinene tolerance to *E. coli.*

2.3. Creating and Testing a Sub-library from pCC1FOS-96

To efficiently screen for beneficial genes, we created a sub-library using the fosmid pCC1FOS-96 as the starting template. pCC1FOS-96 DNA was partially digested using the restriction enzyme Sau3A1. Cut sites for Sau3A1 (GATC) appear frequently in the sequence, so we used a partial digest, controlling the length of the reaction and the units of enzyme added. We extracted DNA in the 4-12kb range, selecting this size to allow for complete genes or the possibility of multiple genes without creating inserts that would be prohibitive for cloning or those that return so many genes that it would be difficult to determine contributions. We cloned the inserts into the medium copy vector pBbA5k (Lee, Krupa et al. 2011), using *E. coli* MG1655 for all subsequent experiments. In order to screen for genes that conferred tolerance, we repeated the competition procedure, stressing cells with 0.05% pinene over the course of 96 hours, as we did with the fosmid library (Figure 6). We conducted the same control experiments as in the fosmid library, replacing the negative control fosmid with pBbA5k-rfp, which expresses red fluorescent protein and does not confer pinene tolerance. Upon completion of the competition experiment, all colonies sampled contained identical plasmids, which we refer to as pBbA5k-96. We extracted and retransformed this plasmid into fresh *E. coli* MG1655 cells to control for the possibility that improvements in tolerance were due to genomic mutations, however the freshly transformed cells retained improved tolerance. This indicated that the presence of the pBbA5k-96 plasmid was responsible for conferring tolerance.

2.4. YceI Improves Pinene Tolerance in *E. coli*

The plasmid pBbA5k-96 contains two complete genes from *M. aquaeolei* with additional truncated genes on each side (Figure 7). The complete genes encode for

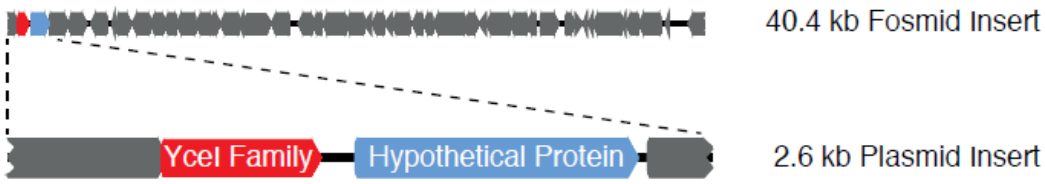


Figure 7: The isolated fosmid and plasmid inserts from the *M. aqueolei* competition experiment

a YceI-family protein, Maqu_1680 [GenBank:YP_958951.1], and a hypothetical protein, Maqu_1681 [GenBank:YP_958952.1]. To determine which gene or genes were responsible for providing increased pinene tolerance, we subcloned the two complete genes individually and together into pBbA8k medium copy vectors. The plasmid contains an arabinose-inducible P_{BAD} promoter, allowing us to control the level of gene expression. Using 100 µM arabinose we tested the tolerance of cells containing plasmids with the individual and combined genes (Figure 8). Under no pinene stress, cells with YceI grew comparably to the negative control, while all cells expressing the hypothetic protein (alone, in combination with YceI, and on the original pBbA5k-96) experienced a reduction in growth. When exposed to 0.15% pinene, only cells with the YceI protein showed improvements in growth relative to the negative control. Statistical analysis was performed in GraphPad Prism using a one-way ANOVA with a Dunnett test to compare each group of samples with the control in both cases (with and without biofuel). Our tests

show that there is a benefit to having YceI in the samples when pinene is present (p-value < 0.001). The Dunnett test also showed that there was no statistical benefit for having the hypothetical protein.

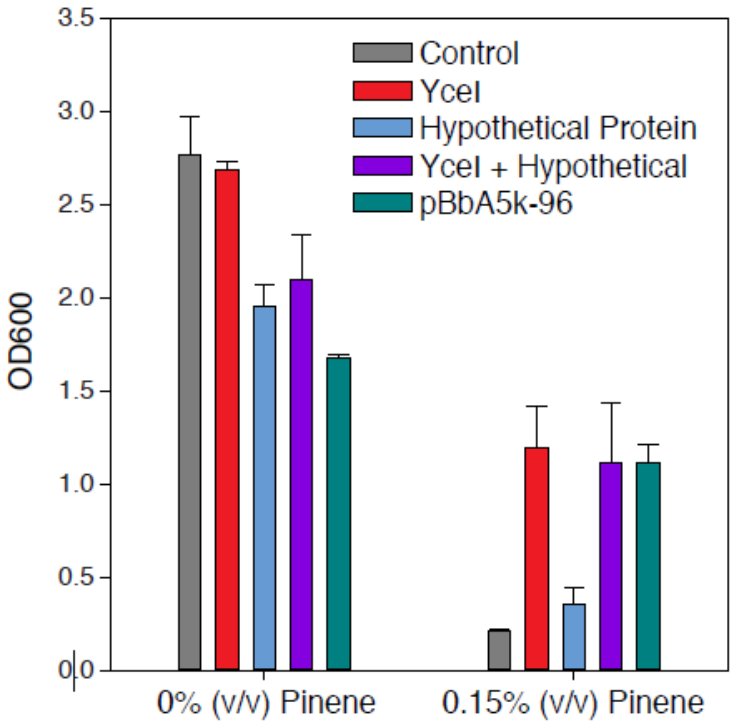


Figure 8. Testing the tolerance of cells containing plasmids with the individual genes identified in the pinene tolerance selection. Measurements of the cell densities were taken at 600nm (OD600). The control is pBbA8k-rfp; other plasmids use the pBbA8k vector and contain either *yceI*, the gene encoding the hypothetical protein, or both. pBbA5k-96 contains both genes as well as the truncated genes at the start and end of the plasmid insert. Error bars represent one standard deviation from the mean.

All cells expressing YceI—including those with YceI alone, in combination with the hypothetical protein, and on pBbA5k-96—showed improvements in tolerance and were statistically significant (p-value < 0.001, one-way ANOVA with Dunnett test). Therefore, we concluded that *yceI* was the sole gene responsible for the increased survival during the competition experiment as well as the increased tolerance observed at 0.15% pinene versus the control.

YceI-family proteins are diverse and remain largely uncharacterized. The YceI-like protein TT1927B from *Thermus thermophiles* HB8 functions as an isoprenoid transport or storage protein and may serve as a part of an unknown isoprenoid metabolic pathway (Handa, Terada et al. 2005). The crystal structure for the protein was solved with the protein complexed with its ligand, a C_{40} isoprenoid. HP1286 from *Helicobacter pylori* is another YceI-like protein and is overexpressed in acid stress (Sisinni, Cendron et al. 2010). It binds to amphiphilic compounds containing approximately 22 carbon atoms. HP1286 is secreted by the cell, potentially to sequester and supply fatty acids from the environment for metabolism or to detoxify and protect the cell from the antimicrobial properties of fatty acids. *E. coli* harbors a periplasmic YceI-like protein that responds to basic, high pH conditions, but its interaction with monoterpene biofuels has not otherwise been well characterized (Stancik, Stancik et al. 2002). The evidence provided suggests that YceI is a periplasmic beta-barrel protein that binds with the hydrophobic biofuel compounds inside of the barrel, thus preventing the biofuel from causing stress. YceI from *M. aquaeolei* shares 36% identity with TT1927b, a 31%

identity with HP1286, and a 35% identity with the YceI-like protein from *E. coli*. Previously studied proteins in the YceI family have a multitude of functions, ranging from a role in the isoprenoid metabolic pathway to involvement in acid and base stress.

We next tested different concentrations of pinene to determine the highest level cells expressing *yceI* could withstand (Figure 9). For all tested pinene levels up to 0.25%, cells expressing *yceI* grew better than the control.

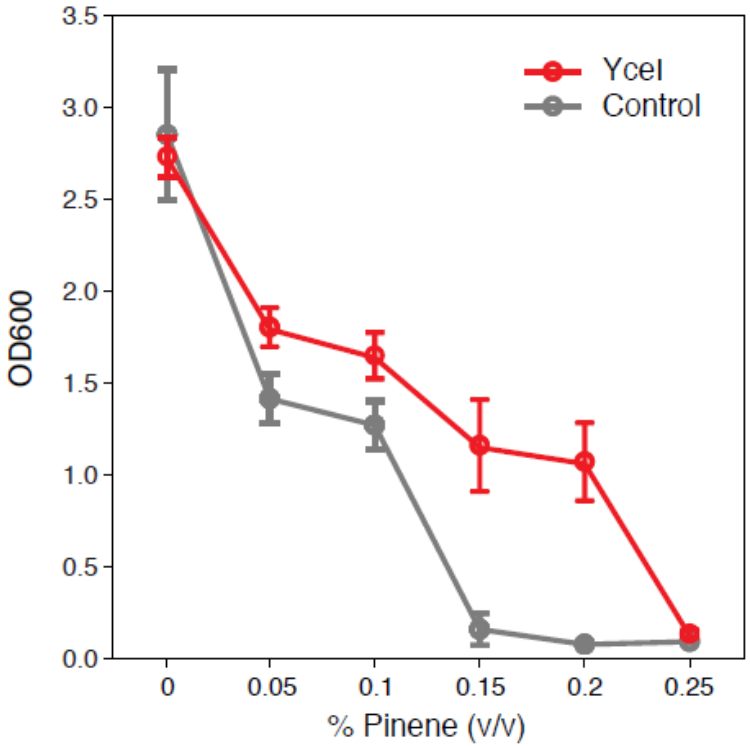


Figure 9. Final OD values under conditions with increasing pinene levels with measurements being taken after 12 hours of growth. Error bars represent one standard deviation from the mean.

We also tested whether overexpression of *yceI* was toxic to cells. Other biofuel tolerance mechanisms, such as efflux pumps are known to provide a benefit under stress, but are toxic if overexpressed, leading to a trade off in expression levels for optimal cell survival (Turner and Dunlop 2014) . We tested varying levels of inducer ranging from 0 to 1000 µM arabinose and observed no growth impact compared to the control at any concentration of inducer (Figure 10).

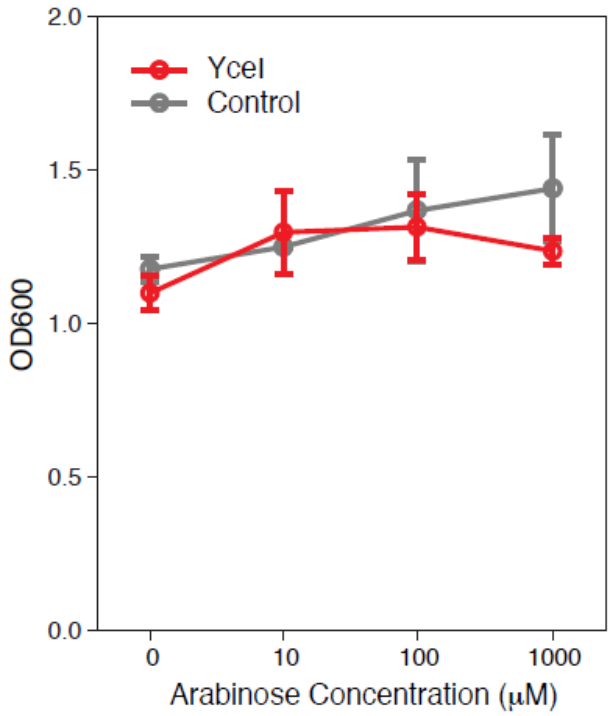


Figure 10. Measuring the toxicity of *yceI* expression compared to the control by inducing expression with arabinose. Error bars represent one standard deviation from the mean.

For all concentrations of arabinose, growth of cells expressing the *yceI* gene were shown through a one-way ANOVA test to be statistically equivalent to the corresponding

negative control at each arabinose level. This implies that there is no toxicity incurred when expressing *yceI* highly.

Due to the similar chemical structure of pinene to the bio-jet fuel precursors terpinolene, terpinene, and limonene (Brennan, Turner et al. 2012), we decided to test tolerance using these chemicals as well. Cells expressing *yceI* were stressed in varying concentrations of the three monoterpene fuels. Our results indicate that *yceI* was beneficial in improving growth when cells were exposed to terpinolene (Figure 11A).

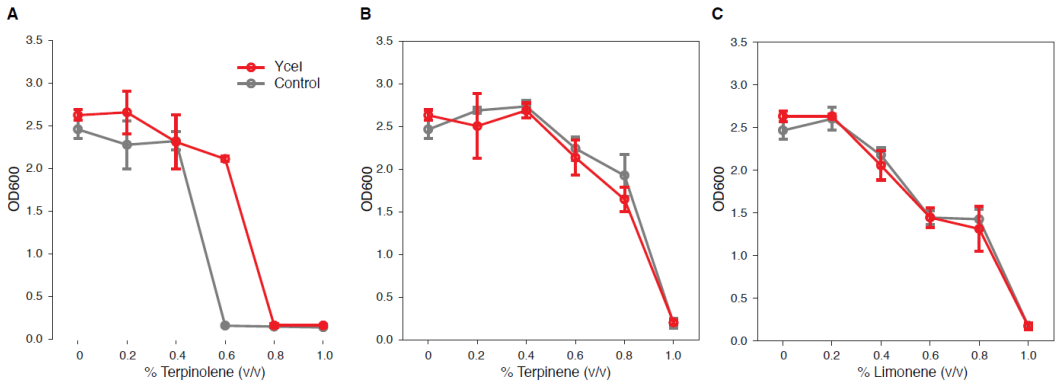


Figure 11. Testing the tolerance effects of *yceI* in the bio-jet fuel precursors (A) terpinolene, (B) terpinene, and (C) limonene. Error bars correspond to one standard deviation from the mean, measured 12 hours after the addition of biofuel.

However, for the two other fuel precursors tested, limonene and terpinene, we saw no significant improvement over the control (Figures 11B-C). It was surprising that we saw such differences in the tolerance benefit of YceI between the four fuels considering that all four of the fuels are monoterpenes ($C_{10}H_{16}$). Further analysis of the chemical structure

of the fuels shows why that may be the case. The arrangement and locations of double bonds are different for each of the fuels which could change the solubility of these chemicals in water (Figure 12). The solubility in water of isoprenoids, specifically limonene and pinene, has been investigated in the past (Fichan, Larroche et al. 1999). They found that the solubiltity of limonene in water, although low, is approximately four times greater than that of pinene.

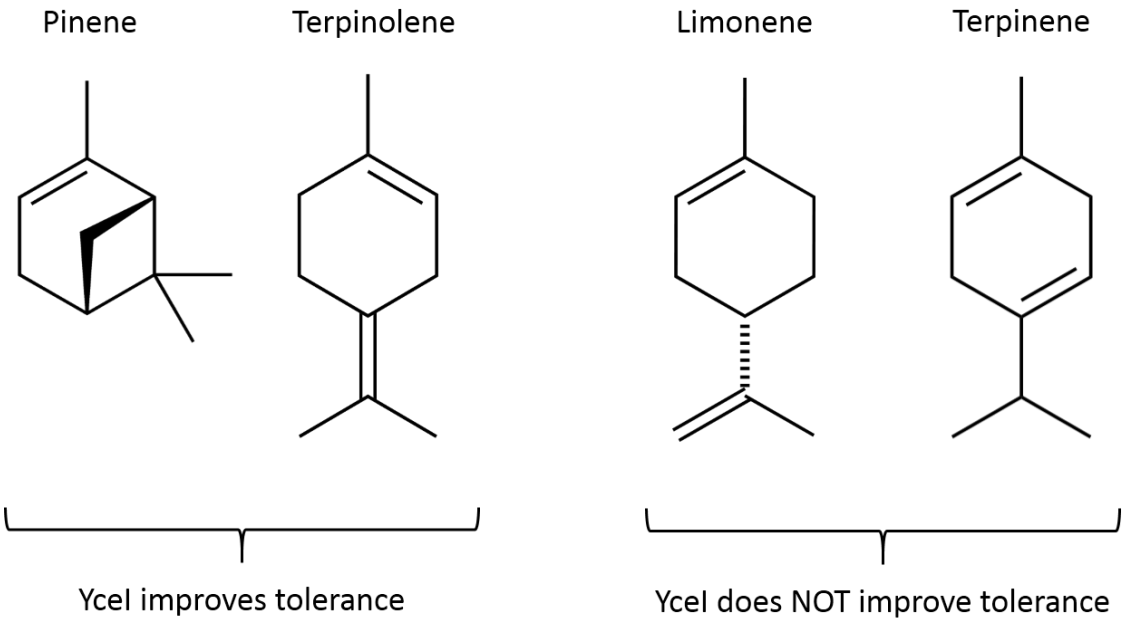


Figure 12. Chemical structures of the four monoterpenes tested in our experiments.

Based on the increased solubility of limonene in water compared to pinene, it is possible that this difference in solubility could be linked to the effectiveness of YceI in mitigating the stress imparted by the fuel.

2.5. Methods— *Marinobacter aquaeolei* Experiments

Marinobacter aquaeolei VT8 genomic DNA was obtained from ATCC (#700491D-5) and was used for the construction of the fosmid library. The degree of genomic DNA fragmentation was determined via gel electrophoresis on a 0.7% agarose gel run for 20 hours at 30 volts. The gel was then stained and imaged. The genomic DNA had an average length of ~40 kb, which we used directly as the insert in the library construction process. The fosmid library was created using the CopyControl Fosmid Library Production Kit with the pCC1FOS vector (Epicentre). A control fosmid was also created using the ~42 kb fragment of human X-chromosome DNA provided as a control in the kit. *E. coli* EPI300-T1 cells were mixed with the prepared phage particles at different concentrations to determine an appropriate titer, and we selected a 1:10 dilution of the phage particles. The *M. aquaeolei* library and control fosmid were plated on Luria Bertani (LB) agar plates containing 12.5 µg/ml of chloramphenicol, yielding approximately 1,000 colonies per plate after overnight growth at 37C. The colonies containing fosmids with *M. aquaeolei* DNA were scraped off the plate using a razor blade

and 50 µl aliquots of the library were stored in a 20% glycerol solution at -80C for later use. An individual colony from the control fosmid plate was used to create a glycerol stock.

2.5.1 Selection Procedure—Fosmid Library

LB supplemented with 12.5 µg/ml of chloramphenicol was used to prepare two cultures containing *E. coli* EPI300-T1 cells harboring the *M. aquaeolei* fosmid library. A third culture containing the human DNA control fosmid in *E. coli* EPI300-T1 was prepared in a similar manner. One of the *M. aquaeolei* library cultures and the control were stressed with 0.05% pinene (v/v), while the other *M. aquaeolei* library culture received no exposure to pinene. Each of the cultures was diluted 1:100 into LB plus chloramphenicol and pinene, where applicable, every 12 hours over the course of a 96 hour period. Every 24 hours, samples from each culture were plated on LB agar plates with chloramphenicol. Three colonies from each of the 0, 24, 48, and 96-hour time points were selected and their fosmids were extracted as follows: Cells were grown in 5ml of LB medium plus chloramphenicol and 10µl of 500X CopyControl Fosmid Autoinduction solution from the Epicentre Kit. The culture was grown for 16 hours and the fosmid DNA was then extracted using a Qiagen QIAprep Spin Miniprep Kit. The fosmid samples sequenced using the pCC1 forward (5′- GGATGTGCTGCAAGGCGATTAAGTTGG - 3′) and reverse (5′- CTCGTATGTTGTGTGGAATTGTGAGC - 3′) primers indicated in

the Epicentre Kit protocol. The resulting sequences were aligned with the *M. aquaeolei* genome. The converged fosmid from the 96-hour time point, which we refer to as pCC1FOS-96, was saved for later use.

2.5.2 *Marinobacter aquaeolei* Plasmid Library Construction

The fosmid from the converged 96-hour time point (pCC1FOS-96) was used as the starting material for plasmid library construction. 500 µg of the fosmid DNA was partially digested at 37C using varying concentrations (0.13U, 0.25U, 0.38U, and 0.50U) of the enzyme Sau3A1 (New England Biolabs). By reducing the digestion time to five minutes, we achieved a DNA insert length centered around 4-12 kb. The digested fosmid DNA from each of the four reactions was then gel extracted from the 4-12 kb range using the Qiagen QIAquick Gel Extraction Kit (Wargo, Szwergold et al. 2008). We used pBbA5k (Lee, Krupa et al. 2011) as the plasmid vector, which has a medium copy p15A origin of replication, lacUV5 promoter upstream of the cloning site, and a kanamycin resistance gene. To prepare the vector, the pBbA5k-rfp plasmid was double digested using BamHI and BglII and gel extracted. The pBbA5k vector and prepared library inserts were ligated using T4 DNA ligase (Fermentas) at a 3:1 insert to vector ratio. The ligated mixture was transformed into *E. coli* MG1655 and plated onto LB agar plates containing 50 µg/ml kanamycin. Plates were incubated overnight at 37C and the resulting

~500 colonies were scraped off of the plates using a razor blade and stored at -80C in a 20% glycerol solution, as described above.

2.5.3 Selection Procedure—Plasmid Library

The selection experiment run using the plasmid library was the same as the fosmid selection described above, with the following changes: pBbA5k-rfp in *E. coli* MG1655 was used as the control, 50 µg/ml of kanamycin was used in place of chloramphenicol, and sequencing was performed on 3 samples from the pinene-treated plasmid library plate at 96 hours using the forward (5'-GGAATTGTGAGCGGATAACAATTTC-3') and reverse (5'-CGTTTTATTTGATGCCTGGAGATCC-3') primers for the pBbA5k vector, as given in (Lee, Krupa et al. 2011). The converged to plasmid after the 96 hours was saved for later use and named pBbA5k-96.

2.5.4 Subcloning Genes from the Converged Plasmid

Two *M. aquaeolei* genes encoding for the YceI family protein [GenBank:YP_958951.1] and the hypothetical protein [GenBank:YP_958952.1] were subcloned using the pBbA8k BioBrick vector (Lee, Krupa et al. 2011), which has a p15A

origin or replication, P_{BAD} promoter, and kanamycin resistance cassette. The DNA fragments were cloned using the Gibson Assembly Protocol (Gibson, Young et al. 2009). Inserts were prepared using PCR and the vector was prepared by digesting pBbA8k-rfp with BamHI and BglII and gel extracting. Individual colonies of each construct were isolated and cloning success was verified via sequencing using the forward (5'-CTACTGTTTCTCCATACCCGTTTTTTTGG-3') and reverse (5'-CGTTTTATTTGATGCCTGGAGATCC-3') primers for the pBbA8k vector (Lee, Krupa et al. 2011).

2.5.5 Tolerance testing—*Marinobacter aquaeolei* Converged Samples

Overnight cultures were grown for 16 hours in LB containing antibiotics and varying levels of arabinose (as required). Cultures for tolerance testing were prepared by first preparing a pre-culture, where we inoculated 5 ml of LB containing antibiotics and arabinose (as required) with 50 µL of the overnight culture. Cultures were grown until they reached an OD600 reading of 0.2, at which point varying levels of biofuel were added to each. The chemicals used for tolerance testing were obtained from Sigma Aldrich (α-pinene P45680, γ-terpinene 86478, limonene 183164, and terpinolene W304603). Growth measurements were taken after 12 hours under biofuel stress. All experiments were performed in triplicate.

CHAPTER 3: HETEROLOGOUS SIGMA FACTOR EXPANDS THE SEARCHABLE SPACE FOR BIOFUEL TOLERANCE MECHANISMS

3.1. RpoD from *Pseudomonas putida* Unlocks Diversity

In this study we applied the genomic library screening approach to identify genes from *P. putida* that improve pinene tolerance in *E. coli*. *P. putida* and other *Pseudomonas* species are well-studied for their resistance to solvents (Ramos, Duque et al. 1998, Rojas, Duque et al. 2001, Rühl, Schmid et al. 2009), and a targeted study has identified genes specific to pinene tolerance (Dunlop, Dossani et al. 2011). However, we asked whether we could identify additional tolerance genes by co-expressing the sigma factor RpoD from *P. putida* and whether results were specific to the level of sigma factor in the selection experiment (Fujita, Hanaura et al. 1995). We hypothesized that differences in overall gene expression may play an important role in uncovering genes like efflux pumps that are sensitive to overexpression. To accomplish this we co-transformed *E. coli* with a *P. putida* genomic library and a plasmid containing *rpoD* from *P. putida* (*rpoD*$_{P. putida}$). We were able to isolate many pinene tolerance genes from *P. putida*, including both new and known mechanisms. Furthermore, inducing the expression of *rpoD*$_{P. putida}$ dramatically increased the diversity of tolerance mechanisms recovered. However, we observed an upper limit on this effect under high induction. We found that this method of varying heterologous sigma factor expression is effective for both plasmid and fosmid transgenic libraries.

We began by constructing a plasmid library using genomic DNA from *P. putida* KT2440. We targeted insert sizes ranging from 4-12 kb to allow us to isolate complete genes or gene clusters. We cloned the inserts into a p15A origin medium-copy vector (~10-30 copies/cell). Using *E. coli* as a host, we then co-transformed the *P. putida* library with a second plasmid. We constructed two libraries that were used in subsequent selection experiments. In the first library, we co-transformed the *P. putida* plasmids with a plasmid containing an arabinose-inducible version of red fluorescent protein (*rfp*), which serves as a control. In the second library, we co-transformed the *P. putida* plasmids with an arabinose-inducible copy of $rpoD_{P.\ putida}$.

We first exposed cells from the *P. putida* library with *rfp* and no $rpoD_{P.\ putida}$ to 0.05% (v/v) pinene. Cells were exposed to pinene over the course of 60 h, with serial dilutions into fresh medium with pinene every 12 h. At the final time step, the cultures were plated and *P. putida* library plasmids from individual colonies were sequenced. All samples we sequenced returned overlapping inserts containing a similar region of the *P. putida* genome (Figure 13, Appendix 1).

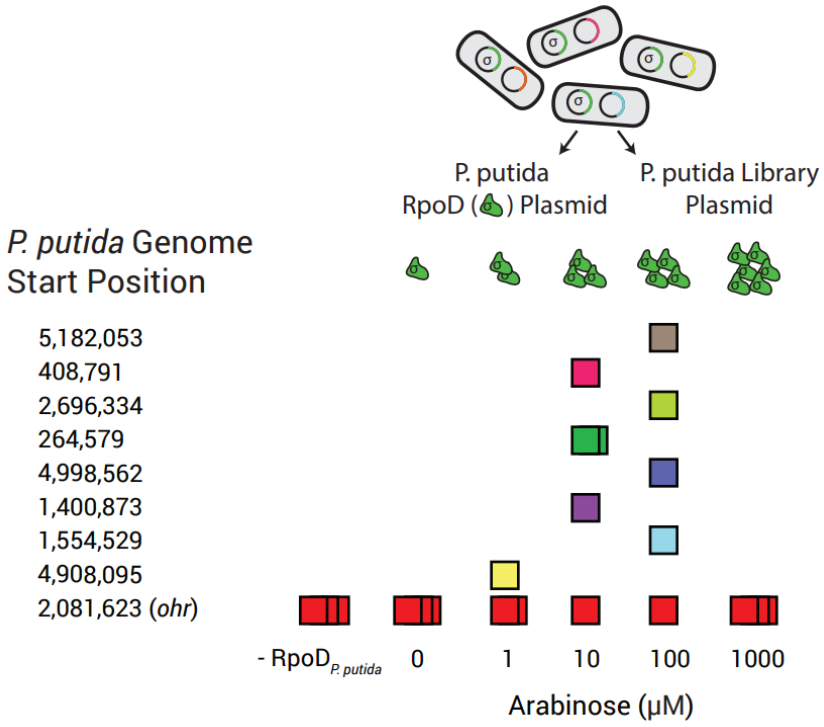


Figure 13. The sigma factor RpoD$_{P.\ putida}$ enables expression of heterologous genes from a *P. putida* library in *E. coli*. Results of library selection under 0.05% pinene with and without *rpoD*$_{P.\ putida}$. Each square represents a single sample isolated and sequenced after pinene treatment. Colors correspond to different genomic regions isolated. Further details are provided in Appendix 1. The presence and absence of *rpoD*$_{P.\ putida}$ and the level of arabinose induction are listed in the figure. In the schematic the sigma factor RpoD$_{P.\ putida}$ is denoted σ.

From the overlapping sequences we identified a single gene, *ohr* (organic hydroperoxide resistance), that has a known role in *Pseudomonas* for protecting the cell against reactive oxygen species (Lesniak, Barton et al. 2002). In *Pseudomonas aeruginosa*, Ohr functions by converting hydroperoxides to less toxic metabolites in order to reduce oxidative stress. Monoterpene stress can elicit intracellular accumulation

of reactive oxygen species (Liu, Zhu et al. 2013), therefore Ohr may help to mitigate the oxidative stress caused by the presence of pinene. To confirm its role in providing tolerance, we subcloned *ohr* into a plasmid, placing it under the control of an IPTG-inducible promoter. We verified that the presence of *ohr* was sufficient to provide tolerance to 0.05% pinene compared to the negative control (Figure 14). Inducing with 1000 μM IPTG further improved growth in the presence of pinene.

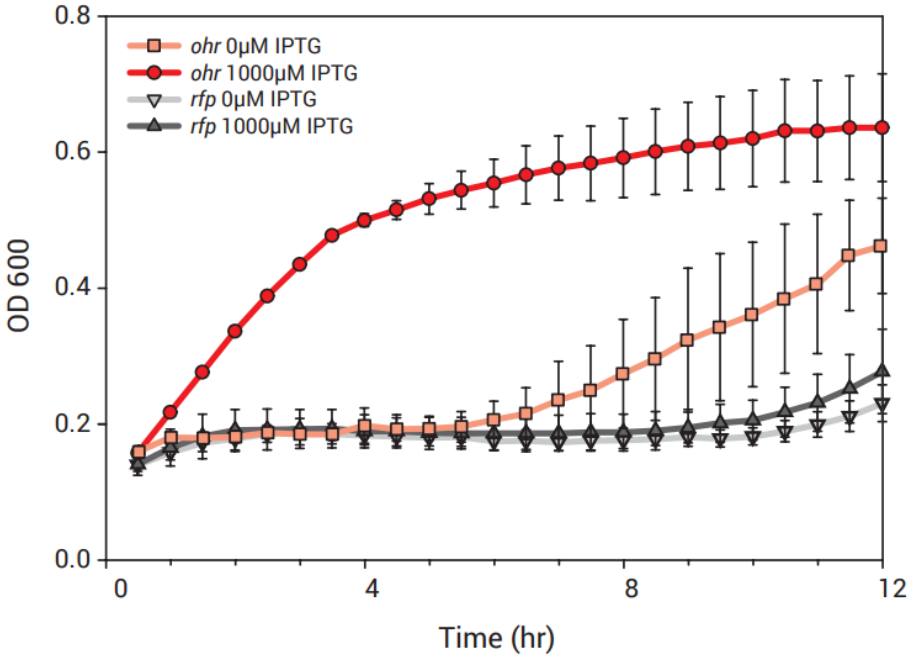


Figure 14. Ohr provides pinene tolerance without $RpoD_{P.\ putida}$. Strains with *ohr* (pBbA5a-ohr) and a negative control with *rfp* (pBbA5a-rfp) were grown in the presence of 0.05% (v/v) pinene. The mean of n = 3 biological replicates is shown; error bars are standard deviations.

Although we were able to demonstrate that *ohr* from *P. putida* helps improve pinene tolerance in *E. coli*, there were notable absences in the results of our selection

experiment. In particular, *P. putida* has several known tolerance mechanisms (Ramos, Duque et al. 1997), including at least three efflux pumps that provide resistance to organic solvents (Dunlop, Dossani et al. 2011). We hypothesized that we might be missing candidate tolerance genes due to poor expression because the native sigma factors in *E. coli* were unable to express the heterologous genes at a beneficial level.

To test this, we repeated the selection experiment using the second *P. putida* plasmid library, which contains the $rpoD_{P.\ putida}$ plasmid. We varied expression of $rpoD_{P.\ putida}$ by inducing with 0, 1, 10, 100, or 1000 µM arabinose. Interestingly, we saw an induction-dependent change in the diversity of the selection results (Figure 13). At 0 µM arabinose our selection returned only plasmids that contained *ohr*. As we increased $rpoD_{P.\ putida}$ induction, we began to see more diversity. The new results included known mechanisms such as pumps and membrane modifications, and also previously unidentified tolerance mechanisms (Appendix 1). However, we observed an upper limit to the diversity provided by $rpoD_{P.\ putida}$ induction. At 1000 µM arabinose there was a sharp decrease in diversity, and only *ohr*-containing plasmids were isolated. Some tolerance mechanisms can be toxic when overexpressed, therefore we speculate that the tolerance mechanisms we isolated may be sensitive to the level of gene expression due to $rpoD_{P.\ putida}$ induction. We also performed additional control experiments with the *P. putida* library with *rfp* and no $rpoD_{P.\ putida}$ under different arabinose induction conditions and found that the majority of the samples sequenced contained *ohr* (Appendix 1), indicating that arabinose induction alone is not responsible for the increase in diversity.

We next tested the growth of two samples isolated from the 10 µM arabinose condition. We tested *rpoD*$_{P.\ putida}$ induction in these two samples using 0, 10, 100, and 1000 µM arabinose. One of the samples (10µM_rpoD_B) was susceptible to induction effects, growing well under the 10 µM arabinose conditions it was isolated at, but less well under different arabinose induction levels (Figure 15). Based on where the sequence for this sample lies on the *P. putida* genome, it is likely that the sample contains some, or all, of the ABC transporter Pp_0219-0224 (Appendix 1). Membrane proteins such as ABC transporters are often toxic when overexpressed, which may explain the arabinose-dependent effects (Turner and Dunlop 2014).

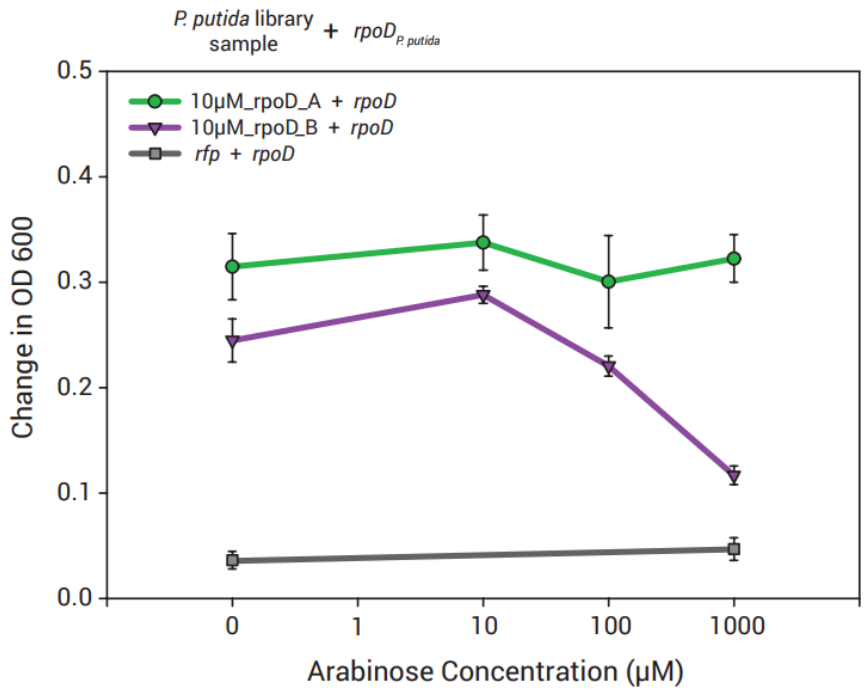


Figure 15. The change in growth after the addition of 0.05% pinene. Samples 10µM_rpoD_A and 10µM_rpoD_B were isolated from the selection experiment with *rpoD*$_{P.\ putida}$ and 10 µM arabinose. The

negative control contains a *rfp* plasmid (pBbA5a-rfp) in place of the *P. putida* library; both are co-transformed with the $rpoD_{P.\ putida}$ plasmid. The key above the figure indicates which library member (or control) is included and whether $rpoD_{P.\ putida}$ is included. Error bars show standard deviations for n = 3 samples.

In contrast, a second sample we tested (10µM_rpoD_A) performed well across all levels of arabinose induction and displayed no toxicity effects, even when $rpoD_{P.\ putida}$ was highly induced. The insert region for this sample contains some of the genes that make up the Tol-OprL system (Pp_1218-1224) responsible for maintenance of outer membrane integrity and cell morphology (Llamas, Rodríguez-Herva et al. 2003), which may play a role in tolerance. We expected the two samples to perform well in the 10 µM arabinose conditions where they were isolated and both did, but we were surprised to see good performance under other induction levels in the 10µM_rpoD_A sample. This effect may be explained by the fact that in selection experiments where many library variants compete, modest differences between strains are be amplified. Therefore, small performance deficits may not be visible in single-strain assays. As expected, a negative control containing the $rpoD_{P.\ putida}$ plasmid, but without a *P. putida* insert, performed poorly in the presence of pinene across all induction conditions tested (Figure 15).

Next, we wanted to confirm that $rpoD_{\ P.\ putida}$ was in fact vital for providing tolerance in the isolated plasmid samples. We chose two samples to run detailed tolerance tests on at 0.05% pinene. The first sample contained *ohr* on the insert. Therefore, we expected it to perform equally well with and without $rpoD_{P.\ putida}$ (Figure 16).

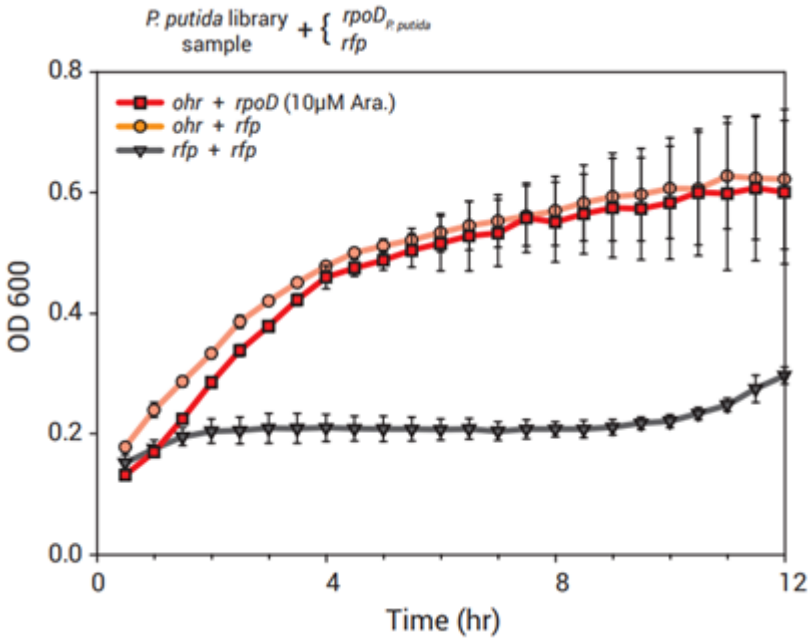


Figure 16. Characterizing the role of *rpoD*$_{P.\ putida}$ in pinene tolerance of library isolates. Samples were grown for 12 hours in 0.05% pinene after an initial preculture of two hours without pinene present. Ara is an abbreviation for arabinose. Error bars show standard deviations from n = 3 samples. Illustrated above is an *ohr* sample with and without *rpoD*$_{P.\ putida}$ and a negative control.

We found that pinene tolerance was not dependent on *rpoD*$_{P.\ putida}$. In contrast, when we removed the *rpoD*$_{P.\ putida}$ plasmid from one of our previously characterized samples (10µM_rpoD_A), we dramatically reduced the pinene tolerance of the strain (Figure 17).

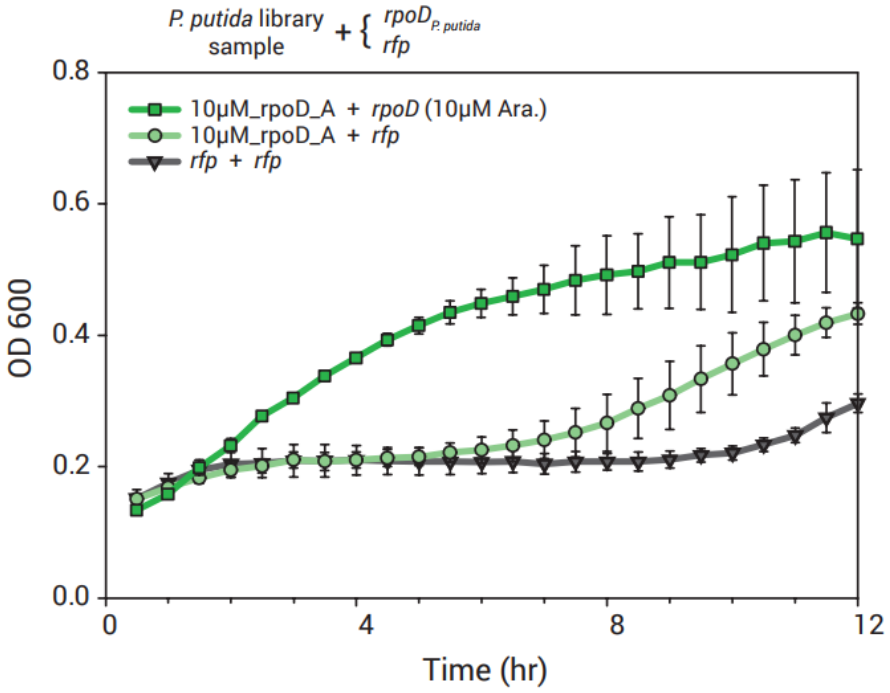


Figure 17: Characterizing the role of *rpoD*$_{P.\ putida}$ in pinene tolerance of library isolates. Samples were grown for 12 hours in 0.05% pinene after an initial preculture of two hours without pinene present. Ara is an abbreviation for arabinose. Error bars show standard deviations from n = 3 samples. Illustrated above is 10μM_rpoD_A which was isolated from the 10 μM arabinose condition during the selection experiment. Growth in pinene is shown with and without *rpoD*$_{P.\ putida}$ in comparison to a *rfp* control.

We also confirmed that expression of *rpoD*$_{P.\ putida}$ was not conferring a tolerance benefit independent of the library plasmids. To do this we tested *E. coli* with and without *rpoD*$_{P.\ putida}$ with 0 μM and 1000 μM arabinose and confirmed that there was no growth benefit to having *rpoD*$_{P.\ putida}$ induction in the absence of the *P. putida* library plasmids (Figure 18).

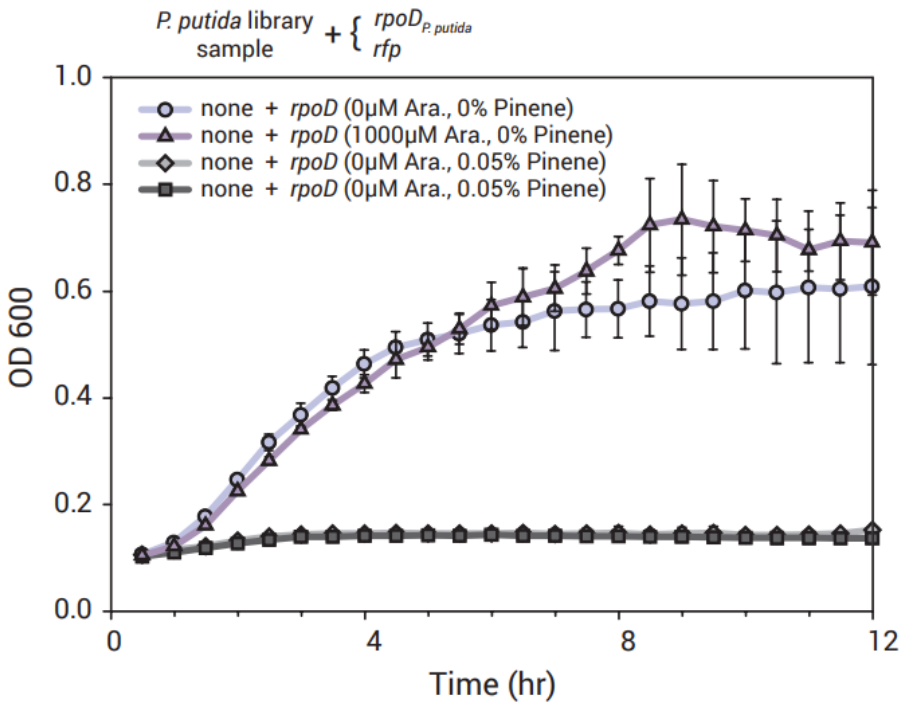


Figure 18: Characterizing the role of *rpoD*$_{P.\ putida}$ in pinene tolerance of library isolates. Samples were grown for 12 hours in 0.05% pinene after an initial preculture of two hours without pinene present. Ara is an abbreviation for arabinose. Error bars show standard deviations from n = 3 samples. The figure above suggests that *rpoD*$_{P.\ putida}$ alone does not improve tolerance. For the induced cases, 1000 μM arabinose was added during the two hour preculture and cells were subsequently grown for 12 hours.

Next we tested the tolerance of samples under increasing pinene concentrations to determine how far they could be pushed before the tolerance mechanisms no longer worked effectively (Figure 19). We found that when the tolerance mechanisms are present and the sigma factor is expressed appropriately, if necessary, we were able to significantly increase pinene tolerance. The negative control, expressing *rfp* in place of a tolerance gene, cannot survive 0.05% pinene, while the other tested samples all grew well at 0.1% pinene and beyond.

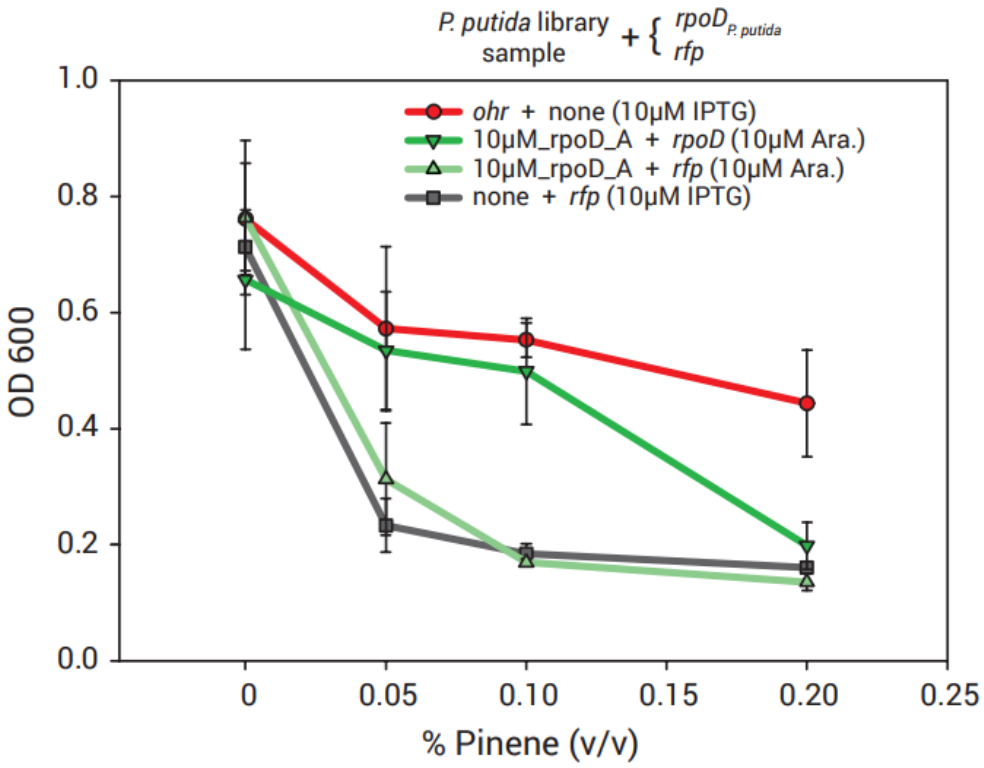


Figure 19: Characterizing the role of *rpoD*$_{P.\ putida}$ in pinene tolerance of library isolates. Samples were grown for 12 hours in 0.05% pinene after an initial preculture of two hours without pinene present. Ara is an abbreviation for arabinose. Error bars show standard deviations from n = 3 samples. The figure above is showing pinene tolerance tests for representative samples. Precultures were grown for two hours before pinene was introduced. The negative control pBbA5a-rfp, and pBbA5a-ohr were induced with 10 µM of IPTG. The other two samples each contained two plasmids: the library plasmid 10µM_rpoD_A, as well as either the *rpoD*$_{P.\ putida}$ plasmid pBbE8k-rpoD or the *rfp* control plasmid pBbE8c-rfp, both of which were induced with 10 µM arabinose.

3.2. RpoD Also Improves Fosmid Library Diversity

We next asked whether our method was also applicable to fosmid libraries. Fosmids can have much larger inserts (~35-40 kb) that are capable of carrying many adjacent genes. Also from a technical standpoint, it can be much easier to generate fosmid libraries with full coverage of the desired genome due to the larger insert size. We constructed a fosmid library using the same *P. putida* genomic DNA as we used in the plasmid library. We reran the selection experiment, comparing the *P. putida* fosmid library with and without arabinose-inducible $rpoD_{P.\ putida}$. We induced $rpoD_{P.\ putida}$ with 1000 µM arabinose in these experiments. As with the plasmid library, without $rpoD_{P.\ putida}$ we selected for many similar fosmids, all of which contained *ohr* (Appendix 2). With $rpoD_{P.\ putida}$, we observed a notable increase in the diversity of winners (Figure 20). In contrast to the plasmid library results, we did not observe any toxic effects resulting from the overabundance of $rpoD_{P.\ putida}$, even at 1000 µM arabinose. This is likely due to the difference in copy number between the fosmid (1 copy/cell) and the plasmid library (10-30 copies/cell).

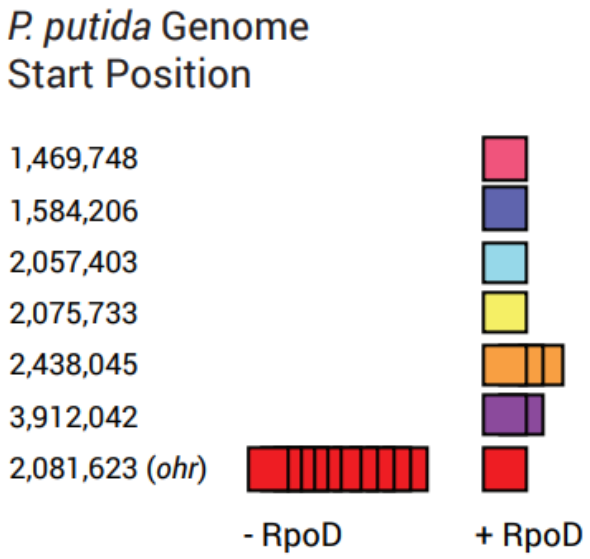


Figure 20: Fosmid library with *rpoD*$_{P.\ putida}$ increases diversity of pinene tolerance genes recovered. Results of fosmid library selection under 0.05% pinene with and without *rpoD*$_{P.\ putida}$. *rpoD*$_{P.\ putida}$ was induced with 1 mM arabinose. Each square represents a single sample isolated and sequenced after pinene treatment. Colors correspond to different genomic regions isolated. Further details are provided in Appendix 2.

In the fosmid selection experiment, subsequent tolerance testing demonstrated that surviving strains displayed significantly higher tolerance to pinene than the *rfp* negative control (Figure 21). Among the fosmid inserts we noted examples of efflux pumps that have known roles in solvent resistance, including Pp_1271 (Verhoef, Ballerstedt et al. 2010), the TtgABC pump (Godoy, Molina-Henares et al. 2010), and Pp_3456 (Turner and Dunlop 2014). We also isolated regions with no obvious candidates for known tolerance genes, which offer an intriguing starting point for future study (Appendix 2).

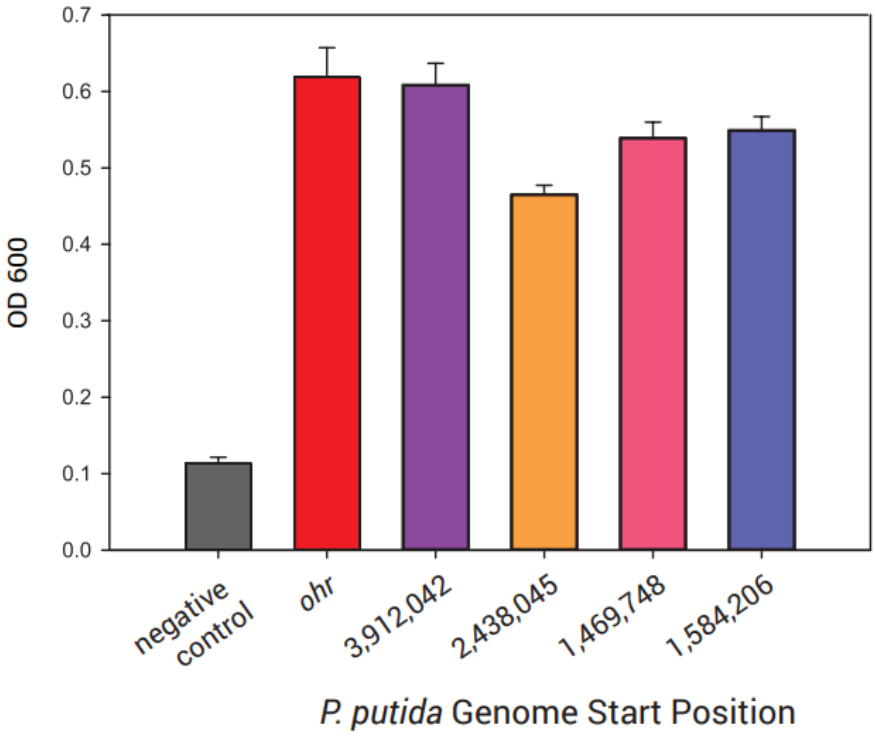


Figure 21: Tolerance of five representative samples obtained after the selection experiment and a negative control. Samples were grown for two hours before 0.05% pinene was added and *rpoD*$_{P. putida}$ was induced with 1 mM of arabinose. Error bars show standard deviations from n = 3 samples.

3.3. Analysis of RpoD Improvements

In this study, we used genomic DNA from *P. putida* and constructed plasmid and fosmid libraries to select for *P. putida* genes that improve tolerance to pinene when expressed in *E. coli*. Importantly, we found that introduction of the sigma factor *rpoD*$_{P. putida}$ greatly expanded the functional genomic space of our libraries. At moderate levels

of expression, the results of the selection experiments contained plasmids from diverse regions of the *P. putida* genome. However, high expression of *rpoD*$_{P.\ putida}$ resulted in decreased diversity. The consensus sequences for RpoD$_{P.\ putida}$ and RpoD$_{E.\ coli}$ are the same (-35: TTGACA, -10: TATAAT (Patten and Glick 2002, Shimada, Yamazaki et al. 2014)), suggesting that other details of how the sigma factors help to initiate transcription play a primary role in the differences we observed here.

In our selection experiments we identified Ohr as a protein capable of reducing pinene toxicity without necessitating *rpoD*$_{P.\ putida}$. Based on studies in *P. aeruginosa*, Ohr may work by reducing pinene to a less toxic metabolite (Lesniak, Barton et al. 2002). It is unclear whether this mechanism would render pinene ineffective as a bio-jet fuel precursor, therefore other tolerance mechanisms may prove to be better in a synthetic biology and metabolic engineering context. Identifying additional genes, such as efflux pumps that mitigate toxicity but do not change the chemical composition of the biofuel, will be especially valuable in tolerance engineering.

Studies regarding the maximum tolerable concentration of pinene *E. coli* can withstand vary widely in previously published reports due to differences in methodology and experimental conditions (Mukhopadhyay 2015). A previous study from our group showed 0.05% pinene is toxic, but that improvements up to 0.2% tolerance could be achieved with the addition of heterologous tolerance genes from *Marinobacter aquaeolei* (Tomko and Dunlop 2015). These results are comparable to what we observed in the

present study and suggest that a future research may explore synergistic effects provided by multiple tolerance mechanisms. We note that the gene candidates returned and tolerance levels achieved are likely to depend on the experimental methods used in the selection.

3.4 Methods—*Pseudomonas putida* Plasmid Library Construction

We constructed the plasmid library using *P. putida* KT2440 (ATCC 47054) genomic DNA, which was isolated using the Omega Mollusc DNA kit (Omega BioTek) after pretreatment with lysozyme to improve extraction efficiency. 10 µg of genomic DNA was digested with the restriction enzyme Sau3A1 (Fisher BioReagents) which cuts the sequence GATC. We varied the concentration of restriction enzyme (7.5 – 12.5U in steps of 2.5U) to control the degree of fragmentation. The reactions were run for 10 minutes at 37C, followed by heat inactivation at 65C for 20 minutes. The resulting cut DNA was separated using gel electrophoresis and sequences in the 4-12 kb region were gel extracted (Qiagen QIAquick Gel Extraction Kit) and combined together to serve as the insert in the plasmid library. The cloning vector was prepared from pBbA5a-rfp (Lee, Krupa et al. 2011). This vector has a medium copy p15A origin of replication, lacUV5 promoter upstream of the cloning site, and an ampicillin resistance gene. We used the BamHI and BglII restriction sites to remove rfp and isolated our vector via digestion and subsequent gel extraction. To prevent recombination, the vector was treated at 37C three

times with 1U of shrimp alkaline phosphatase (New England Biolabs) added 30 minutes apart from each other, followed by a heat inactivation step for 5 minutes at 65C. In order to remove any residual recircularized vector, the resulting mixture was then ligated together (without the addition of an insert) using T4 DNA ligase (Fermentas) and loaded on a gel. The linearized vector was gel extracted and used for library production.

We used a vector-to-insert ratio of 1:3 to construct the plasmid library. The resulting mixture was transformed into MegaX DH10B T1 Electrocomp Cells (ThermoFisher Scientific). Three batches of transformants were plated across a total of 12 Luria Bertani (LB) agar plates containing 100 μg/ml ampicillin. This yielded ~30,000 colonies, which were then scraped off with a razor blade and suspended in approximately 10 ml of LB medium. Using the Clark-Carbon equation, we determined that we have >99.9% probability of full coverage of the *P. putida* genome (conservatively assuming an insert size of 4kb, and 50% empty vector background) (Clarke and Carbon 1976). We then isolated the plasmid library using a Qiagen QIAprep Spin Miniprep Kit and retransformed it into E. coli EPI300-T1 with pBbE8k-rpoD or pBbE8k-rfp.

3.4.1 Construction of the *rpoD*$_{P.\ putida}$ Plasmid

We amplified the *rpoD*$_{P.\ putida}$ gene from *P. putida* KT2440 (ATCC 47054) genomic DNA using PCR (F: ATGTCCGGAAAAGCGC, R: TCACTCGTCGAGGAAGGA) and cloned it into the pBbE8k vector (Lee, Krupa et al. 2011) using Gibson assembly (Gibson, Young et al. 2009). pBbE8k has a high-copy colE1 origin of replication, arabinose-inducible pBAD promoter, and contains a kanamycin resistance gene. We refer to this plasmid below as pBbE8k-rpoD.

We removed the *rpoD*$_{P.\ putida}$ plasmid for experiments in Figs. 16, 17, and 19 to test the tolerance of specific library samples without the heterologous sigma factor. To do this we extracted the plasmid DNA from specific library samples, which also contained the *rpoD*$_{P.\ putida}$ plasmid. We retransformed this DNA into EPI300-T1 electrocompetent cells along with another plasmid (pBbE8c-rfp) with the same origin of replication as the *rpoD*$_{P.\ putida}$ plasmid (pBbE8k-rpoD). The new plasmid had a chloramphenicol resistance marker and replaces *rpoD*$_{P.\ putida}$ with rfp, but is otherwise identical. After transforming samples with the new plasmid, we plated the samples on carbenicillin and chloramphenicol plates to screen for colonies that contained the *P. putida* library plasmid and pBbE8c-rfp. We further confirmed that the cultures had lost the *rpoD*$_{P.\ putida}$ plasmid by plating cells on kanamycin and observed no growth.

3.4.2 *Pseudomonas putida* Fosmid Library Construction

We constructed the fosmid library using *P. putida* genomic DNA, isolated as described previously. After preparation, the DNA had an average length centered around ~40kb which we deemed to be suitable for direct use in the fosmid construction process based on guidance from the fosmid library kit (CopyControl Fosmid Library Production Kit (Epicentre) with the pCC1FOS vector). The vector was ligated with our *P. putida* insert DNA and was packaged in the provided phage particles and prepared for infection following manufacturer's instructions. A control fosmid (pCC1FOS-Con) was also created using the ~42 kb fragment of human X-chromosome DNA provided in the kit.

We infected two strains of E. coli EPI300-T1 with the *P. putida* fosmid library: one with and one without the pBbE8k-rpoD plasmid. The two libraries were created using a 1:10 dilution of the prepared phage particles. The host strain, E. coli EPI300-T1 was infected and plated on six LB agar plates containing 12.5 µg/ml chloramphenicol, supplemented with 50 µg/ml of kanamycin for the samples that contained the $rpoD_{P.\ putida}$ plasmid. The plates were grown overnight at 37C. The plates were then scraped with a razor blade as described above. Both of the fosmid libraries, with and without the $rpoD_{P.\ putida}$ plasmid, contained ~1,500 clones. Using the Clark-Carbon equation, this number of clones has a >99.9% probability that the library provides full coverage of the *P. putida* genome (Clarke and Carbon 1976).

3.4.3 Selection Procedure—*Pseudomonas Putida* Libraries

The protocols for the individual plasmid and fosmid selection experiments were identical. Library cultures were started in LB medium at a 1:100 ratio. The LB medium also contained appropriate antibiotics, 12.5 µg/ml chloramphenicol for the fosmid library and 100 µg/ml carbenicillin for the plasmid library. Samples containing the pBbE8k-rpoD or pBbE8k-rfp plasmids were supplemented with 30 µg/ml kanamycin. Arabinose was added when the cultures were inoculated and concentrations ranged from 0-1000 µM depending on the experiment to induce expression of $rpoD_{P.\ putida}$. Separate samples containing the control plasmid (pBbA5a-rfp) and control fosmid (pCC1FOS-Con) were also grown in parallel to account for the possibility of a genomic mutation leading to increased tolerance characteristics. In all cases the controls under pinene stress displayed no growth or very poor growth delayed by many hours, indicating that genomic mutations were not responsible for the tolerance improvements. Each condition in the selection experiment was performed in triplicate. All samples were grown at 37C for 2 hours, at which point 0.05% pinene (v/v) (Sigma-Aldrich) was added to each culture. The samples were allowed to grow for an additional 10 hours and were then diluted into fresh medium with antibiotics and arabinose as indicated above. From this point forward, each culture was grown for 4 hours before the addition of 0.05% pinene, then for 8 hours under pinene stress. The 4 hour pre-incubation period was necessary to allow the cells time to recover after the pinene stress. After 60 hours, the samples were plated and sequenced.

3.4.4 Tolerance testing—*Pseudomonas putida* Converged Samples

To further eliminate the possibility that pinene tolerance was due to a genomic mutation, the plasmids from individual samples from the final time step of the selection experiment were isolated and retransformed into fresh E. coli EPI300-T1 cells. Cultures were grown overnight in selective LB medium, as described above. We diluted the overnight cultures into a 24-well plate with an optically-clear adhesive film covering to prevent cross contamination and evaporation. Varying levels of arabinose were added and the samples were grown for two hours before adding the desired pinene concentration (usually 0.05%, unless otherwise indicated). OD600 measurements were taken every 30 minutes for 12 hours of additional growth under pinene stress using a BioTek Synergy H1 Hybrid Plate Reader.

CHAPTER 4: ADDITIONAL EXPERIMENTS AND FUTURE IDEAS

4.1. Additional Studies Using a *Pseudomonas aeruginosa* Library

In addition to the experiments performed using the *M. aquaeolei* and *P. putida* libraries, we also screened a *Pseudomonas aeruginosa* library for tolerance genes. We studied *Pseudomonas aeruginosa* due to its prominence as a pathogen and its use of efflux pumps to survive in the presence of a variety of solvents (Ramos, Duque et al. 2002). *P. aeruginosa* strains expressing wild-type levels of their multidrug efflux system, MexAB-OprM, showed increased tolerance to *n*-hexane and *p*-xylene versus the strain with those genes deleted (Li, Zhang et al. 1998). We used a *P. aeruginosa* plasmid library provided by Dr. Wargo at The University of Vermont for our additional tolerance screening experiments (Wargo and Hogan 2009). Fragments of the genomic DNA of *P. aeruginosa* on plasmids is contained in the library. We transformed this library into *E. coli* MG1655 for our initial biofuel tolerance tests.

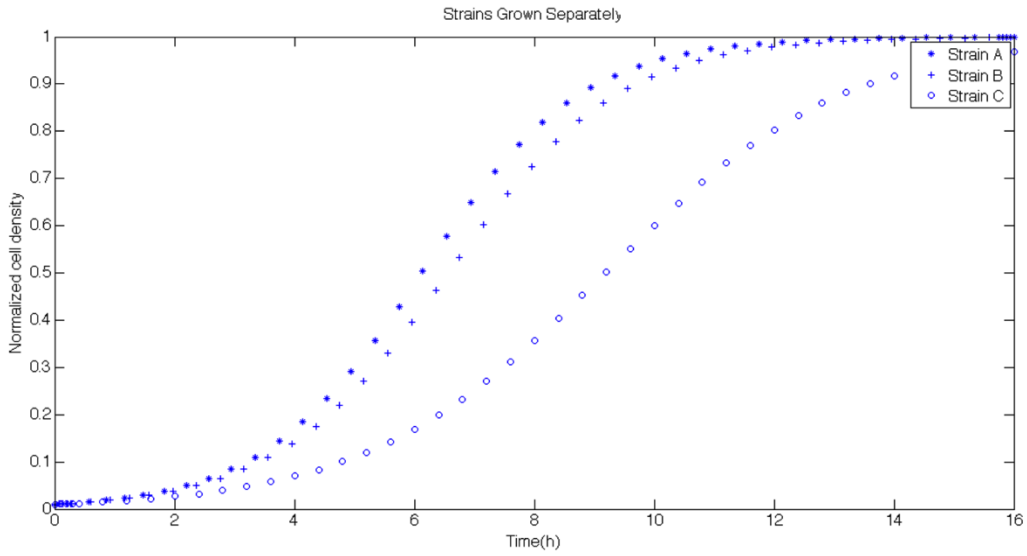
4.2. Mathematical Modeling Can Guide the Competition Experiment Design

Mathematical modeling using the competitive Lotka-Volterra equation was used to verify the concept of the competition experiment and the modeling work was derived from previous studies performed by Dunlop et al (Strogatz 2005, Dunlop, Dossani et al. 2011).

$$\frac{dN_i}{dt} = d_i N_i \left(1 - \sum_{j=1}^{C} \frac{d_j N_j}{d_i K}\right)$$

In the equation shown above, for the i^{th} strain, N_i is the cell density, d_i is the growth rate and K is the carrying capacity of the culture. Figure 22A shows a scenario where three different strains of bacteria with varying growth rates are grown independently of each other. In this scenario, there is no competition between the strains and each strain reaches its maximum cell density at a different time. In Figure 22B the three strains compete against each other in the same culture. The strains with the faster growth rates outperform the slower strains (A>B>C).

A



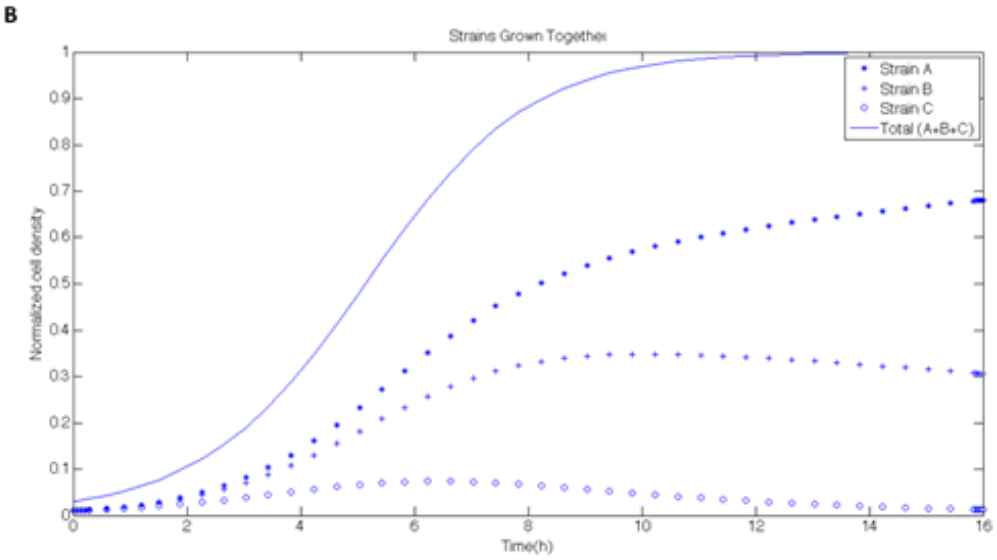


Figure 22. Model using the competitive Lotka-Volterra equation (A) Shows the scenario when there is no competition between the strains, with each strain growing independently. (B) Growing the strains in the same culture provides competition and variation in the final cell densities.

From this idea, applying a stress (such as adding biofuel) to cells expressing our library, should result in faster growing cells outcompeting the others. The model was also used to determine roughly how long to run the competition for before convergence to a single genotype would occur. Shown in Figure 23 is a model where C = 100 strains of bacteria and d_i was randomly distributed with a mean and standard deviation of μ = 0.75 and σ = 0.02. The carrying capacity was also held constant at K=1. Over the course of 96 hours, the spread between the individual cell densities becomes large. This means that the population of cells is converging to the fastest growing variants.

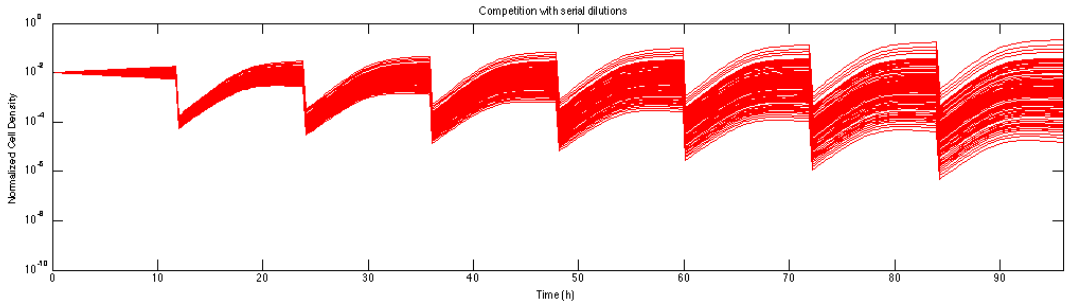


Figure 23. Model using the competitive Lotka-Volterra equation to compete 100 strains of bacteria with varying growth rates over the course of 96 hours. 1:100 serial dilutions occur every 12 hours.

The results of the modeling suggest that running the competition with our library for 96 hours, with 1:100 serial dilutions happening every 12 hours would be good for isolating primarily the best growing samples.

4.3. Comparing the Competition and the Modeling Results

From the modeling results presented in section 4.2 it was expected that 96 hours would be required for the *M. aquaeolei* fosmid library culture to start being dominated by cells that contain a single shared fosmid. We saw in Table 1 from section 2.2 that after 24 hours the three fosmids selected from the competition experiment contained different fosmids. From the model, we can determine the probability of selecting the best performing fosmid when sampling from the culture after 24 hours P(pCC1FOS-96) = 0.029. In this case, the model matches well with the experimental data because the

probability of selecting the fosmid pCC1FOS-96 at the 24 hour time step was very low and none of the samples sequenced at that time step were pCC1FOS-96.

At the 48 hour time step, all three of the sequenced fosmids from the competition were identical. Looking at the model, the probability of selecting pCC1FOS-96 at the 48 hour time step was P(pCC1FOS-96) = 0.068, also a relatively low probability. In fact, the probability that all three randomly selected fosmids were pCC1FOS-96 is $P = 0.068^3 = 0.00031$. Based on what the model showed, we were not expecting to see convergence to a single fosmid at this point.

The model showed convergence happening at a later time step. This is because the standard deviation of the d_i values (the different growth rates of the strains in the library) was too low. What happened in the experiment was that after 48 hours there was enough of a competitive advantage for the cells containing the pCC1FOS-96 fosmid to begin dominating the culture. Having a higher standard deviation in the model would account for the fact that some cells in the library have an extreme competitive advantage over the others. The change made to better match the model to our experimental results is illustrated in Figure 24.

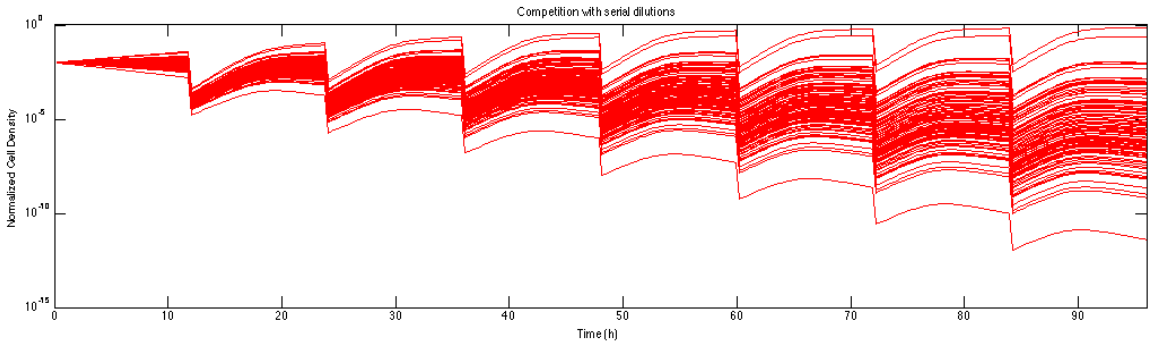


Figure 24. Showing an updated version of the model where the mean of the d_i distribution = 0.75 and the standard deviation has been changed to 0.05.

In this case, the model has clear "winners" of the competition at a much earlier time step (~48 hours), similar to what we observed experimentally.

4.4. Performing the Competition with a *Pseudomonas aeruginosa* Library

We did another test of the competition experiment with the *P. aeruginosa* plasmid library. The plasmid library was transformed into MG1655 *E. coli* cells and the resulting colonies were then pooled together to form the starting population of cells that would make up the library. During the experimental setup phase, it was important to try and prevent any cell growth within the population prior to the addition of biofuel. This is because tolerance mechanisms could be costly and detrimental to the cell in the absence of biofuel. These cells could very quickly be outcompeted by others in the population if biofuel was not present. Therefore, we transformed the plasmid library into *E. coli* and

used colonies directly from the agar plates for our experiments. This prevented any unwanted competition between cells before we started our testing.

We ran the competition experiment in the presence of 0.2% limonene (v/v) for 96 hours with serial dilutions into fresh media every 12 hours. This concentration of limonene was selected because it inhibited growth without completely killing *E. coli* cells not containing the library. Along with the population grown in the presence of biofuel, a separate population was grown in the absence of biofuel as a control to represent the natural competition occurring within the population. Cultures were plated every 24 hours and at the end of the competition, plasmids from five colonies on the 96 hour plate were isolated and sequenced to determine what *P. aeruginosa* DNA they contained. Interestingly, all of the sequenced plasmids from the 96 hour biofuel exposed plate contained the same five genes from *P. aeruginosa.* Each potential tolerance gene was then subcloned into a separate vector in order to determine which of the genes was responsible for the increased limonene tolerance observed in *E. coli.* We used a vector (pBbA5k) from the BioBricks series vectors, placing the gene downstream of a LacI-repressed promoter (Lee, Krupa et al. 2011). We supplemented the medium with 100 µM of IPTG to induce expression. We did tolerance testing on each gene and, as shown in Figure 25, PA2850 was much better than the rest at improving *E. coli's* tolerance to 0.2% limonene. The negative control (MG1655/pBbA5k-rfp) contained red florescent protein (*rfp*) instead of transgenic DNA. It showed poor growth when exposed to limonene. The positive control, the winning plasmid obtained after 96 hours of the

competition experiment was retransformed into fresh MG1655 cells to eliminate the possibility that any observed tolerance increases were because of mutations on the genome. The positive control plasmid containing genes PA2847-PA2851 grew well in the presence of biofuel and showed similar tolerance levels to the cells expressing only gene PA2850.

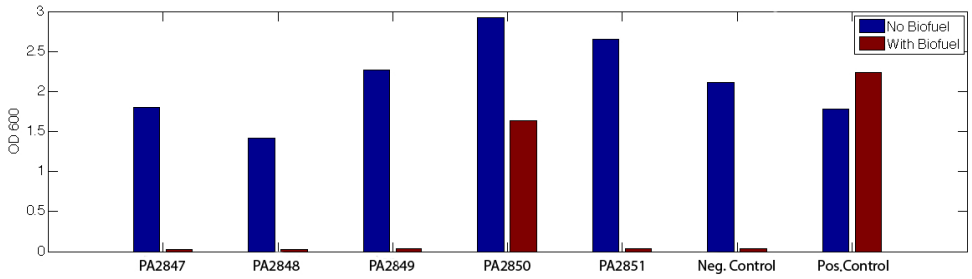


Figure 25: Tolerance testing in 0.2% limonene of individual genes recovered from the *P. aeruginosa* library after the competition experiment. Limonene was added at time zero and gene expression was induced with 100µM of IPTG at the same time. OD600 measurements were taken after 16 hours of growth. The negative control (pBbA5k-rfp) contains only a red florescent protein gene. The positive control contains all five genes recovered from the competition experiment.

4.5. Ohr from *P. aeruginosa* Increase Limonene Tolerance in *E. coli*

PA2850 encodes an organic hydroperoxide resistance protein Ohr (YP_790323.1), and was clearly the gene responsible for the improved limonene tolerance. As previously discussed in Chapter 3 with *P. putida*, Ohr works to neutralize oxidative species produced

during bacterial aerobic respiration or by the host immune system as a defense mechanism (Lesniak, Barton et al. 2002). It may function by converting the oxidative molecule into a less toxic form. In this case, Ohr may alleviate oxidative stress in *E. coli* caused by the addition of limonene. Further testing of *ohr* in Figure 26 showed improved growth versus the *rfp* negative control up to 0.4% limonene, validating the role that *ohr* plays in alleviating biofuel stress.

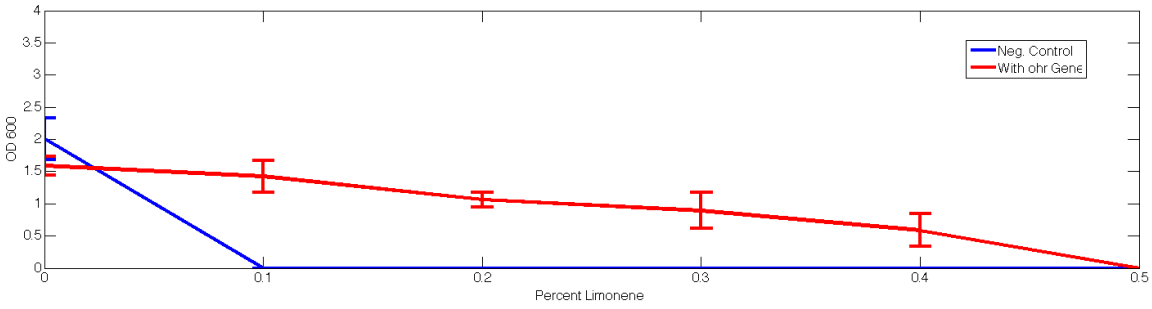


Figure 26: Testing the tolerance range of pBbA5k-ohr in *E. coli* MG1655 versus the negative control (pBbA5k-rfp). Both samples were induced with 100 µM of IPTG. Limonene was added at time zero and OD600 measurement were taken after 16 hours of growth. The experiment was performed in triplicate with the error bars representing standard deviation.

4.6. Additional Experimental Directions Considered

In this dissertation, we used a library approach to screen the genomes of three microbes that have adapted to survive in harsh environments, *M. aquaeolei*, *P. putida*

and *P. aeruginosa* for genes that impart tolerance to potential bio-jet fuel precursors. We showed that expression of the gene *ohr* from *P. aeruginosa* and *yceI* from *M. aquaeolei* successfully increased the tolerance of *E. coli* MG1655 to limonene and pinene/terpinolene respectively. Also, with the addition of the heterologous sigma factor *rpoD* from *P. putida* we were able to improve the expression of genes in out *P. putida* libraries. By accomplishing this, the diversity of samples isolated after a selection experiment greatly increased, when *rpoD* was expressed moderately.

In the future, it would be beneficial to test both *ohr* and *yceI* in limonene and pinene production strains, respectively, to quantify the potential titer increase that could be achieved. Also, combining these two biofuel tolerance genes with other known tolerance mechanisms has the potential to synergistically increase cell survival at higher biofuel concentrations. This has been shown in the past to be a successful strategy for boosting tolerance to higher levels than could achieved by one mechanism alone (Turner and Dunlop 2014) .

This method of library construction and competition could also be used to discover additional tolerance genes. Other hydrocarbon resistant microbes such as *A. borkumensis* or the strains isolated during the Deepwater Horizon disaster, would make excellent sources of genetic material for library construction. Construction of this library was attempted previously in our lab, but extracting high quality genomic DNA suitable

for library construction was difficult. However, this remains to be an interesting target for researchers in the future.

4.7. Testing Alternative Heterologous Sigma Factors

Another interesting experiment we tried, although it ultimately did not yield any new tolerance genes, was to probe the effect of other heterologous sigma factors responsible for aiding in the transcription of gene families. In this dissertation, we focused on the general sigma factor $rpoD_{P.\ putida}$. Many other sigma factors play important roles and are triggered in response to events such as stress or heat shock. Examples include σ^{22} (also called *algT*, *algU*, or *rpoE*) related to osmotic stress (Schnider-Keel, Lejbølle et al. 2001, Kahlon 2016), σ^{32} (*rpoH*) the heat shock sigma factor (Grossman, Straus et al. 1987), and σ^{38} (*rpoS*) the starvation/stationary phase sigma factor (Loewen, Hu et al. 1998). In theory, expressing these sigma factors with heterologous plasmid and fosmid libraries could help to access genomic space that was previously inaccessible. These sigma factor genes were individually put on a plasmid, in the same manner that pBbE8k-rpoD constructed, and the individual plasmids were inserted into cells containing the *P. putida* plasmid library. We used the following primer sequences shown in Table 2 to amplify the sigma factor genes from the *P. putida* genome.

σ Factor	Forward Primer Site	Reverse Primer Site
rpoE	5’ ATGCTAACCCAGGAAGAGGA 3’	5’ CGTTGTTGCAGGAAACCTGA 3’
rpoH	5’ ATGACCACATCGTTGCAAC 3’	5’ TCAGGCAGCGATCAGTGC 3’
rpoN	5’ ATGAAACCATCGCTCGTCCT 3’	5’ CTACATCAGTCGCTTGCGTT 3′
rpoS	5’ ATGGCTCTCAGTAAAGAAGTGCC 3’	5’ CAGCGAGTCATTGTTCCAGTAG 3’

Table 2: The primer site sequences used to amplify the alternative sigma factor genes from *P. putida*.

A selection experiment was run, again similar to what was done with the previous *rpoD* experiments, and all of the samples sequenced at the final time step contained the tolerance gene *ohr* (Table 3). This indicates that the alternative sigma factors we tested had no effect in expanding the accessibility of genes in our *P. putida* plasmid library.

Selection with Alternative Sigma Factor Samples	Tolerane Gene Candidate
0μM_rpoE_A	*ohr*
0μM_rpoH_A	*ohr*
0μM_rpoN_A	*ohr*
0μM_rpoS_A	*ohr*
10μM_rpoE_A	---
10μM_rpoE_B	*ohr*
10μM_rpoH_A	*ohr*
10μM_rpoH_B	*ohr*
10μM_rpoN_A	---
10μM_rpoN_B	*ohr*
10μM_rpoS_A	*ohr*
10μM_rpoS_B	*ohr*
100μM_rpoE_A	*ohr*
100μM_rpoE_B	*ohr*
100μM_rpoE_C	*ohr*
100μM_rpoH_A	*ohr*
100μM_rpoH_B	*ohr*
100μM_rpoH_C	*ohr*
100μM_rpoN_A	*ohr*
100μM_rpoN_B	*ohr*
100μM_rpoN_C	*ohr*
100μM_rpoS_A	*ohr*
100μM_rpoS_B	---
100μM_rpoS_C	*ohr*

Table 2: Additional sigma factors were tested in conjunction with the *P. putida* plasmid library. The four new libraries were stressed with pinene (0.05% v/v) with varying levels of *rpoE, rpoH, rpoN* and *rpoS* expression controlled by altering arabinose levels from 0, 10 and 100μM. In the table "---" means that sequencing did not yield a conclusive result for the particular sample.

4.8. Combining Multiple Libraries for Better Selection of Tolerance Genes

Another set of experiments that were attempted, but never produced any noteworthy results were based around the idea of combining multiple libraries together. The main obstacle that prevented this experiment from moving forward was the technical

challenge of getting the required number of clones for full library coverage. With a single library, the number of clones required can be calculated easily using the Clark-Carbon equation, as discussed previously. However, for this experiment we wanted each cell to carry one piece from two separate libraries. For instance, when combining the *P. putida* plasmid and fosmid libraries, each cell would contain a unique plasmid and a fosmid. In order to do this with all potential combinations of plasmids and fosmids represented, millions of colonies would need to be collected. Low efficiencies (~25,000 clones/transformation) in the transformation step prevented us from completing this experiment. Nevertheless, we are still hopeful that combining the two libraries and then running a selection experiment could produce interesting results. We theorized that by designing the experiment in this manner, we could isolate genes that worked synergistically, but were not close enough to each other in the genome to be captured by a single plasmid or fosmid in a library. More specifically, we thought that a possible outcome would be that the fosmid would perhaps contain a tolerance mechanism from *P. putida* and the plasmid would contain the appropriate sigma factor to enhance the expression of that tolerance mechanism. The other likely outcome of this experiment could have been increased tolerance due to combinatorial effects of combining multiple tolerance mechanisms. For instance, perhaps the selection would have produced a plasmid that contained *ohr* and a fosmid that contained an efflux pump system like Pp_3456. That scenario is unlikely to happen however without also introducing $rpoD_{P. putida}$ into the cells as well.

CHAPTER 5: CONCLUSION

5.1. Summary of Important Results

In this dissertation we have shown the advantages and available benefits of screening genomic libraries for biofuel tolerance genes. Selecting a good candidate microbe to screen played an important role in discovering useful tolerance genes. The candidate must have adaptations that allow it to survive in harsh environments. For this work we screened three microbes, *M. aquaeolei, P. putida* and *P. aeruginosa* for potential biofuel tolerance genes due to their propensity to thrive in stressful, hydrocarbon rich environments.

The work with the *M. aquaeolei* fosmid library lead to the discovery of *yceI* as a gene that imparts tolerance to pinene and terpinolene when expressed in *E. coli*. It is possible that the beta-barrel YceI protein binds with the hydrophobic biofuel compounds and alleviates stress through sequestration. A significant advantage to *yceI* is that high expression of this gene does not have any deleterious effects like an efflux pump would. A future target for engineering is to express *yceI* in a biofuel production strain to determine what the potential increase in titer could be. Also combining *yceI* with other known tolerance genes to produce combinatorial tolerance effects may also be worth pursuing.

The work with the *P. putida* library initially did not show the diversity of tolerance genes that were known to exist from the literature. We inserted a plasmid into the library cells that contained the sigma factor $rpoD_{P.\ putida}$ controlled via an arabinose responsive promoter. We immediately saw an increase in the diversity of competition

"winners" when RpoD$_{P.\ putida}$ was present in moderate amounts. A number of unique known and unknown tolerance genes were uncovered and tested to determine what tolerance benefit they each imparted. This work was done using both plasmid and fosmid libraries with unique and different tolerance genes arising from both libraries.

Additional work with genomic libraries was performed and may give other researchers ideas about future directions to continue in this field. We screened a *P. aeruginosa* plasmid library in limonene and uncovered the tolerance gene *ohr*, similar to what was seen in the experiments with the *P. putida* library. No other tolerance genes were identified via that library, however there was no heterologous sigma factor expression done in this case. We also performed tests using the *P. putida* plasmid library in combination with alternative sigma factors (*rpoE, rpoH, rpoN* and rpoS) from *P. putida.* The expression of these sigma factor genes did not lead to an increased diversity of winners of the competition experiment. Again, *ohr* was the only tolerance gene isolated when using the alternative sigma factors instead of *rpoD*$_{P.putida}$.

5.2. Final Remarks

Given the importance of liquid fuels in our society, it is critical to ensure that we have reliable and economical sources of these fuels for the future. Identifying better ways to produce advanced biofuels will help to lower the cost of their production and make biofuels more competitive with petroleum-derived fuels in the future. With particular

attention to detail, biofuels can be produced so that the net amount of greenhouse gasses emitted is lower compared to their petroleum sourced counterparts. A step in this direction is essential if our society wishes to stop the harmful phenomenon of global warming.

References:

Alonso-Gutierrez, J., R. Chan, T. S. Batth, P. D. Adams, J. D. Keasling, C. J. Petzold and T. S. Lee (2013). "Metabolic engineering of Escherichia coli for limonene and perillyl alcohol production." Metabolic Engineering **19**: 33-41.

Alonso-Gutierrez, J., E.-M. Kim, T. S. Batth, N. Cho, Q. Hu, L. J. G. Chan, C. J. Petzold, N. J. Hillson, P. D. Adams and J. D. Keasling (2015). "Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering." Metabolic engineering **28**: 123-133.

Alper, H., J. Moxley, E. Nevoigt, G. R. Fink and G. Stephanopoulos (2006). "Engineering yeast transcription machinery for improved ethanol tolerance and production." Science **314**(5805): 1565-1568.

Alsaker, K. V., C. Paredes and E. T. Papoutsakis (2010). "Metabolite stress and tolerance in the production of biofuels and chemicals: Gene-expression-based systems analysis of butanol, butyrate, and acetate stresses in the anaerobe Clostridium acetobutylicum." Biotechnol Bioeng: n/a-n/a.

Atsumi, S., T.-Y. Wu, I. M. P. Machado, W.-C. Huang, P.-Y. Chen, M. Pellegrini and J. C. Liao (2010). "Evolution, genomic analysis, and reconstruction of isobutanol tolerance in Escherichia coli." Mol. Syst. Biol. **6**: 449.

Boyarskiy, S., S. D. López, N. Kong and D. Tullman-Ercek (2016). "Transcriptional feedback regulation of efflux protein expression for increased tolerance to and production of n-butanol." Metabolic engineering **33**: 130-137.

Brennan, T. C. R., C. D. Turner, J. O. Krömer and L. K. Nielsen (2012). "Alleviating monoterpene toxicity using a two-phase extractive fermentation for the bioproduction of jet fuel mixtures in Saccharomyces cerevisiae." Biotechnol Bioeng **109**(10): 2513-2522.

Camilli, R., C. M. Reddy, D. R. Yoerger, B. A. S. Van Mooy, M. V. Jakuba, J. C. Kinsey, C. P. McIntyre, S. P. Sylva and J. V. Maloney (2010). "Tracking hydrocarbon plume transport and biodegradation at Deepwater Horizon." Science **330**(6001): 201-204.

Chubukov, V., A. Mukhopadhyay, C. J. Petzold, J. D. Keasling and H. G. Martín (2016). "Synthetic and systems biology for microbial production of commodity chemicals." npj Systems Biology and Applications **2**: 16009.

Chuck, C. J. and J. Donnelly (2014). "The compatibility of potential bioderived fuels with Jet A-1 aviation kerosene." Appl Energy **118**: 83-91.

Clarke, L. and J. Carbon (1976). "A colony bank containing synthetic Col El hybrid plasmids representative of the entire E. coli genome." Cell **9**(1): 91-99.

Doshi, R., T. Nguyen and G. Chang (2013). "Transporter-mediated biofuel secretion." Proceedings Of The National Academy Of Sciences Of The United States Of America **110**(19): 7642-7647.

Dunlop, M. J. (2011). "Engineering microbes for tolerance to next-generation biofuels." Biotechnol Biofuels **4**: 32.

Dunlop, M. J., Z. Y. Dossani, H. L. Szmidt, H. C. Chu, T. S. Lee, J. D. Keasling, M. Z. Hadi and A. Mukhopadhyay (2011). "Engineering microbial biofuel tolerance and export using efflux pumps." Mol. Syst. Biol. **7**: 487.

Dunlop, M. J., J. D. Keasling and A. Mukhopadhyay (2010). "A model for improving microbial biofuel production using a synthetic feedback loop." Systems and synthetic biology **4**(2): 95-104.

Fichan, I., C. Larroche and J. Gros (1999). "Water solubility, vapor pressure, and activity coefficients of terpenes and terpenoids." Journal of Chemical & Engineering Data **44**(1): 56-62.

Fiocco, D., V. Capozzi, P. Goffin, P. Hols and G. Spano (2007). "Improved adaptation to heat, cold, and solvent tolerance in Lactobacillus plantarum." Appl. Microbiol. Biotechnol. **77**(4): 909-915.

Fischer, C. R., D. Klein-Marcuschamer and G. Stephanopoulos (2008). "Selection and optimization of microbial hosts for biofuels production." Metabolic Engineering **10**(6): 295-304.

Foo, J. L., H. M. Jensen, R. H. Dahl, K. George, J. D. Keasling, T. S. Lee, S. Leong and A. Mukhopadhyay (2014). "Improving microbial biogasoline production in Escherichia coli using tolerance engineering." MBio **5**(6): e01932.

Foo, J. L. and S. S. J. Leong (2013). "Directed evolution of an E. coli inner membrane transporter for improved efflux of biofuel molecules." Biotechnology For Biofuels **6**(1): 81.

Fortman, J. L., S. Chhabra, A. Mukhopadhyay, H. Chou, T. S. Lee, E. Steen and J. D. Keasling (2008). "Biofuel alternatives to ethanol: pumping the microbial well." Trends Biotechnol. **26**(7): 375-381.

Fujita, M., Y. Hanaura and A. Amemura (1995). "Analysis of the rpoD gene encoding the principal sigma factor of Pseudomonas putida." Gene **167**(1): 93-98.

Gaida, S. M., N. R. Sandoval, S. A. Nicolaou, Y. Chen, K. P. Venkataramanan and E. T. Papoutsakis (2015). "Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous genomic libraries." Nature communications **6**.

Gibson, D. G., L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison and H. O. Smith (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases." Nat Methods **6**(5): 343-345.

Godoy, P., A. J. Molina-Henares, J. De La Torre, E. Duque and J. L. Ramos (2010). "Characterization of the RND family of multidrug efflux pumps: in silico to in vivo confirmation of four functionally distinct subgroups." Microbial biotechnology **3**(6): 691-700.

Grossman, A. D., D. B. Straus, W. A. Walter and C. A. Gross (1987). "Sigma 32 synthesis can regulate the synthesis of heat shock proteins in Escherichia coli." Genes & Development **1**(2): 179-184.

Handa, N., T. Terada, Y. Doi-Katayama, H. Hirota, J. R. H. Tame, S.-Y. Park, S. Kuramitsu, M. Shirouzu and S. Yokoyama (2005). "Crystal structure of a novel polyisoprenoid-binding protein from Thermus thermophilus HB8." Protein Sci. **14**(4): 1004-1010.

Harrison, M. E. and M. J. Dunlop (2014). "Synthetic feedback loop model for increasing microbial biofuel production using a biosensor." Synthetic biology applications in industrial microbiology: 58.

Harvey, B. G., M. E. Wright and R. L. Quintana (2009). "High-density renewable fuels based on the selective dimerization of pinenes." Energy & Fuels **24**(1): 267-273.

Hazen, T. C., E. A. Dubinsky, T. Z. DeSantis, G. L. Andersen, Y. M. Piceno, N. Singh, J. K. Jansson, A. Probst, S. E. Borglin, J. L. Fortney, W. T. Stringfellow, M. Bill, M. E. Conrad, L. M. Tom, K. L. Chavarria, T. R. Alusi, R. Lamendella, D. C. Joyner, C. Spier, J. Baelum, M. Auer, M. L. Zemla, R. Chakraborty, E. L. Sonnenthal, P. D'haeseleer, H.-Y. N. Holman, S. Osman, Z. Lu, J. D. Van Nostrand, Y. Deng, J. Zhou and O. U. Mason (2010). "Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria." Science (New York, NY) **330**(6001): 204-208.

Holtwick, R., F. Meinhardt and H. Keweloh (1997). "cis-trans isomerization of unsaturated fatty acids: cloning and sequencing of the cti gene from Pseudomonas putida P8." Applied and environmental microbiology **63**(11): 4292-4297.

Hong, M.-E., K.-S. Lee, B. J. Yu, Y.-J. Sung, S. M. Park, H. M. Koo, D.-H. Kweon, J. C. Park and Y.-S. Jin (2010). "Identification of gene targets eliciting improved alcohol tolerance in Saccharomyces cerevisiae through inverse metabolic engineering." Journal of Biotechnology **149**(1-2): 52-59.

Huu, N. B., E. B. Denner, D. T. Ha, G. Wanner and H. Stan-Lotter (1999). "Marinobacter aquaeolei sp. nov., a halophilic bacterium isolated from a Vietnamese oil-producing well." Int J Syst Bacteriol **49 Pt 2**: 367-375.

Johnson, T. J., J. L. Gibbons, L. Gu, R. Zhou and W. R. Gibbons (2016). "Molecular genetic improvements of cyanobacteria to enhance the industrial potential of the microbe: A review." Biotechnology progress **32**(6): 1357-1371.

Junker, F. and J. L. Ramos (1999). "Involvement of the cis/trans isomerase Cti in solvent resistance of Pseudomonas putida DOT-T1E." J. Bacteriol. **181**(18): 5693-5700.

Kahlon, R. S. (2016). Pseudomonas: Molecular and Applied Biology, Springer.

Kang, M.-K., J.-H. Eom, Y. Kim, Y. Um and H. M. Woo (2014). "Biosynthesis of pinene from glucose using metabolically-engineered Corynebacterium glutamicum." Biotechnology letters **36**(10): 2069-2077.

Korman, T. P., P. H. Opgenorth and J. U. Bowie (2017). "A synthetic biochemistry platform for cell free production of monoterpenes from glucose." Nature Communications **8**: 15526.

Lee, S. K., H. Chou, T. S. Ham, T. S. Lee and J. D. Keasling (2008). "Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels." Current Opinion in Biotechnology **19**(6): 556-563.

Lee, T. S., R. A. Krupa, F. Zhang, M. Hajimorad, W. J. Holtz, N. Prasad, S. K. Lee and J. D. Keasling (2011). "BglBrick vectors and datasheets: A synthetic biology platform for gene expression." Journal of Biological Engineering **5**(1): 12.

Lee, T. S., R. A. Krupa, F. Zhang, M. Hajimorad, W. J. Holtz, N. Prasad, S. K. Lee and J. D. Keasling (2011). "BglBrick vectors and datasheets; a synthetic biology platform for gene expression." Journal of Biological Engineering **5**(1): 12.

Lesniak, J., W. A. Barton and D. B. Nikolov (2002). "Structural and functional characterization of the Pseudomonas hydroperoxide resistance protein Ohr." The EMBO journal **21**(24): 6649-6659.

Lesniak, J., W. A. Barton and D. B. Nikolov (2002). Structural and functional characterization of the Pseudomonas hydroperoxide resistance protein Ohr. EMBO J. **21:** 6649-6659.

Li, X. Z., L. Zhang and K. Poole (1998). "Role of the multidrug efflux systems of Pseudomonas aeruginosa in organic solvent tolerance." J. Bacteriol. **180**(11): 2987-2991.

Liu, J., Y. Zhu, G. Du, J. Zhou and J. Chen (2013). "Response of Saccharomyces cerevisiae to D-limonene-induced oxidative stress." Applied microbiology and biotechnology **97**(14): 6467-6475.

Llamas, M. A., J. J. Rodríguez-Herva, R. E. Hancock, W. Bitter, J. Tommassen and J. L. Ramos (2003). "Role of Pseudomonas putida tol-oprL gene products in uptake of solutes through the cytoplasmic membrane." Journal of bacteriology **185**(16): 4707-4716.

Loewen, P. C., B. Hu, J. Strutinsky and R. Sparling (1998). "Regulation in the rpoS regulon of Escherichia coli." Canadian journal of microbiology **44**(8): 707-717.

Mukhopadhyay, A. (2015). "Tolerance engineering in bacteria for the production of advanced biofuels and chemicals." Trends in microbiology **23**(8): 498-508.

Nicolaou, S. A., S. M. Gaida and E. T. Papoutsakis (2010). "A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation." Metabolic Engineering **12**(4): 307-331.

Nicolaou, S. A., S. M. Gaida and E. T. Papoutsakis (2012). "Exploring the combinatorial genomic space in Escherichia coli for ethanol tolerance." Biotechnol J **7**(11): 1337-1345.

Paget, M. S. and J. D. Helmann (2003). "The σ 70 family of sigma factors." Genome biology **4**(1): 1.

Patten, C. L. and B. R. Glick (2002). "Regulation of indoleacetic acid production in Pseudomonas putida GR12-2 by tryptophan and the stationary-phase sigma factor RpoS." Canadian journal of microbiology **48**(7): 635-642.

Peabody, G. L., J. Winkler and K. C. Kao (2014). "Tools for developing tolerance to toxic chemicals in microbial systems and perspectives on moving the field forward and into the industrial setting." Current Opinion in Chemical Engineering **6**: 9-17.

Ramos, J. L., E. Duque, M.-T. Gallegos, P. Godoy, M. I. Ramos-Gonzalez, A. Rojas, W. Teran and A. Segura (2002). "Mechanisms of solvent tolerance in gram-negative bacteria." Annu. Rev. Microbiol. **56**: 743-768.

Ramos, J. L., E. Duque, P. Godoy and A. Segura (1998). "Efflux pumps involved in toluene tolerance in Pseudomonas putida DOT-T1E." Journal of bacteriology **180**(13): 3323-3329.

Ramos, J. L., E. Duque, J.-J. Rodríguez-Herva, P. Godoy, A. Haïdour, F. Reyes and A. Fernández-Barrero (1997). "Mechanisms for solvent tolerance in bacteria." Journal of Biological Chemistry **272**(7): 3887-3890.

Reyes, L. H., M. P. Almario and K. C. Kao (2011). "Genomic library screens for genes involved in n-butanol tolerance in Escherichia coli." PLoS ONE **6**(3).

Rojas, A., E. Duque, G. Mosqueda, G. Golden, A. Hurtado, J. L. Ramos and A. Segura (2001). "Three efflux pumps are required to provide efficient tolerance to toluene in Pseudomonas putida DOT-T1E." J. Bacteriol. **183**(13): 3967-3973.

Ruegg, T. L., E.-M. Kim, B. A. Simmons, J. D. Keasling, S. W. Singer, T. S. Lee and M. P. Thelen (2014). "An auto-inducible mechanism for ionic liquid resistance in microbial biofuel production." Nat Commun **5**: 3490.

Rühl, J., A. Schmid and L. M. Blank (2009). "Selected Pseudomonas putida strains able to grow in the presence of high butanol concentrations." Applied and environmental microbiology **75**(13): 4653-4656.

Sarria, S., B. Wong, H. García Martín, J. D. Keasling and P. Peralta-Yahya (2014). "Microbial synthesis of pinene." ACS Synth Biol **3**(7): 466-475.

Schneiker, S., V. A. P. M. dos Santos, D. Bartels, T. Bekel, M. Brecht, J. Buhrmester, T. N. Chernikova, R. Denaro, M. Ferrer, C. Gertler, A. Goesmann, O. V. Golyshina, F. Kaminski, A. N. Khachane, S. Lang, B. Linke, A. C. McHardy, F. Meyer, T. Nechitaylo, A. Pühler, D. Regenhardt, O. Rupp, J. S. Sabirova, W. Selbitschka, M. M. Yakimov, K. N. Timmis, F.-J. Vorhölter, S. Weidner, O. Kaiser and P. N. Golyshin (2006). "Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium Alcanivorax borkumensis." Nat Biotechnol **24**(8): 997-1004.

Schnider-Keel, U., K. B. Lejbølle, E. Baehler, D. Haas and C. Keel (2001). "The Sigma Factor AlgU (AlgT) Controls Exopolysaccharide Production and Tolerance towards Desiccation and Osmotic Stress in the Biocontrol Agent Pseudomonas fluorescensCHA0." Applied and environmental microbiology **67**(12): 5683-5693.

Schobert, H. H. (2013). The Chemistry of Hydrocarbon Fuels, Butterworth-Heinemann.

Shimada, T., Y. Yamazaki, K. Tanaka and A. Ishihama (2014). "The whole set of constitutive promoters recognized by RNA polymerase RpoD holoenzyme of Escherichia coli." PLoS One **9**(3): e90447.

Sisinni, L., L. Cendron, G. Favaro and G. Zanotti (2010). "Helicobacter pylori acidic stress response factor HP1286 is a YceI homolog with new binding specificity." FEBS J **277**(8): 1896-1905.

Stancik, L. M., D. M. Stancik and B. Schmidt (2002). "pH-dependent expression of periplasmic proteins and amino acid catabolism in Escherichia coli." Journal of Bacteriology.

Strogatz, S. (2005). Nonlinear dynamics and chaos with applications to physics, biology, chemistry, and engineering. Cambridge, Mass, Westview Press.

Tashiro, M., H. Kiyota, S. Kawai-Noma, K. Saito, M. Ikeuchi, Y. Iijima and D. Umeno (2016). "Bacterial production of pinene by a laboratory-evolved pinene-synthase." ACS synthetic biology **5**(9): 1011-1020.

Tomas, C. A., N. E. Welker and E. T. Papoutsakis (2003). "Overexpression of groESL in Clostridium acetobutylicum results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program." Applied and Environmental Microbiology **69**(8): 4951-4965.

Tomko, T. A. and M. J. Dunlop (2015). "Engineering improved bio-jet fuel tolerance in Escherichia coli using a transgenic library from the hydrocarbon-degrader Marinobacter aquaeolei." Biotechnology for biofuels **8**(1): 1.

Turner, W. J. and M. J. Dunlop (2014). "Trade-Offs in Improving Biofuel Tolerance Using Combinations of Efflux Pumps." ACS Synth Biol.

Uchiyama, T. and K. Miyazaki (2009). "Functional metagenomics for enzyme discovery: challenges to efficient screening." Current opinion in biotechnology **20**(6): 616-622.

Verhoef, S., H. Ballerstedt, R. J. Volkers, J. H. de Winde and H. J. Ruijssenaars (2010). "Comparative transcriptomics and proteomics of p-hydroxybenzoate producing Pseudomonas putida S12: novel responses and implications for strain improvement." Applied microbiology and biotechnology **87**(2): 679-690.

Wargo, M. J. and D. A. Hogan (2009). "Identification of genes required for Pseudomonas aeruginosa carnitine catabolism." Microbiology **155**(7): 2411-2419.

Wargo, M. J., B. S. Szwergold and D. A. Hogan (2008). "Identification of two gene clusters and a transcriptional regulator required for Pseudomonas aeruginosa glycine betaine catabolism." J. Bacteriol. **190**(8): 2690-2699.

Warren, R. L., J. D. Freeman, R. C. Levesque, D. E. Smailus, S. Flibotte and R. A. Holt (2008). "Transcription of foreign DNA in Escherichia coli." Genome research **18**(11): 1798-1805.

Weber, H., T. Polen, J. Heuveling, V. F. Wendisch and R. Hengge (2005). "Genome-Wide Analysis of the General Stress Response Network in Escherichia coli: S-Dependent Genes, Promoters, and Sigma Factor Selectivity." J. Bacteriol. **187**(5): 1591-1603.

Wood, K. B. and P. Cluzel (2012). "Trade-offs between drug toxicity and benefit in the multi-antibiotic resistance system underlie optimal growth of E. coli." BMC systems biology **6**(1): 1.

Woodruff, L. B. A., J. Pandhal, S. Y. Ow, A. Karimpour-Fard, S. J. Weiss, P. C. Wright and R. T. Gill (2013). "Genome-scale identification and characterization of ethanol tolerance genes in Escherichia coli." Metabolic Engineering **15**(C): 124-133.

Zhang, C., X. Hui, Y. Lin and C.-J. Sung (2016). "Recent development in studies of alternative jet fuel combustion: Progress, challenges, and opportunities." Renewable and Sustainable Energy Reviews **54**: 120-138.

Zingaro, K. A., S. A. Nicolaou, Y. Yuan and E. T. Papoutsakis (2014). "Exploring the heterologous genomic space for building, stepwise, complex, multicomponent tolerance to toxic chemicals." ACS Synth Biol **3**(7): 476-486.

APPENDIX

Appendix 1

Selection without RpoD	**Insert Start Position**	**Insert End Position**	**Potential Tolerance Gene Candidates** (Based on location and insert size)
0µM_rfp_A1	2,076,184	--	*ohr*
0µM_rfp_B1	2,082,864	--	*ohr*
0µM_rfp_C1	2,076,798	--	*ohr*
1µM_rfp_A1	2,040,459	2,045,404	*pyrF:* COG284 (PyrF) Orotidine-5'-phosphate decarboxylase [F] *Pp_1816:* COG2130 Putative NADP-dependent oxidoreductases [R] *Pp_1817:* COG1028 (FabG) Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases) [IQR] *Pp_1818:* hypothetical protein
1µM_rfp_C1	2,076,798	--	*ohr*
10µM_rfp_B1	2,074,779	--	*ohr*
10µM_rfp_C2	2,074,764	--	*ohr*
100µM_rfp_A1	2,084,339	--	*ohr*
100µM_rfp_C2	5,479,207	5,481,748	*prmA:* COG2264 (PrmA) Ribosomal protein L11 methylase [J] *Pp_4819:* hypothetical protein
Selection with RpoD	**Insert Start Position**	**Insert End Position**	**Potential Tolerance Gene Candidates** (Based on location and insert size)
0µM_rpoD_A1	2,078,625	--	*ohr*
0µM_rpoD_B1	2,075,337	--	*ohr*
0µM_rpoD_C1	2,074,784	--	*ohr*
1µM_rpoD_A2	4,908,095	4,994,115	*bkdB:* COG508 (AceF) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes [C]

			lpdV: COG1249 (Lpd) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes [C] *Pp_4405:* COG2199 FOG: GGDEF domain [T] *Pp_4312:* hypothetical protein *Pp_4313:* hypothetical protein *Pp_4314:* COG3791 Uncharacterized conserved protein [S] *Pp_4315:* COG384 Predicted epimerase, PhzC/PhzF homolog [R] *Pp_4316:* COG1052 (LdhA) Lactate dehydrogenase and related dehydrogenases [CHR] *Pp_4317:* hypothetical protein *Pp_4318:* hypothetical protein *Pp_4319:* hypothetical protein *Pp_4320:* hypothetical protein **(Approximate insert size: 4kb)**
1µM_rpoD_B1	2,076,133	--	*ohr*
1µM_rpoD_C1	2,074,766	--	*ohr*
10µM_rpoD_A1 (referred to as 10µM_rpoD_A in the main text)	1,400,873	3,611,923	*tolA:* biopolymer transport protein TolA *tolB:* COG823 (TolB) Periplasmic component of the Tol biopolymer transport system [U] *oprL:* COG2885 (OmpA) Outer membrane protein and related peptidoglycan-associated (lipo)proteins [M] *Pp_1224:* COG1729 Uncharacterized protein conserved in bacteria [S] *Pp_1225:* COG602 (NrdG) Organic radical activating enzymes [O] *Pp_1226:* COG603 Predicted PP-loop superfamily ATPase [R] *Pp_1227:* COG3965 Predicted Co/Zn/Cd cation transporters [P] *Pp_3143:* COG583 (LysR) Transcriptional regulator [K] *Pp_3144:* COG1760 (SdaA) L-serine deaminase [E] *Pp_3145:* conserved hypothetical protein *Pp_3146:* COG665 (DadA) Glycine/D-amino acid oxidases (deaminating) [E] *Pp_3147:* COG687 (PotD) Spermidine/putrescine-binding periplasmic protein [E] **(Approximate insert size: 3kb)**
10µM_rpoD_B1 (referred to as 10µM_rpoD_B in the main text)	264,579	3,325,971	*Pp_0219:* COG2011 (AbcD) ABC-type metal ion transport system, permease component [P]

			Pp_0220: COG1135 (AbcC) ABC-type metal ion transport system, ATPase component [P] *Pp_0221:* COG1464 (NlpA) ABC-type metal ion transport system, periplasmic component/surface antigen [P] *Pp_0222:* COG2141 Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases [C] *Pp_0223:* COG1960 (CaiA) Acyl-CoA dehydrogenases [I] *Pp_0224:* COG1960 (CaiA) Acyl-CoA dehydrogenases [I] **(Approximate insert size: 2.5kb)**
10µM_rpoD_C1	254,219	264,579	*Pp_0204:* COG2188 (PhnF) Transcriptional regulators [K] *Pp_0205:* COG1053 (SdhA) Succinate dehydrogenase/fumarate reductase, flavoprotein subunit [C] *Pp_0206:* COG1146 Ferredoxin [C] *Pp_0207:* COG715 (TauA) ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components [P] *Pp_0208:* COG600 (TauC) ABC-type nitrate/sulfonate/bicarbonate transport system, permease component [P] *Pp_0209:* COG1116 (TauB) ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component [P] *Pp_0210:* COG1413 FOG: HEAT repeat [C] *Pp_0211:* COG3536 Uncharacterized protein conserved in bacteria [S] *Pp_0212:* hypothetical protein *gabD:* COG1012 (PutA) NAD-dependent aldehyde dehydrogenases [C]
10µM_rpoD_B2	2,074,773	--	*ohr*
10µM_rpoD_C3	405,059	408,791	*Pp_0337:* COG5001 Predicted signal transduction protein containing a membrane domain, an EAL and a GGDEF domain [T] *aceF:* COG508 (AceF) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes [C]
100µM_rpoD_B1	1,554,529	3,769,293	*Pp_1362:* COG469 (PykF) Pyruvate kinase [G] *Pp_1363:* conserved hypothetical protein *Pp_1364:* conserved hypothetical protein

			Pp_1365: COG2925 (SbcB) Exonuclease I [L] *Pp_1366:* transcriptional regulator *Pp_1367:* COG788 (PurU) Formyltetrahydrofolate hydrolase [F] *Pp_1368:* COG392 Predicted integral membrane protein [S] *Pp_3329:* COG4325 Predicted membrane protein [S] *Pp_3330:* COG1629 (CirA) Outer membrane receptor proteins, mostly Fe transport [P] *Pp_3331:* conserved hypothetical protein **(Approximate insert size: 3kb)**
100µM_rpoD_A1	2,078,598	--	*ohr*
100µM_rpoD_B3	4,994,115	4,998,562	*bkdB:* COG508 (AceF) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes [C] *lpdV:* COG1249 (Lpd) Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes [C] *Pp_4405:* COG2199 FOG: GGDEF domain [T]
100µM_rpoD_C3	2,693,079	2,696,334	*csuC:* COG3121 (FimC) P pilus assembly protein, chaperone PapD [NU] *csuD:* COG3188 (FimD) P pilus assembly protein, porin PapC [NU]
100µM_rpoD_A5	5,182,053	5,188,548	*Pp_4561:* conserved hypothetical protein *Pp_4562:* conserved hypothetical protein *Pp_4563:* COG2996 Uncharacterized protein conserved in bacteria [S] *Pp_4564:* conserved hypothetical protein *aceK:* COG4579 Isocitrate dehydrogenase kinase/phosphatase [T] *Pp_4566:* membrane protein *Pp_4567:* hypothetical protein *Pp_4568:* membrane protein
1000µM_rpoD_A1	2,082,447	--	*ohr*
1000µM_rpoD_B1	2,086,702	--	*ohr*
1000µM_rpoD_C1	2,086,702	--	*ohr*

Appendix 2

Selection without RpoD	**Insert Start Position**	**Insert End Position**	**Known Tolerance Gene Candidates** (Based on location and insert size)
FosA1	2,041,821	2,082,192	*ohr*
FosA2	2,041,821	2,082,192	*ohr*
FosB1	3,594,543	2,096,380	*ohr*
FosB2	3,594,543	2,096,594	*ohr*
FosC1	2,069,717	2,111,474	*ohr*
Sample with RpoD	**Insert Start Position**	**Insert End Position**	
rpoD_FosA1	3,912,065	3,945,734	*Pp_3456*
rpoD_FosA2	2,057,403	2,539,218	
rpoD_FosA3	1,584,206	1,544,143	*ttgABC*
rpoD_FosB1	2,438,045	2,399,033	
rpoD_FosB2	2,438,045	2,399,033	
rpoD_FosB3	2,438,045	--	
rpoD_FosC1	1,469,748	1,431,620	*Pp_1271*
rpoD_FosC2	3,912,042	3,945,729	
rpoD_FosC3	2,075,733	2,040,789	