Phage Assisted Evolutions for Engineering Biosynthetic Pathways

Phage Assisted Evolutions for Engineering Biosynthetic Pathways

A dissertation presented

by

Timothy Brian Roth

to

The Chemical Biology Program at the Graduate School of Arts and Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemical Biology

Harvard University

Cambridge, Massachusetts

April 2018

Dissertation Advisor: Professor David R. Liu Timothy Brian Roth

Phage Assisted Evolutions for Engineering Biosynthetic Pathways

Abstract

Phage Assisted Continuous Evolution (PACE) and its derivative techniques are powerful tools for evolving improved and altered enzymatic activities. Prior to this work, no one had demonstrated the applicability of these techniques for metabolic pathway enzyme engineering. Here I demonstrate ways in which bacterial sensors of metabolically relevant phenotypes can be optimized and applied to the PACE system to evolve greater *in vivo* pathway activity based on individual improvements to target enzyme activities. First, I optimized a system for sensing polyhydroxyalkanoate inclusion bodies to evolve increased production of polyhydroxybutyrate (PHB) in *E. coli*. I demonstrated increased production either by evolving a single enzyme in a pathway or by evolving an entire transcript containing all three enzymes responsible for PHB production. I also engineered the *E. coli* formaldehyde repressor circuit to have greater sensitivity and used this sensor in a non-continuous variant of PACE to evolve a series of mutant *Bacillus methanolicus* methanol dehydrogenase genes in *E. coli*. I characterized this set of evolved dehydrogenases *in vitro* to show that they have significantly improved kinetic parameters, and, with collaborators in the Stephanopoulos lab, I also demonstrated that they enable us to incorporate roughly twice as much methanol through the ribulose monophosphate pathway in *E. coli* compared to the state-of-the-art methanol dehydrogenases under equivalent conditions. Together, the results of these projects demonstrate that in addition to its many previously reported capabilities, PACE can also be used to evolve a diverse set of metabolic enzyme activities.

Table of Contents

List of Figures

Acknowledgments

Liz Roth-Johnson, for always knowing exactly what I was experiencing throughout graduate school even when I didn't realize it. You've continued to be a trailblazer to me even past academia and I will never stop admiring your tenacity and accomplishments in science, education, and beyond.

Perry Roth-Johnson, for being a constant, positive presence even in the face of Liz and my interminable food science dialogues. I couldn't have asked for a better brother.

Jim and Molly Roth, my loving parents, without whom I honestly have no idea where I would be. It is an immense privilege to receive the love and support of a family throughout one's education, and the gravity of this realization has only increased with each year I've been in graduate school.

Travis Blum, **Tina Wang**, and **Ben Thuronyi**, three post-docs who have been there to help me find confidence in my work for the good results and the bad. I aspire to someday be as organized, tenacious, and committed to science as you three.

Humphrey Hu, Lisa 'Paco' Yan, Eric Yuan, and **Wei 'is fat' Chang**, for being a constant online presence to remind me that I will always be a failure and I will never have to be alone.

Daniel and **Colleen Cohen**, for being kind and supportive friends and roommates even as the fire department smashed in the windows of our apartment on New Year's Eve. Consider this penance for never buying a Jumbotron spot for your wedding.

Stephen Hinshaw and **Belinda Wang**, for allowing me to immediately land on my feet after watching the fire department smash in the windows of my apartment on New Year's Eve and for graciously allowing me space to write even as I descended into the neurotic depths only a graduate thesis can induce.

Pia Sörensen, Mai Nguyen, Casey Roehrig, and Charlie Margarit, for allowing me so many opportunities to indulge in my passion for food and education and for teaching me so much about how to be a better educator.

Ahmed Badran and **David Thompson**, for teaching me damn near everything I know about molecular biology, and **Ben Woolston** for reminding me of how much I have yet to learn in fields I aim to explore next.

To all of my co-conspirators in the Liu Lab, from the spiciest of memelords to the dorkiest of dungeon crawlers, none of this would have been possible without your constant, positive presence in the lab and beyond. To all current graduate students, there is always a light at the end of the tunnel, regardless of what your pessimistic senior labmates might imply. To Alix Chan, Bill Kim, Zhen Chen, and Johnny Hu: Cheers to the light, I hope it brings us all the best. To all current, former, and future post-docs and graduate students of which there are too many to name: you have taught me more than I will ever be able to repay, so I will do my best to pay it forward to whomever I may work with next.

To my classmates in the Chemical Biology program and beyond, especially to **Sami Farhi** who has put up with me far more than any one person should have to.

Last and certainly not least, thank you to **David R. Liu** for giving me so much free rein with my research projects, and who understands more than most the value and benefit of fostering a positive lab environment. The group you have brought together is truly special, and I'm honored to have been a part of it.

Introduction

Microbial chemistry has been a key component of civilization for millennia[1,2]. In more recent years, as our ability to specifically engineer microbial chemistries at finer and finer scales has improved dramatically, society has developed complex bio-refining strategies, with entire industries centered around using microbes to produce commodity chemicals from inexpensive agricultural or petrochemical products and byproducts[3]. The field of metabolic engineering focuses on altering the complex networks of microbial chemistries to produce industrially relevant yields of commodity chemicals, fuels, and pharmaceuticals, ideally from cheap and renewable resources as opposed to the standard petrochemical pipelines they compete with. To this end, researchers adopt many strategies to improve three general components of an engineered microbial culture: titer, rate, and yield. As with all process chemistry, there are many routes to improve these factors – from broad considerations such as growth media and degree of oxygenation, the choice of microbial host, and the shape and size of a bioreactor to specific considerations such as the pathway enzymes used for production of a small molecule of interest, their relative localization to each other and cellular machinery, and the trade-offs they have between overproduction of desired product and impact on biomass accumulation in the host[4,5]. Many effective engineering strategies exist to control these first two factors, but improving enzyme activities is still frequently a difficult, expensive, and/or time-consuming endeavor[6].

Of the many selections and screens available for engineering improved enzymatic activities, *in vivo* techniques are notable for their relatively simple workflow. Designing *in vivo* selections for small molecule production in bacteria requires developing a way of continuously monitoring production or consumption of a target molecule by a target enzyme, a method of generating a diverse library of enzyme variants to test, and a way to select for higher activity variants over lower activity variants. The Liu lab's Phage Assisted Continuous Evolution (PACE) system is an ideal

platform on which to base such selections, as it allows for rapid *in vivo* library generation and selection based on a genetically-encoded sensor in *E. coli* coupled to the M13 bacteriophage life-cycle. This technique has been used to evolve a wide variety of phenotypes—RNA transcription[7–9], protein-protein interactions[10], aminoacyl-tRNA synthetase specificity[11], protease specificity[12]—but to date no one has used the system to evolve enzymes involved in small-molecule metabolism.

In this work, I describe two different selections I developed to evolve a variety of metabolic enzymes for improved activity based on PACE technologies. The first evolutionary target is a simple pathway for the synthesis of polyester inclusion bodies in bacteria. Polyhydroxyalkanoate (PHA) biosynthesis has long been a target product for metabolic engineering, and it makes a tempting target for phage-based evolution given the intracellular and relatively inert nature of these polymers. It also presents three potential evolutionary targets and a unique opportunity to co-evolve an entire pathway given its relatively compact and well-characterized component parts.

The second evolutionary target is methanol dehydrogenase, a key enzyme for the growth of bacterial cultures on single-carbon feedstocks such as methane gas or methanol. This enzyme has been implicated as a rate-determining step for methanol feeding in *E. coli*, and it has been proposed that a sufficiently active MDH variant would greatly improve flux through the ribulose monophosphate pathway required for efficient assimilation of methanol feedstocks into central metabolites.

These targets both present unique challenges for the PACE system and thus represent a significant expansion of what the Liu Lab's Phage Assisted Continuous Evolution (PACE) platform can accomplish. With this thesis, I hope to demonstrate that such phage-assisted *in vivo* evolutions can successfully evolve metabolic phenotypes toward key applications in the field of metabolic engineering. While more work is still required to bring these two projects to full fruition, I hope to demonstrate that PACE and its derivatives are valuable alternatives to the standard *in vitro* library assemblies and plate-based screens that are the standard in the industry.

Chapter 1: Phage-Assisted Evolution of Polhydroxybutyrate Pathways

Section 1.1: Introduction to Polyhydroxyalkanoate Pathway Engineering

Polyhydroxyalkanoates (PHAs) are a class of polyesters that are produced by a wide variety of microorganisms. First discovered and isolated in 1926[13], these hydrophobic inclusion bodies allow organisms that synthesize them to stockpile cellular carbon in times when it is overly abundant relative to other nutrients, a practice analogous to fat and glycogen synthesis in animals or starch granule formation in plants. PHA synthesis typically occurs through a relatively simple set of shunts to central metabolic processes that produce acetyl-CoA and fatty acids, with the specific chemical makeup of the polymers determined by both the chemical specificity of the PHA polymerase/synthase step, as well as the available pool of substrates present in the cell (**Figure 1**). The required substrates for all known polymerizations are hydroxyacyl-CoAs, with different PHA synthases typically having selective preference for different acyl chain lengths, typically produced during fatty acid biosynthesis and degradation[14] or through the two-step reaction mediated by thiolase and reductase genes encoded on the same operon as the synthase itself[15]. For example, the PHA synthase from *Cupriavidus necator* (formerly *Ralstonia eutropha*), PhaC, primarily catalyzes the polymerization of (*R*)-3-hydroxybutyryl-CoA produced from acetyl-CoA by the PhaA and PhaB genes encoded immediately downstream[16], whereas PHA synthases from *Pseudomonas* species preferentially catalyze the polymerization of 3-hydroxyacyl-CoAs with chain lengths of 6 to 10 carbons[17–20]. Engineering of these synthases has expanded this specificity to include lactic acid[21], 2-hydroxybutyrate[22], glycolate[23–25], and 4-hydroxybutyrate[23,26–28], enabling production of the corresponding polymers and co-polymers of each. Native producers of PHAs have a series of regulatory mechanisms to control expression of these PHA synthases, as well as dedicated, upstream components and a host of polymer-granule-associated factors[29]. The broadest class of these granule-associated proteins are the phasins, which can control granule morphology, modulate PHA synthase activity, and further regulate gene expression in response to granule synthesis to minimize their impact on other cellular functions[30–35]. In addition to phasins, regulatory proteins

control expression of PHA production operons in organisms such as *C. necator* and *Paracoccus denitrificans* by binding to promoter-encoding DNA sequences upstream of phasin and PHA biosynthetic genes. Typically, these repressors will de-repress their cognate promoters only in the presence of hydrophobic granule surfaces[31,34,36,37]. Finally, depolymerase proteins will associate to granule surfaces to allow the re-circulation of carbon metabolites for the organisms producing them[32,38–40], or in some cases allowing separate organisms to predate PHA-producing bacteria[41]. These depolymerases form the basis for PHA biodegradation, making PHA-derived materials an enticing alternative to petroleum-derived plastics, which typically do not have effective natural degradation pathways.

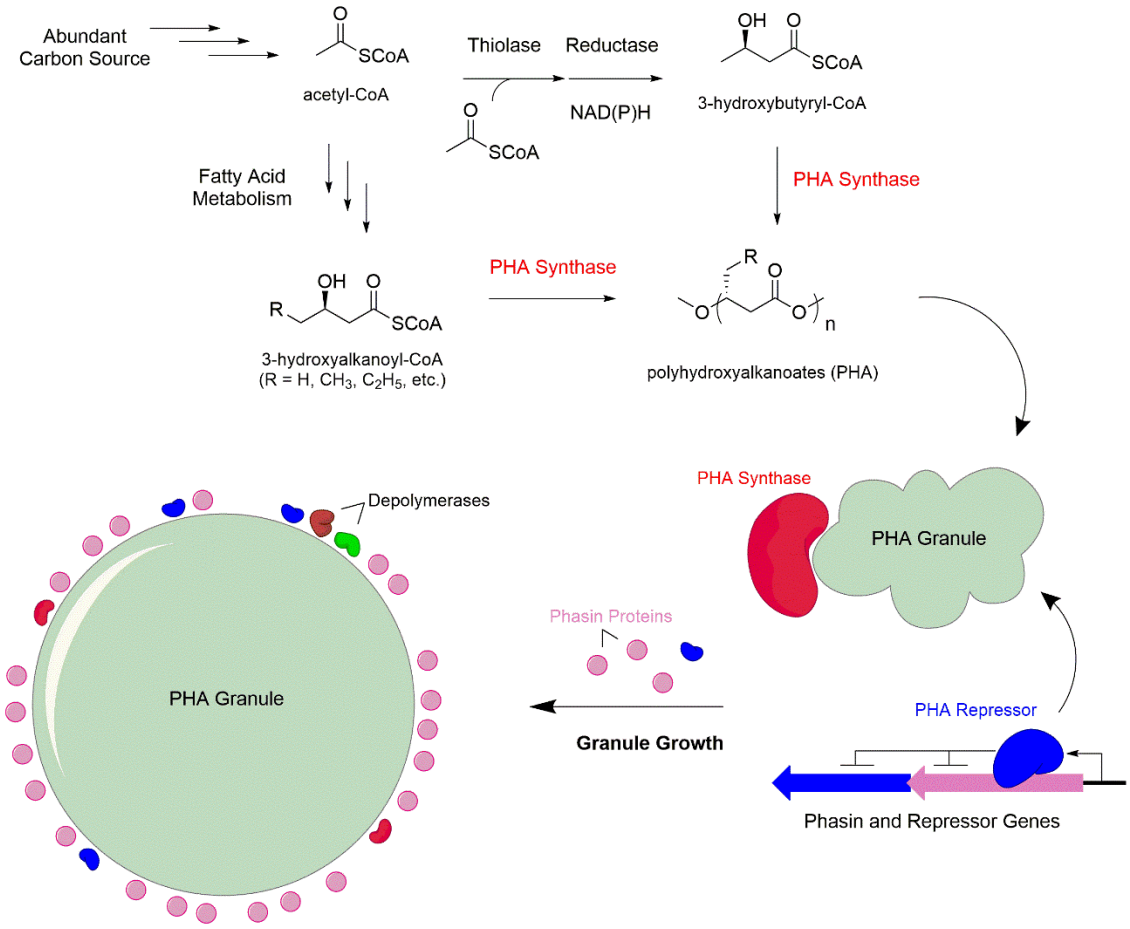


Figure 1: PHA granules are synthesized via several pathways and form inclusion bodies regulated by a wide array of proteins.

Due to the ecological impact and relative biosynthetic simplicity of PHAs, their production has been of interest to the bioengineering community for several decades. Pathway engineering and directed evolution of PHA production pathways has been a relatively straightforward process, with researchers commonly using established random mutagenesis approaches coupled to plate-based screens or selections[3]. As an example, Amara *et al.* used an *in vivo* random mutagenesis approach to isolate a mutant *Aeromonas punctata* PHA synthase that, with a single amino acid change, gave a nearly 5-fold increase in activity and a 20% gain in PHA accumulation with greater specificity for (*R*)-3-hydroxybutyryl-CoA as a substrate[42]. Another random mutagenesis approach, this time done *in vitro* using error-prone PCR, was used by Kichise *et al.* to find that a single-amino-acid change in a PHA synthase from *Aeromonas caviae* significantly increased its specific *in vitro* activity as well as the organism's overall ability to accumulate PHA *in vivo*[43]. More recent work by Sheu *et al.* took a more targeted approach at engineering these synthases through recombination of different synthase domains to create chimeric enzymes with the optimized phenotypes from different PHA-producing organisms[44].

Accurate detection of PHA accumulation and composition via chemical analysis requires digestion of the polymer granules, which can be separated from cell pellets or directly processed from dry cell mass for chromatographic quantitation[13,45,46]. High-throughput methods have been developed for this process[47] but are typically time and energy intensive compared to what more modern screen and selection techniques would require. To circumvent this, researchers hoping to quickly assess the PHA phenotype of a diverse population rely on visual cues for PHA production. Most simply, colony opacity under ambient light can be assessed and PHA-producing colonies picked and characterized, as demonstrated in a recent study screening for optimized RBS sequences in a PHA pathway in *E. coli*[48]. However, this requires growth on solid media and sufficient time to produce at least 10% dry cell weight of polymer to detect. As a result, most screens and selections use hydrophobic dyes such as Nile Red to stain granules directly and improve quantifiable contrast

between producers and non-producers[49–52]. Bacteria encoding diversified PHA-production gene libraries can be grown in solid or liquid-phase media containing dye or washed in concentrated dye solutions after producing polymers and monitored for total, relative granule content by monitoring fluorescence of the applied dye. This technique enables efficient techniques such as automated colony-picking and flow cytometry to pick out PHA overproducers. However, the relatively non-specific staining technique limits these approaches by requiring a relatively long period of polymer accumulation before significant contrast can be seen compared to non-PHA accumulating individuals in a population, due in part to the background staining of lipids in the cell membrane and growth media[52]. This limits the speed of these selections but aligns well with the goals of the industry—the rate of polymer production is typically far less important than yield, which is inherently limited compared to other commodity molecules produced microbially by the strictly intracellular nature of PHA inclusions. As a result, most selections are interested in altering the specificity of synthases toward the production of novel PHA materials from known metabolites to broaden the range of high-value polymer targets accessible to the field. If PHA is not present at incredibly high biomass percentages, it is prohibitively expensive to produce industrially compared to existing alternatives. However, *in vitro* applications of PHA enzyme cascades still stand to benefit from improved pathway kinetics. For example, a recent study used PHB synthesis as a model system to demonstrate *in vitro* redox cofactor recycling as a key component in maximizing flux in production[53]. Improving the stability, specific activity, or general kinetic behavior of PHA synthases and their upstream components could lend further viability to these *in vitro* enzyme assemblies**.**

Engineering work has also sought to capitalize on the granule-binding domains of PHA-associated factors. Several studies have been published using granule-binding domains from phasins, PHA-associated repressors and synthases as protein purification affinity tags. Target proteins are fused to these granule-associating factors via inteins and expressed in cells synthesizing PHA granules. The fusion products are co-purified with the protein-decorated

granules, and intein self-cleavage liberates the target proteins from the granule-bound domains[8,9]. These binding domains have also been proposed as a mechanism for surface functionalization of PHA nanoparticles for drug delivery[54], or even for more general biosurfactant applications in the absence of PHA granules given their highly amphiphilic nature[55].

Despite these parallel engineering efforts to produce PHAs and harness the properties of their associated factors, I could not find any publications using genetically-encoded biosensors to monitor and evolve PHA production in growing cell populations. While PHA production is a relatively easy phenotype to monitor *in vivo* using the methods described above, a genetically encoded sensor for granule formation would allow for simple and efficient monitoring of polymer production in real-time. Optimized reporters have been used to dynamically regulate pathway component expression and optimize yields[56], and sensitive reporter systems would provide an attractive, potentially high-contrast alternative to dye-based fluorescence detection in existing selection techniques. Furthermore, using PHA inclusions as a terminal readout would allow a variety of upstream pathway components to be evolved either individually or in combinations that ultimately provide a PHA synthase with viable substrate. This would be a versatile tool for engineering improved thiolases and reductases, or even for improved flux through fatty acid biosynthesis machinery and related process upstream of PHA granule formation (**Figure 1**). The promiscuity of PHA granule binding exhibited by some PhaR proteins[57] also suggests their utility in expanding the scope of PHA synthase activities beyond known moieties, helping make these biological syntheses more viable as competitors to petrochemical polymer production platforms. One of the fastest and most efficient platforms for applying *in vivo* biosensors for directed evolution is Phage Assisted Continuous Evolution, which to date has been used to evolve a broad range of phenotypes including RNA polymerases[7–9], proteases with altered specificity[12], DNA-binding domains[58], and protein-protein interactions[10]. This chapter will describe my efforts to develop a generalized PACE platform for a reporter-based selection for PHA synthesis in *E. coli* as illustrated

in **Figure 2.** The final section of this chapter will detail current efforts and propose strategies for more complex selections to expand the activity of these PHB pathways beyond known biopolymer production.

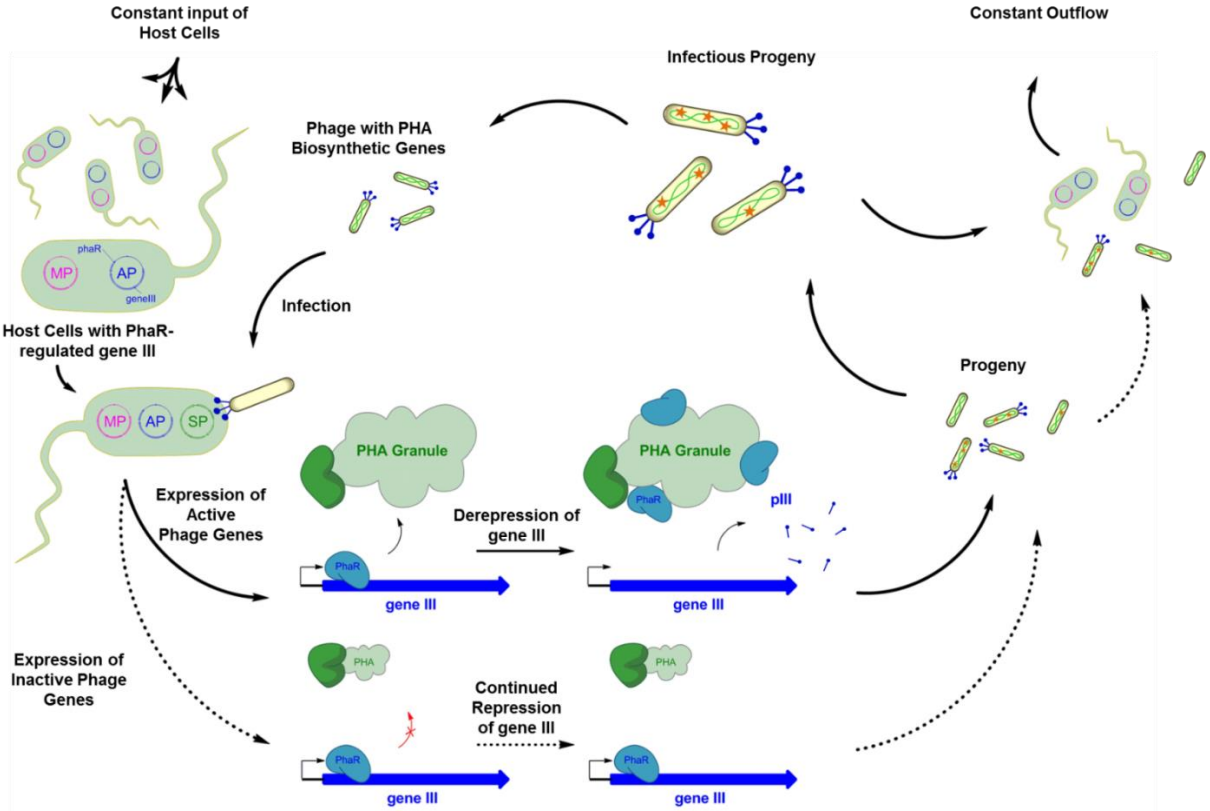


Figure 2 Generalized PACE selection for PHA biosynthesis.
Host cells are engineered with both the inducible Mutagenesis Plasmid (MP) and Accessory Plasmid (AP). The AP contains a genetic circuit that represses expression of essential phage gene III in the absence of hydrophobic granules surfaces and de-represses when such surfaces are present in sufficient quantities in the cell due to the activity of PHA production pathways. Selection Phage (SP) have been engineered to lack gene III and instead contain one or more essential genes involved in PHA biosynthesis. Active, phage-encoded pathway components will produce PHA granules, de-repress gene III from the AP, and allow production of pIII and subsequently infectious phage progeny, continuing the phage life-cycle. Inactive pathways will fail to produce infectious progeny and eventually wash out of the system. Pathway gene diversity is created by the mutagenic properties of the MP, allowing new pathway gene variants to be tested with each phage life cycle. Stars indicate mutations acquired during a single round of the phage life cycle.

Section 1.2: Development of a Polhydroxybutyrate (PHB) Biosensor

There are several classes of proteins in PHA-producing microorganisms that coat the surface of PHA granules to regulate their size and number, as well as to help separate them from the cytoplasmic environment. Of these, PHA repressor (PhaR) proteins are known to control the expression of other granule-associated proteins based on the presence or absence of PHA in the cell (**Figure 1**). The PhaR protein from *Paracoccus denitrificans* has been shown to bind *in vitro* to a wide range of hydrophobic polymers, including polyhydroxybutyrate (PHB), polystyrene, and polylactic acid, suggesting that it has a broad range of interactions based on polymer surface properties rather than specific PHA interactions[57]. In the absence of known, specific binders of PHA derivatives, this promiscuity makes PhaR ideal for evolving enzymes toward novel PHA production. I obtained the *P. denitrificans phaR* (*phaR*) gene from the Taguchi lab at the School of Engineering of Hokkaido University in Japan, which I then PCR amplified and subcloned downstream of a tetracycline inducible promoter on a low-copy vector, as shown in **Figure 3A**. On this same plasmid, I included a luciferase reporter under control of the previously identified PhaR DNA binding sequence[36]. I also obtained a codon-optimized version of the *Cupriavidus necator* (formerly *Ralstonia eutropha*) polyhydroxybutyrate (PHB) biosynthetic pathway from Michelle Chang's research group at University of California (UC) Berkeley. This pathway contains a codon-optimized version of the *phaABC* cassette from *C. necator*, along with *phaP* at the end of the transcript, which encodes the major phasin protein from this organism (**Figure 3B**). In place of its native promoter and operator sequences, this pathway is under control of the arabinose-inducible P_{BAD}/*araC* regulatory system, allowing for tunable expression of the pathway during experiments. I used luciferase assays based on the *Xenorhabdus luminescens xluxAB* reporter genes as previously described[7,8] to test the ability of PhaR to regulate gene expression in the presence or absence of PHB production from this model pathway (**Figure 3C**).

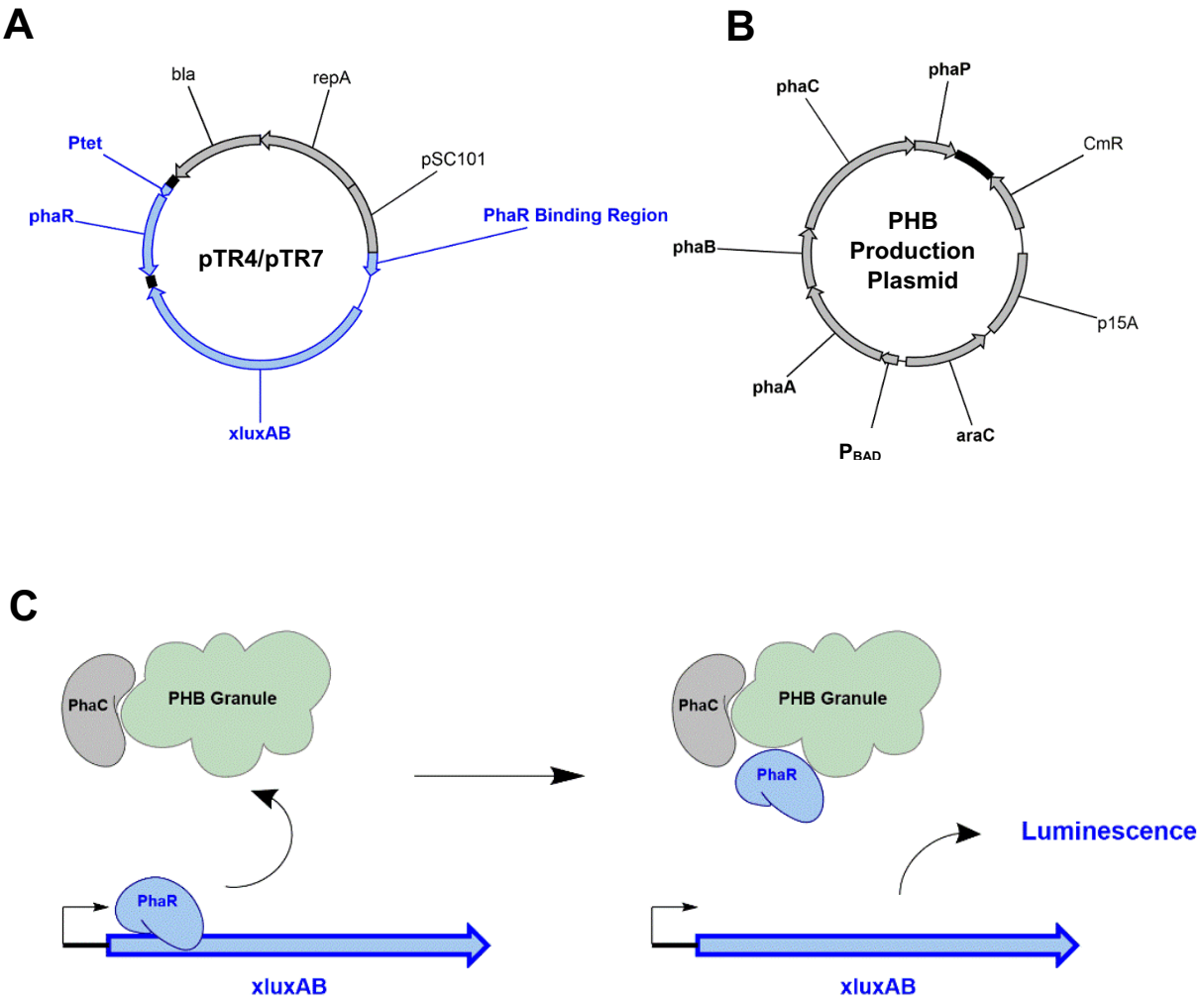


Figure 3: Reporter assay for PHB production in *E. coli*.
A. General map of the pTR4/pTR7 series reporter constructs expressing the *phaR* gene from *Paracoccus denitrificans* and **B.** General structure of the plasmid used to synthesize PHB during initial reporter characterization. bla and CmR represent antibiotic resistance markers, whereas repA, pSC101, and p15A represent origins of replication. **C.** Illustration of the PhaR-based sensor for PHB production. When PHB is synthesized, a direct and irreversible binding of PhaR will de-repress the luciferase reporter gene and allow transcription/translation of luciferase, which will produce a readable luminescence output.

I first placed the luciferase reporter gene downstream of the wild-type binding sequence from the *P. denitrificans* genome, which I then placed downstream of a series of previously characterized *E. coli* constitutive promoters[59]. I based this strategy on previous GFP-reporter constructs for PhaR characterization, which used a T7 promoter upstream of the native binding sequence to screen for inactive PhaR variants[60]. This approach assumes negligible activity from the promoter embedded within the binding sequence itself, relying solely on the binding of the PhaR repressor protein downstream of the σ^{70} factor that recruits the RNA polymerase. Successful repression thus impedes the progress of the RNA polymerase after it binds to the target promoter rather than directly obstructing the DNA sequence to which it initially binds. As seen in **Figure 4A**, preliminary tests with this setup showed promising results, giving well above 10-fold change in luminescence from the repressed state (*phaR* expression induced) to the non-repressed state (no *phaR* expression induced) for weak upstream promoters, although this effect lessened considerably as the constitutive promoter strength was increased.

To improve upon these results and attempt to find a promoter with both higher transcription rates and lower background, I tried mapping the upstream promoter series tested directly to the embedded promoter sequence from the native binding pocket. Ideally, this would lower background transcription rates by forcing direct competition for DNA binding by PhaR and σ^{70}, but at the risk of interfering with PhaR binding kinetics to residues involved both in σ^{70} recruitment and PhaR binding. Modified embedded promoter sequences were tested as indicated in **Table 1**, and overall it seems clear that for strong promoter activities, embedded promoters provide a far superior dynamic range than separate promoters placed upstream of a native binding sequence. The modified PhaR binding sequences used in plasmids pTR4pB and pTR4pD were the most promising, as they provide both a good dynamic range (about 50-fold increase over the repressed state upon expression of a PHB synthetic pathway) and two of the highest signals for the non-repressed state. Despite having a lower dynamic range than pTR4 or pTR4pA, this high signal

in the presence of PHB granules is a far more desirable phenotype for applications like PACE that require significant protein production from a regulated PhaR circuit.

Table 1: AP variants used to test PhaR Repression and De-Repression.
The bases highlighted in blue and red correspond to putative σ^{70} -35 and -10 consensus sequences, respectively, while bases highlighted in green and underlined represent changes made to the wild-type promoter sequence. All other components of each construct are identical.

Construct	Upstream Promoter	PhaR binding sequence (5' → 3')
pTR4	none	GGGCGTAAAATTTTTCTGCACCGCAGCAAGAAAACC**TTGCAA**TGCTGCGGTGCAGAAAG**TATATG**TT
pTR7-1	pro1*	GGGCGTAAAATTTTTCTGCACCGCAGCAAGAAAACC**TTGCAA**TGCTGCGGTGCAGAAAG**TATATG**TT
pTR7-3	pro3*	GGGCGTAAAATTTTTCTGCACCGCAGCAAGAAAACC**TTGCAA**TGCTGCGGTGCAGAAAG**TATATG**TT
pTR7a	proA*	GGGCGTAAAATTTTTCTGCACCGCAGCAAGAAAACC**TTGCAA**TGCTGCGGTGCAGAAAG**TATATG**TT
pTR7b	proB*	GGGCGTAAAATTTTTCTGCACCGCAGCAAGAAAACC**TTGCAA**TGCTGCGGTGCAGAAAG**TATATG**TT
pTR7d	proD*	GGGCGTAAAATTTTTCTGCACCGCAGCAAGAAAACC**TTGCAA**TGCTGCGGTGCAGAAAG**TATATG**TT
pTR4p	none	GGGCGTAAAATTTTTCTGCACCGCAGCAAGAAAACC**TTGCAA**TGCTGCGGTGCAGAAAG**TATAAT**TT
pTR4p1*	none	GGGCGTAAAATTTTTCTGCACCGCAGCAAGAAAACC**TTTACG**TGCTGCGGTGCAGAAAG**GTATCT**TT
pTR4p3*	none	GGGCGTAAAATTTTTCTGCACCGCAGCAAGAAAACC**TTTACG**TGCTGCGGTGCAGAAAG**GAGGAT**TT
pTR4pA*	none	GGGCGTAAAATTTTTCTGCACCGCAGCAAGAAAACC**TTTACG**TGCTGCGGTGCAGAAAG**TAGGCT**TT
pTR4pB*	none	GGGCGTAAAATTTTTCTGCACCGCAGCAAGAAAACC**TTTACG**TGCTGCGGTGCAGAAAG**TAATAT**TT
pTR4pC*	none	GGGCGTAAAATTTTTCTGCACCGCAGCAAGAAAACC**TTTACG**TGCTGCGGTGCAGAAAG**TATGAT**TT
pTR4pD*	none	GGGCGTAAAATTTTTCTGCACCGCAGCAAGAAAACC**TTTACG**TGCTGCGGTGCAGAAAG**TATAAT**TT

*Indicates previously characterized insulated promoters or their -35/-10 binding sequences with the following relative strengths: 1<3<A<B<C<D[59].

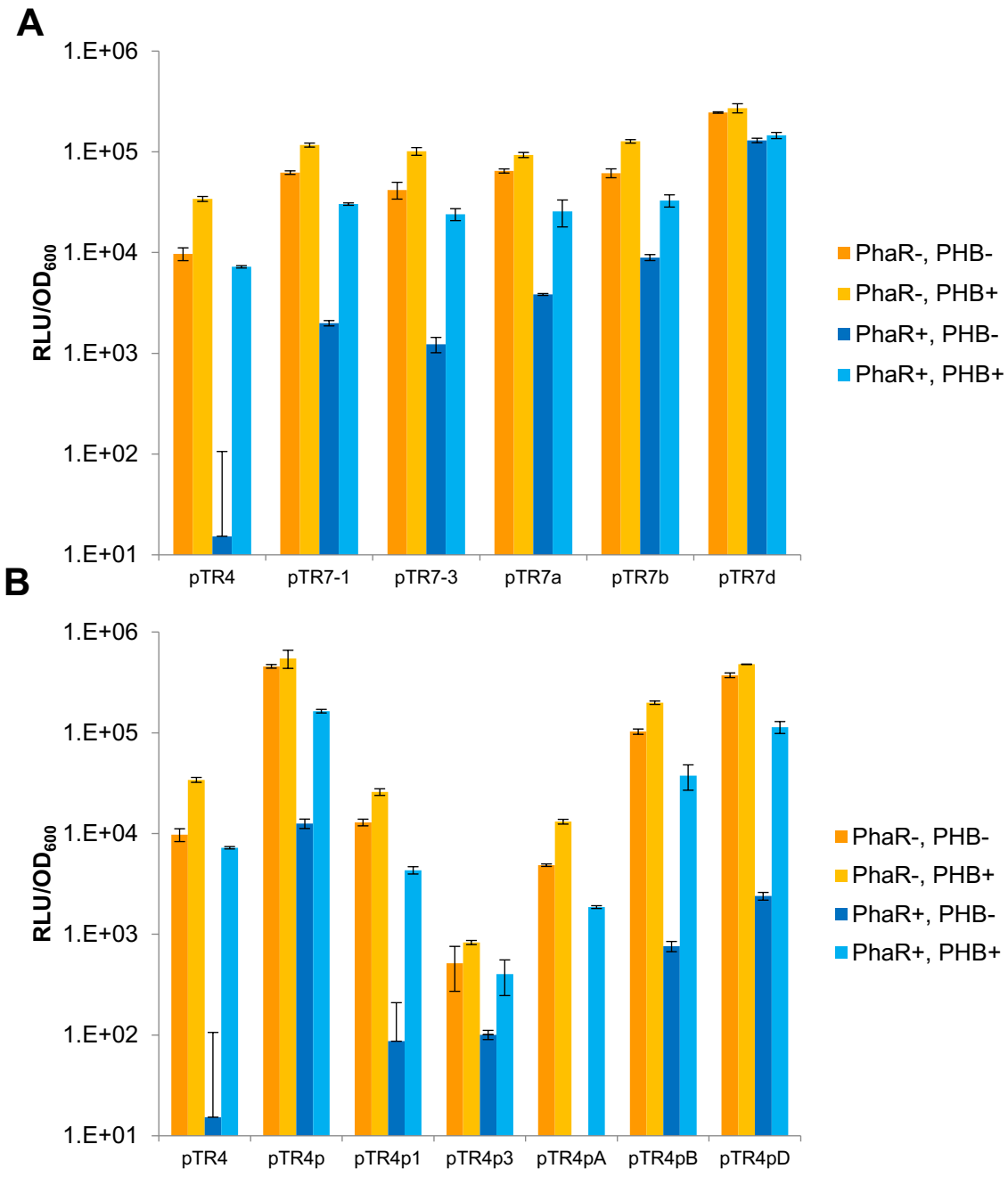


Figure 4: Minimal modification of the wild-type PhaR DNA binding sequence greatly improves the dynamic range of PhaR-regulated gene transcription in E. coli.
Luciferase Reporter Data for promoters placed upstream of the PhaR-binding sequence (**A**) and altered embedded promoters within the binding sequence (**B**). PhaR- indicates no induction of the *phaR* gene; PhaR+ indicates induction of the *phaR* gene by addition of 10ng/uL anhydrotetracycline (ATc) to the growth medium; PHB- indicates no induction of PHB synthesis; PHB+ indicates induction of PHB synthesis by addition of 0.2% (w/v) (L)-arabinose to the growth medium(~13.3mM). Luminescence readings were taken during mid-to-late log phase. Bars show averages and error bars show standard deviations across three technical replicates. RLU, Relative Luminescence Units. OD_{600}, Optical Density at 600nm.

Section 1.3: Linking PHB synthesis to phage propagation

With a set of reporter constructs in hand, I next needed to successfully link phage propagation to the synthesis of PHB. M13 Selection Phage (SP) propagation for systems like PACE relies on many copies of gene III being transcribed and translated into pIII, which in turn demands both a large dynamic range and fine-tuned background expression. If background expression is too low, a 100-fold increase in gene expression will provide barely enough gene III transcription for viable phage particles to form after infection, eventually leading to washout during the constant dilution rate of evolving populations in PACE. Conversely, if background expression is too high (such as with pTR7b or pTR7d constructs as shown in **Figure 4**), gene III transcription will begin prematurely before phage particles can infect host cells, which has been shown to block infection and lead to washout[7,8]. In order to test my series of reporter constructs for viability in PACE, I placed gene III upstream of the luciferase reporter in the pTR4 series of constructs and tested for infectibility using the kanamycin-resistance assay described the original PACE publication[7] (**Figure 5**). From these data, I noted that all strains are highly infectible so long as PhaR is expressed, with no notable differences in infectibility based on the embedded promoter strength used.

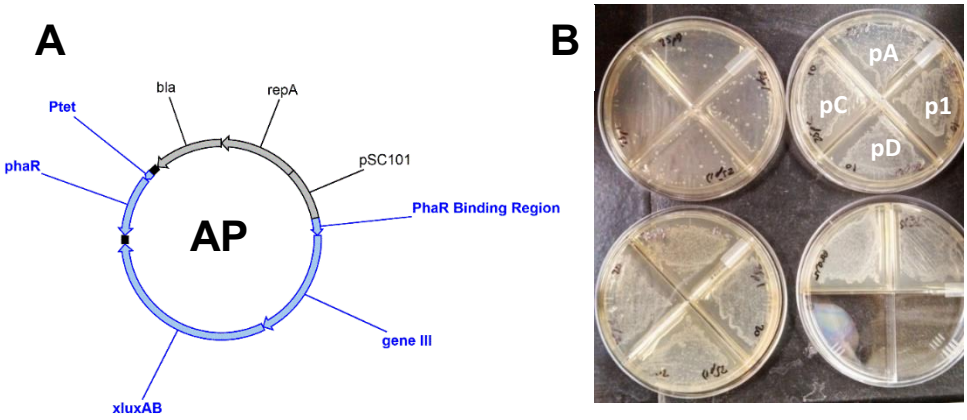


Figure 5: Phage Infectibility Assays for Engineered AP Architectures.
A. The pTR25pX series, equivalent to the pTR4pX series, but with gene III placed upstream of xluxAB as pictured. bla, antibiotic resistance marker; pSC101/repA, origin of replication. **B.** Phage infection of S1030 cells carrying one of the pTR25pX series of APs, with PhaR-binding sequences analogous to the corresponding pTR4pX plasmids listed in *Table 1*. Bottom right plate, positive control strains (no AP). Within each of the other plates, pTR25p1, pTR25pA, pTR25pC, and pTR25pD are positioned as labeled in the top-right plate. Top left plate, 0ng/mL Anhydrotetracycline (ATc), Top right plate, 10ng/mL ATc, bottom left plate 20ng/mL ATc,

After confirming all AP candidates as infectible, I next wanted to see if I could make plaque formation dependent on PHB synthesis. Activity-dependent plaquing is a common, baseline level of activity for PACE compatibility, as generally any SP that can form visible plaques on a lawn of host containing the cognate AP is more than capable of surviving in a continuous flow system. As discussed in the previous paragraph, proper PhaR regulation of gene III should permit infection by these selection phage but prevent their propagation unless a full complement of PHB synthetic genes is achieved (e.g. plated host cells expressing *phaB* and *phaC*, but lacking *pha*A, should show clear plaques when infected with SP containing *phaA*). I first cloned phage containing the full *phaABCP* pathway from the PHB Production Plasmid as picture in **Figure 3**, but with this full cassette, no visible plaques formed on any AP-containing host tested. This was somewhat expected given that larger phage like this ~10kb SP require more DNA replication and packaging than a standard M13 phage particle**.** I next moved to phage containing either just *phaA* or *phaCP*, which have significantly smaller genome sizes. Whichever genes were not placed on the phage were instead placed on a complementary plasmid alongside the AP in the host strain. Both phage were able to form plaques on lawns of host expressing their pathway complement, but not on host expressing the wrong complementary set of pathway genes, providing the basis for a selection for individual pathway enzyme activities based on expression of an appropriate complement in the host.

After observing a plaquing phenotype for these phage, I wanted to move into a continuous flow system to validate their potential for PACE. From my initial plaque assays, I noticed that plaque sizes tended to be larger when *phaP* was encoded and expressed either on the phage or in the host, which corroborated reporter tests I had run previously (**Figure 6**). To give myself the best possible chance at demonstrating propagation under more stringent continuous flow conditions, I added the *phaP* gene to the AP, placing it under control of the *E. coli* phage-shock promoter (PSP)[61]. This circuit thus supplies PhaP only after phage containing PHB pathway genes infect the host,

helping prevent any unpredictable effects due to granule synthesis in hosts that have been constitutively overexpressing *phaP* prior to infection. Consistent with previous results, this PSP-*phaP* AP provided a similar plaquing phenotype to strains constitutively expressing *phaP* and a stronger plaquing phenotype compared to strains without any *phaP* expression.

With this new AP, I focused my initial PACE experiments on a promising alternative to the *phaA* SP previously shown to form activity-dependent plaques. NphT7 is a decarboxylative acetoacetyl-CoA synthase that requires both acetyl-CoA and malonyl-CoA substrates[62]. Unlike thiolases such as PhaA, NphT7 will not hydrolyze its product in a reverse reaction (**Figure 7A**). This irreversibility makes it an attractive option for biosynthetic pathways that include acetoacetyl-CoA as an intermediate (*n*-butanol, isoprenoids, etc.)[63–65]. I initially compared the PHB producing capacity of NphT7 to PhaA through luciferase assays and found that, all things equal, pathways using NphT7 gave consistently lower yields than those using PhaA. This was corroborated in chemical analysis of PHB production, which showed higher calculated PHB content for the PhaA strain than the NphT7 strain (**Figure 7B**).

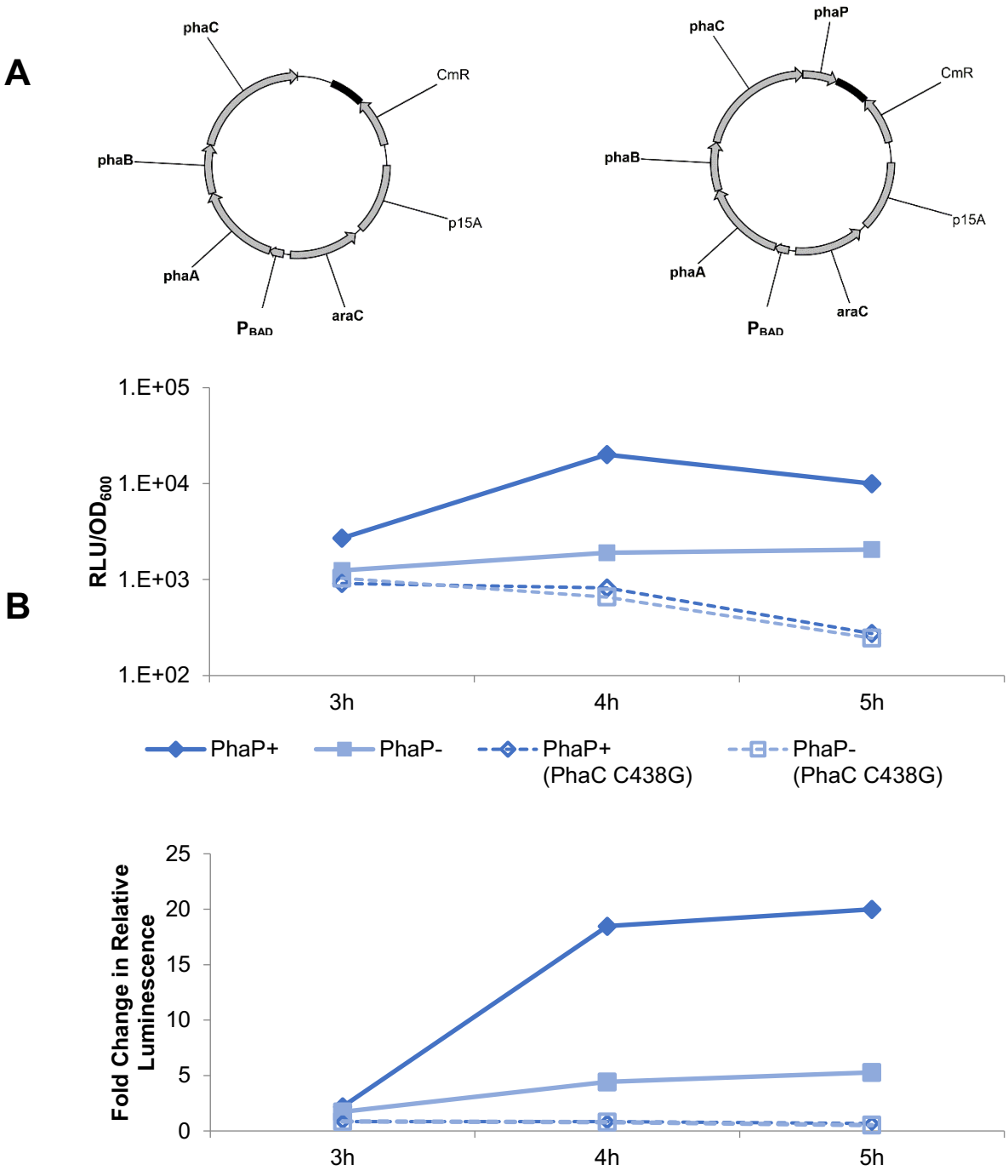


Figure 6: Illustration of the effect of phasin production on PhaR response.
A. Two plasmids were prepared, one containing a transcript with *phaABC*, the other with a transcript containing *phaABCP*. CmR, antibiotic resistance marker; p15A, origin of replication. **B.** Using a reporter construct based on the pTR4 series, four strains were compared for their increase in luminescence over time. The first two contained the plasmids pictured in A, the second two contained the same plasmids with a deactivating point mutation (C438G) in *phaC*. RLU, Relative Luminescence Units. OD_{600}, Optical Density at 600nm.

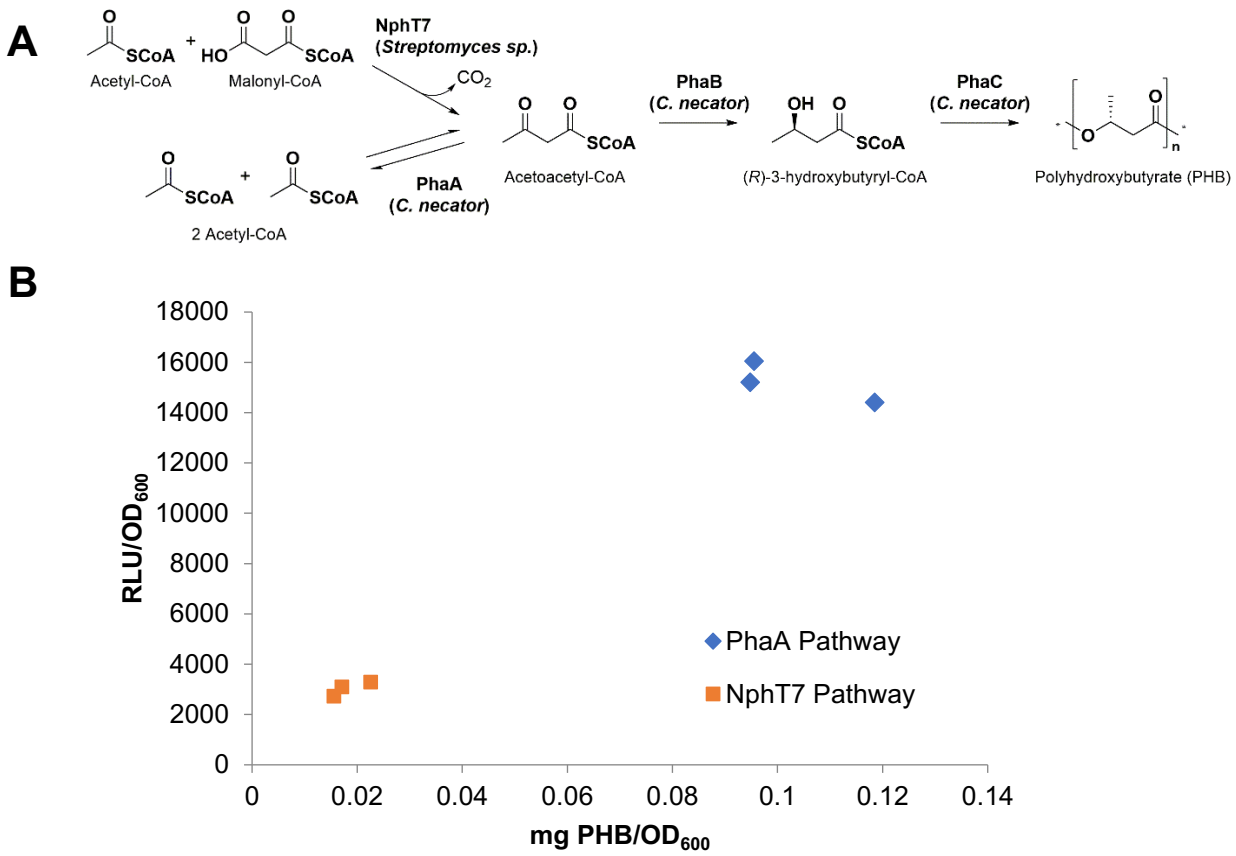


Figure 7: Comparison of PhaA and NphT7 in PHB production pathways.
A. NphT7 facilitates the irreversible production of acetoacetyl-CoA through decarboxylative condensation, whereas thiolases like PhaA condense 2 acetyl-CoA molecules through a process that is highly reversible[62]. **B.** Equal volumes of PHB-producing cultures were measured for luciferase activity. A second subset of the same cultures was processed chemically as reported previously[45,66] to determine total PHB content. RLU, Relative Luminescence Units. OD_{600}, Optical Density at 600nm.

I next tested the ability of this gene to persist in continuous culture as required by PACE. I first placed the *nphT7* gene on the selection phage to create SP7. I then flowed a chemostat containing host cells expressing *phaB* and *phaC* genes from a complementary plasmid into two equal-volume lagoons at a flow rate of 1 lagoon volume per hour. I infected the first lagoon with 10^8 plaque-forming units (pfu) of SP7 and the other with a mixture of 10^8 pfu of SP7 and 10^6 pfu of a selection phage containing the *phaA* gene (SP5). The purpose of this second lagoon was to see if, as expected, the higher activity of PhaA would result in selective enrichment of SP5 over time. The phage titers of the two lagoons were monitored for a 72h period and phage identity (SP7 vs. SP5) was checked via PCR (**Figure 8**). The SP5 *phaA* fragment enriched over the SP7 *nphT7* fragment as

expected, suggesting that the selection efficiently enriches for more active enzyme activities in a continuous flow system and is not inhibited by an overabundance of less active or inactive phage.

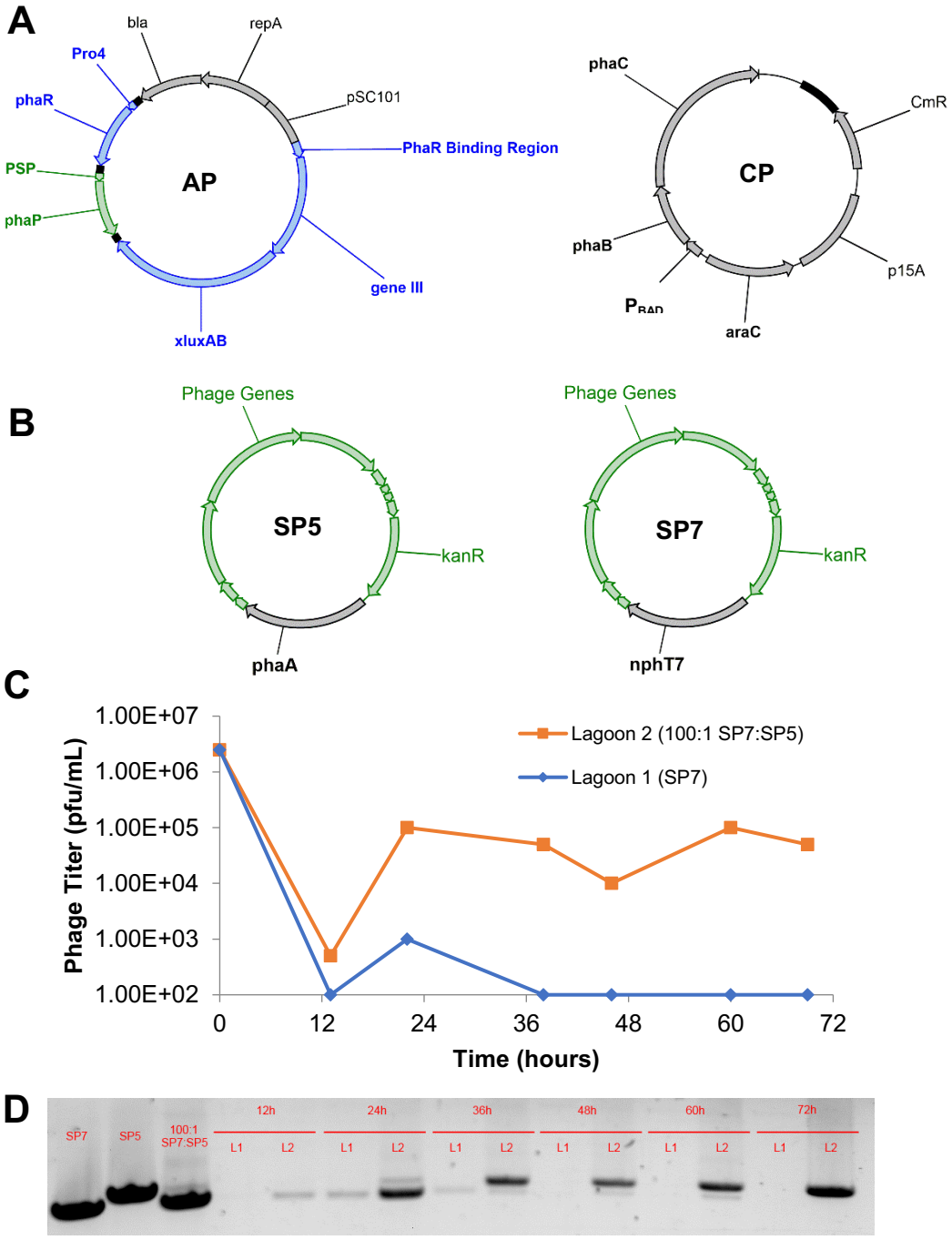


Figure 8: Phage encoding the *pha*A gene from *C. necator* can propagate in continuous culture, but phage encoding the *nph*T7 gene from *Streptomyces sp.* cannot.
A. plasmids cloned into the host cells for this experiment. Arabinose was added to the chemostat to ensure induction of the *phaBC* cassette from the complementary plasmid (CP). bla and CmR, antibiotic resistance markers; pSC101/repA and p15A, origins of replication. **B.** Selection Phage used in this experiment. **C.** Activity-independent phage titers observed in each lagoon. The graph shows the total plaque-forming units (pfu) per mL of the lagoons over time as determined by plaque assays in a strain containing PSP-gene III. **D.** PCR amplification products for lagoon timepoints using primers binding to the selection phage immediately outside the region encoding either *nph*T7 for SP7 or the slightly larger *pha*A fragment for SP5**.**

Section 1.4: Phage-Assisted Continuous Evolution of NphT7

Given the inability of SP7 to propagate in continuous flow culture, I added a mutagenesis plasmid (MP4)[67] to the same host strain and ran a full PACE experiment. A few changes were made to host strain to simplify the target phenotype – most notably, *phaP* was removed from the AP to prevent evolution of a production phenotype dependent on phasins for granule formation. Previous data suggested that expression of *phaP* would increase the sensitivity of the AP response (**Figure 6**), but it was unclear if this correlated with higher PHB production levels or merely biased the signal by modulating the surface-area-to-volume ratio of granules forming in the cell such that the PhaR de-repression events were artificially increased with time. Given the diverse and unclear properties of phasins and their interactions with other granule-associated proteins[35], I opted to omit them from all future selections, which I hypothesized would better ensure a direct relationship between PHB synthesis and PhaR de-repression.

The dilution rate for the lagoon was kept at a rate of 1 lagoon volume per hour as in the previous experiment, but this time arabinose was added to induce mutagenesis from the MP. I kept the PACE running for 280 hours in total, but there were many technical difficulties. First, the initial infection pool washed out within 48 hours, requiring re-infection with the clonal starting population at 72 hours. Additionally, on multiple occasions (as noted in the graph for **Figure 9**), the flow rate was effectively reduced to zero – typically this was due to problems with the chemostat. The two modes of chemostat failure observed were washout of host cells (i.e. reduction of chemostat optical density to a level below log phase) and improper pumping of fresh media into the chemostat, causing the downstream flow to pump only air and allowing the chemostat and lagoon cell populations to reach saturation. Sequencing of several timepoints both immediately before and after these technical issues showed many silent mutations accumulating within the evolving gene of interest, but no consensus of coding mutations. Sequencing of the final population at 280 hours, however, showed a T106S mutation fixed in a majority of individual phage sequenced (**Figure 8C**).

To see if I could push this evolution further – ideally with fewer technical problems – I repeated the PACE experiment with two lagoons; the first was infected with the same starting material from the initial PACE and the second with the final timepoint of the initial PACE. Titers were adjusted such that both lagoons were infected with similar titers of phage. The lagoon containing the evolved phage increased considerably in titer and only struggled when flow rates were increased at 84 hours (**Figure 8B**). As seen previously, the lagoon infected with clonal starting SP7 washed out very quickly, and never recovered in titer even with a chemostat error and subsequent stall in flow rate at 72 hours. This inconsistency with the initial PACE likely stems from the lower starting titer of phage in the lagoon, but most importantly highlights the improvements in the evolved population, which was seeded with a similar initial titer but successfully escaped dilution for a total of 360 hours. Sequencing of this lagoon showed immediate convergence on a T106A mutation – or possibly S106A mutation given the majority sequence of the initial population used to infect this lagoon, though the propensity of the MP4 mutagenesis plasmid used is more toward the former (an adenine to guanine transition) than the latter (a thymine to guanine, which occurs much less frequently)[67]. While evolved phage containing T106 mutations didn't show any major changes in plaque size over those encoding wild-type NphT7, plaques formed by phage containing T106A had distinct white halos around their zones of clearance, possibly reflecting an increased accumulation of PHB in the lawn infected with low titers of selection phage at their diffusion front.

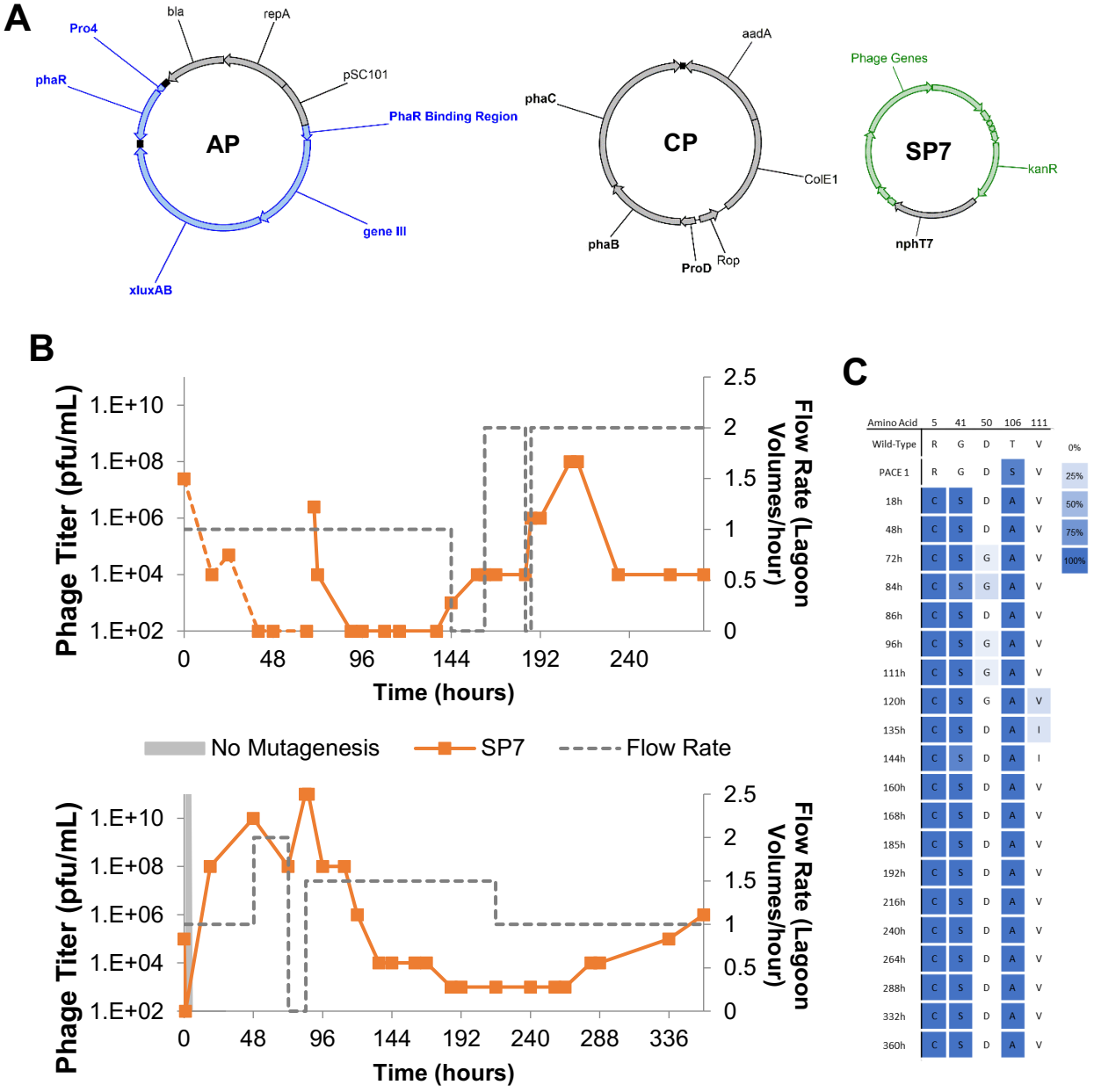


Figure 9: Phage Assisted Continuous Evolution of NphT7.
A. S1030 *E. coli* were transformed with this AP and CP combination and infected with SP7, which carries the target *nphT7* gene. Host cells also contained MP4 to induce mutagenesis (not shown). bla and aadA, antibiotic resistance markers; pSC101/repA and ColE1/rop, origins of replication. **B.** Titers for the first (top) and second (bottom) PACE experiment during which SP7 was evolved. The second experiment was seeded with phage isolated from the first. The gray region at the start of the second PACE indicates that no arabinose was added initially, with mutagenesis only induced several hours after infection. pfu, plaque forming units. **C.** Sequencing data for the *nphT7* gene encoded on SP7 showing the final sequence of the first PACE experiment and the change in sequence over time for the second. Mutations shown are coding mutations that occurred in at least 2 individuals sequenced. Silent mutations not shown. Shading represents the percentage of individuals sequenced for a given time point containing the indicated mutation.

To validate the results of this PACE of NphT7, I PCR amplified the genes encoding the evolved NphT7 enzymes from the first and second PACE experiments (the PACE 1 and 360h rows in **Figure 9C**), which I will refer to as NphT7m1 and NphT7m2, respectively. I placed these mutant genes on a plasmid upstream of the *pha*B and *pha*C genes and downstream of an arabinose-inducible promoter. I then tested the constructs for apparent PHB synthesis activity using luciferase reporter assays, showing that both NphT7m1 and NphT7m2 indeed display a higher average luciferase signal than the wild-type gene. To help reduce the biological noise of these results, which showed significant variability across the 12 biological replicates assayed, I re-grew the 3 highest performing individuals encoding NphT7 (wt) and NphT7m2 and performed a second luciferase assay coupled to chemical analysis of PHB content via acid degradation and HPLC analysis[45,66]. The titer obtained after 6 hours of growth at the highest level of expression for NphT7 was 5.5mg/L of PHB, whereas the corresponding NphT7m2 titer was 34.5mg/L. **Figure 10** summarizes the results obtained from these assays.

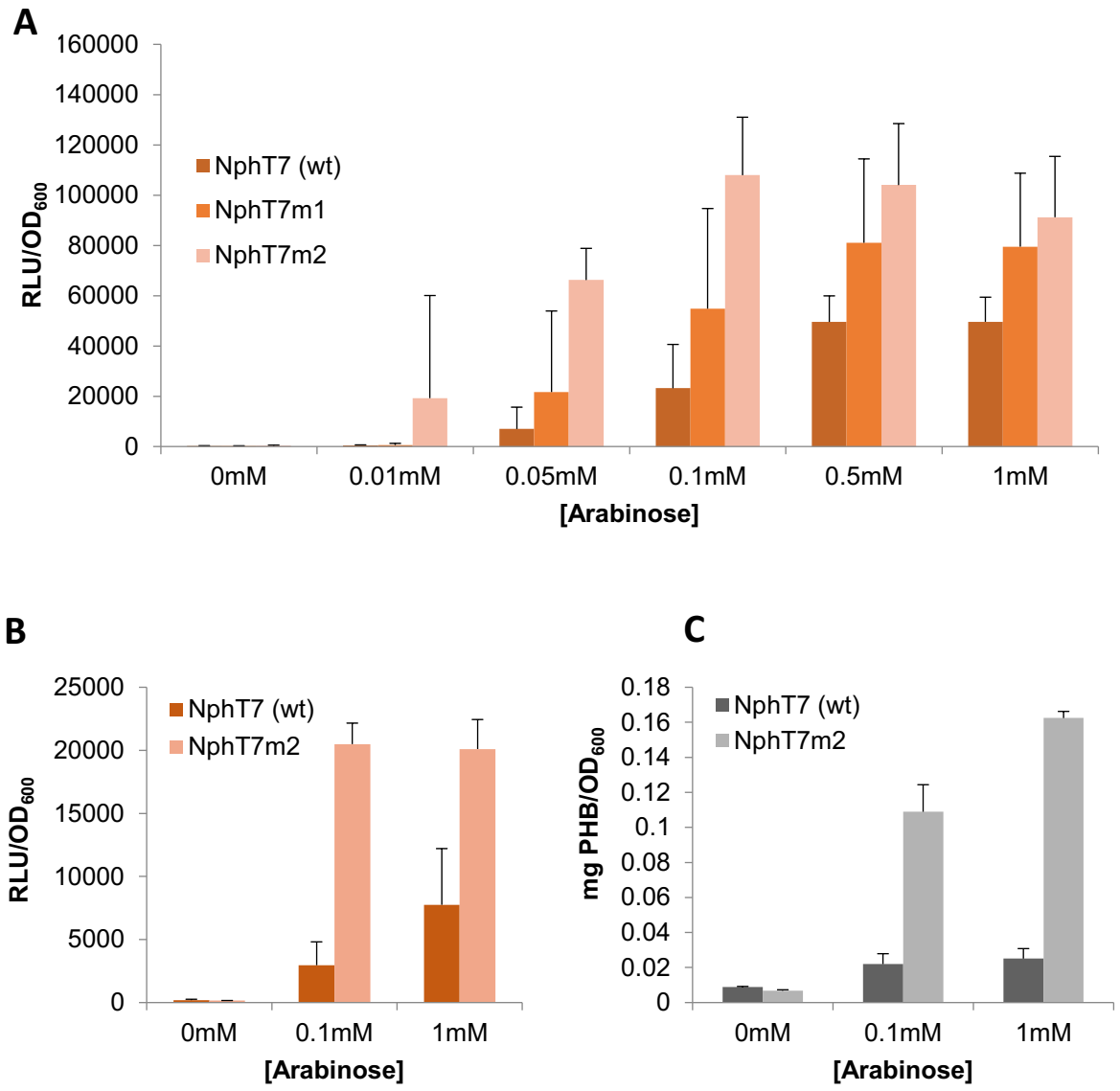
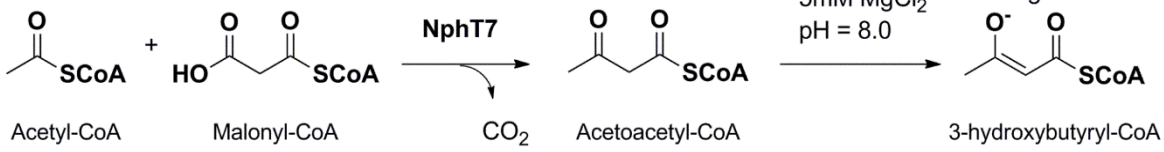


Figure 10: Validation of NphT7 variants evolved in PACE.
A. Evolved NphT7 variants show improved luciferase signals compared to wild-type. RLU, Relative Luminescence Units, normalized to the optical density of the culture at 600nm. Data shown was taken 8h after inoculation of strains into growth medium. Bars show average signal and error bars show standard deviation from 12 biological replicates. **B.** Secondary validation of 3 of the high-performing biological replicates from part A for NphT7 (wt) and NphT7m2. Data shown was taken 6h after inoculation of strains into growth medium. RLU, Relative Luminescence Units. OD_{600}, Optical Density at 600nm. **C.** HPLC validation of results from part B. Error bars in all graphs show the standard deviation between the designated number of biological replicates.

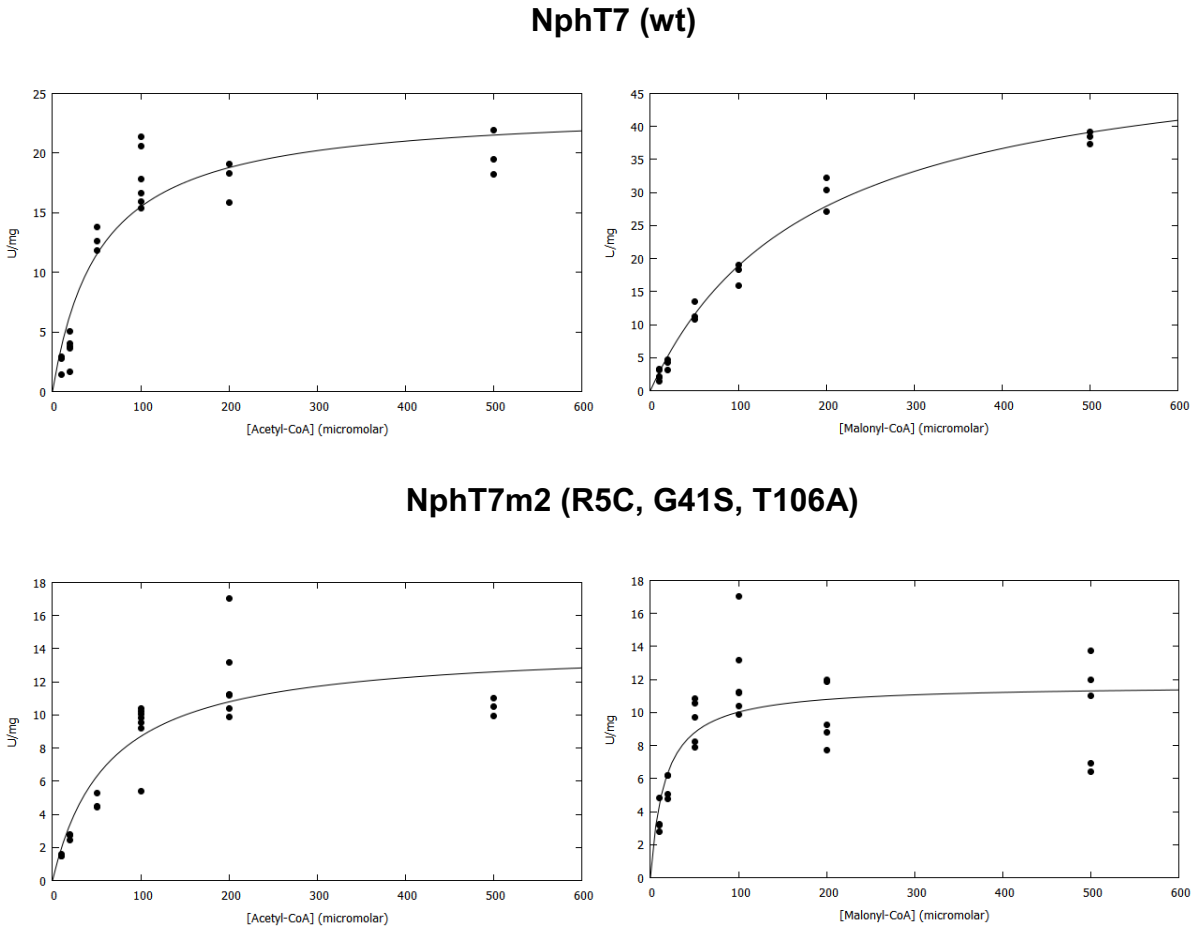
From these data, it seems clear that NphT7m2 has a higher *in vivo* activity than the wild-type gene, but this could be due to a number of factors relating to expression, enzyme stability and solubility, etc. To further identify the nature of this improved activity, I purified 10x His-tagged versions of each enzyme and measured their *in vitro* activity by monitoring synthesis of acetoacetyl-CoA from malonyl-CoA and acetyl-CoA as previously reported[62]. I measured initial velocities for both enzymes as dependent on their two substrates and fit the data to the Michaelis-Menten model to calculate the V_{max} and K_m values for each. Importantly, my data for wild-type NphT7 is not in agreement with previously reported values, which estimate a K_m closer to that observed for the mutant enzyme[62]. Studies using the same steady-state kinetics assay I used report a K_m of about 28μM for the wild-type enzyme, while other studies have corroborated this result with a separate PhaB-coupled kinetics assay, making it unlikely that my values are accurate[68]. It is unclear why there is such a large discrepancy between my data and the published data, but it was apparent from spectrophotometric reads that the linear fit for NphT7m2 with varying malonyl-CoA concentrations tended to be better over a longer period. Repeat experiments should use less enzyme and more consistent readings to minimize the noise of this assay. Further characterization is necessary before drawing any conclusions, but from the values I obtained, NphT7m2 has steady-state kinetic parameters slightly lower than those reported for the wild-type enzyme for both malonyl-CoA binding and acetyl-CoA binding. Assuming these values accurately reflect the properties of this mutant NphT7, improvements in *in vivo* PHB incorporation based on NphT7m2 expression could very likely be related to factors outside of kinetic improvements to the enzyme, such as expression rate, stability/solubility, or similar. These improvements, while important to overall *in vivo* PHB yield, would not be reflected *in vitro*.

Figure 11: Kinetics Assays of purified NphT7 Variants
A. SDS-PAGE gel showing the purified 10x-His-tagged NphT7 enzymes. Crude: crude lysate before centrifugation; insol: insoluble fraction of the lysate after centrifugation; sol: soluble fraction of the lysate post-centrifugation; ft: flow-through after application to the Ni-NTA column; eluant: fraction collected after application of concentrated imidazole to the Ni-NTA column.

B



C



D

	Acetyl-CoA		Malonyl-CoA	
	V_{max} (U/mg)	K_M (μM)	V_{max} (U/mg)	K_M (μM)
NphT7(wt)	23.82 ± 1.79	53.69 ± 12.37	53.39 ± 2.54	182.56 ± 19.68
NphT7m2	14.19 ± 1.39	63.02 ± 19.35	11.68 ± 0.81	16.55 ± 5.61

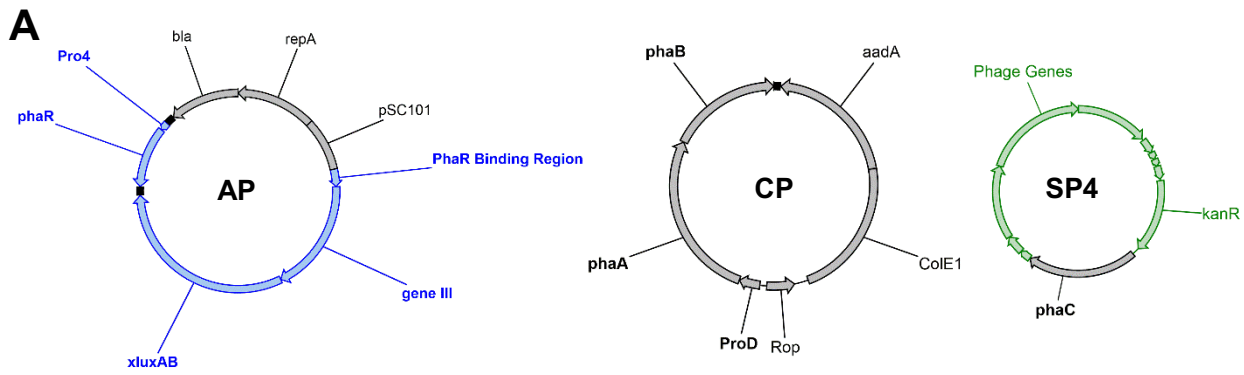
Figure 11 (continued): B. The reaction monitored was formation of the acetoacetyl-CoA enolate form via stabilization by magnesium present in the buffer. Absorbance at 280 correlates with concentration of this species. **C.** Plot of enzyme activities for NphT7 (top) and NphT7m2 (bottom) as a function of either acetyl-CoA concentration or malonyl-CoA concentration. **D.** Summary of calculated parameters for each enzyme from fitting the data to the Michaelis-Menten equation. Error shows the standard deviation for the full data-set for fit.

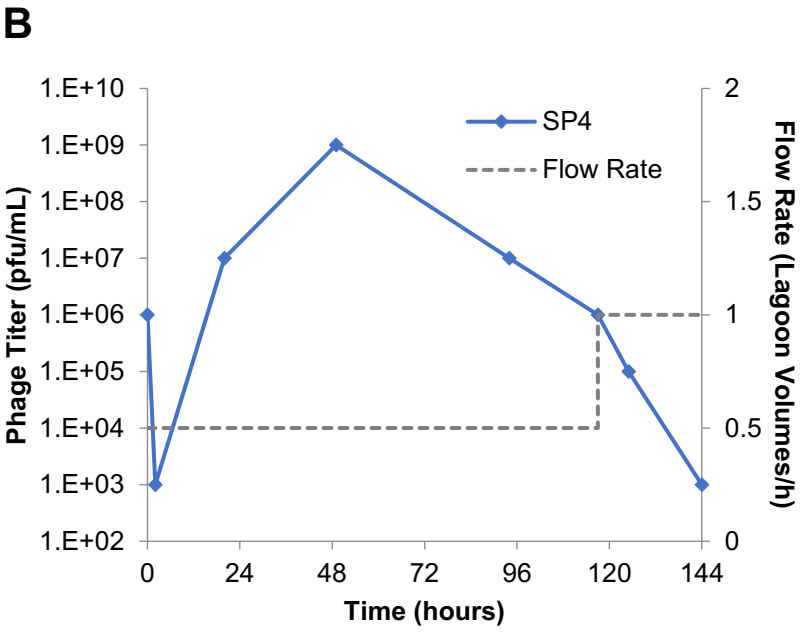
Section 1.5: PACE of wild-type *phaC* activity

While there are many PHA synthases in nature that have been characterized and evolved, I chose to focus on the *C. necator* PhaC for two reasons: first, there is a good body of literature annotating mutations that can improve or limit PHB production[69–72], and second, at the time I started this project there was a distinct lack of complete structural information for this enzyme, making rational design a difficult approach for improving its catalytic efficiency. Partial structures have since been reported, but engineering of this active site remains uncertain territory, with relatively few mutations known to significantly improve catalytic efficiency or *in vivo* PHB production[73–75]. My approach for evolving a selection phage containing the *phaC* gene was essentially the same as my approach for evolving *nphT7*, but with the host encoding a plasmid with the two upstream components to complement the phage PHA synthase, namely *phaA* and *phaB*. Unlike the selection phage I had validated previously, I kept my strategy in place from the NphT7 PACE experiments and remade a version of these selection phage without *phaP*. Additionally, to optimize phasin-free *phaC*-SP propagation, I screened several levels of pathway complement expression for their ability to propagate phage either through plaque assays or overnight infections. From the results in **Table 2**, I decided to go with the highest level of complement expression and attempt PACE (**Figure 12**). An initial setup failed to propagate phage, possibly from too low of an infection titer or similar miscalculation, but a repeat attempt yielded phage propagation at 0.5 lagoon volumes per hour for several days (**Figure 12B**). Sequencing shows very low convergence of coding mutations in PhaC itself, but sequencing of the *kanR* gene upstream of the region of interest shows strong convergence on a stop codon mutation toward the start of the evolution (**Figure 12C**). This is notable given that the *kanR* and *phaC* genes on this SP were designed to transcribe from the same promoter and are in-frame, making this stop codon mutant translatable into a KanR-PhaC fusion protein.

Table 2: Optimization data for pathway complements to *phaC*-SP SP4. Overnight titer expansions were calculated by plaque assays using the positive control strain, which contains a PSP-gene III circuit as previously described. Pro1 < ProA < ProB < ProD[59].

Promoter Driving *phaAB* Expression	SP4 Plaques?	Overnight Titer Expansion (fold increase in pfu/mL)
Pro1	No	10
ProA	No	100
ProB	Yes, Very Faint	1000
ProD	Yes, Very Faint	1000
Positive Control	Yes	10000
No Complement	No	1





C

Gene	KanR		PhaC	
Amino Acid	2	272	152	584
Wild-Type	S	.	V	W
2h	Full Deletion of Both Genes			
20h	R	Q	V	W
49h	R	Q	V	W
94h	R	Q	V	W
117h	R	Q	V	W
125h	R	Q	D	C

Figure 12: Phage Assisted Continuous Evolution of PhaC
A. S1030 cells containing the Accessory Plasmid (AP) and Complementary Plasmid (CP) were infected with Selection Phage encoding *phaC* (SP4). Host cells also contained MP4 to induce mutagenesis (not shown). bla and aadA, antibiotic resistance markers; pSC101/repA and ColE1/rop, origins of replication. **B.** SP4 titers over time as measured via activity-independent plaque assays. pfu, plaque forming units. **C.** Sequencing data of evolving SP4. Dark blue shading shows 100% convergence, light blue shows 50% convergence.

To determine the full effects of this mutation, I subcloned both the *kanR* and *phaC* genes downstream of *phaA* and *phaB* onto a separate plasmid with the full transcript under control of the P_{BAD} promoter. I subsequently cloned a version of this same plasmid that omitted the *kanR* gene entirely or reverted the stop codon to confirm which of three possible scenarios led to improved phage propagation during PACE: if any of the silent mutations in the evolved *phaC* gene were responsible for the observed phenotype, the construct lacking *kanR* should perform best. If the presence of the *kanR* gene itself is enough to improve activity, there should be no difference between the fusion KanR-PhaC construct and the reversion construct. These plasmids were transformed into S1030 cells with the same AP used during PACE to complete and monitor activity of the resultant PHB production pathway. From the data shown in **Figure 13**, I saw that while the fusion construct indeed had the highest average apparent activity, the most notable change in was a very high background activity for the KanR-PhaC fusion construct. Given the large signal even without induction of the fusion protein from the P_{BAD} promoter, I hypothesized that PACE had evolved the stop codon to take advantage of an unannotated promoter embedded within the *kanR* gene itself. Regardless of the specific mechanism, however, there were two important conclusions from this attempted evolution – one being that *in vivo* selections like PACE will often find unexpected solutions and workarounds for improving phage propagation and the other being that multiple genes evolved on the same transcript in PACE have the potential to form fusion constructs if separated by a reasonable multiple of three nucleotides. This is particularly noteworthy for metabolic pathway evolution, since functional fusion constructs for PHB synthesis pathways have been reported previously[76] and may provide promising routes for future evolutions in PACE. While the activities of the KanR and PhaC proteins in this experiment are orthogonal, selection phage can theoretically be designed to allow an evolving population easy access to fusion constructs between enzymes.

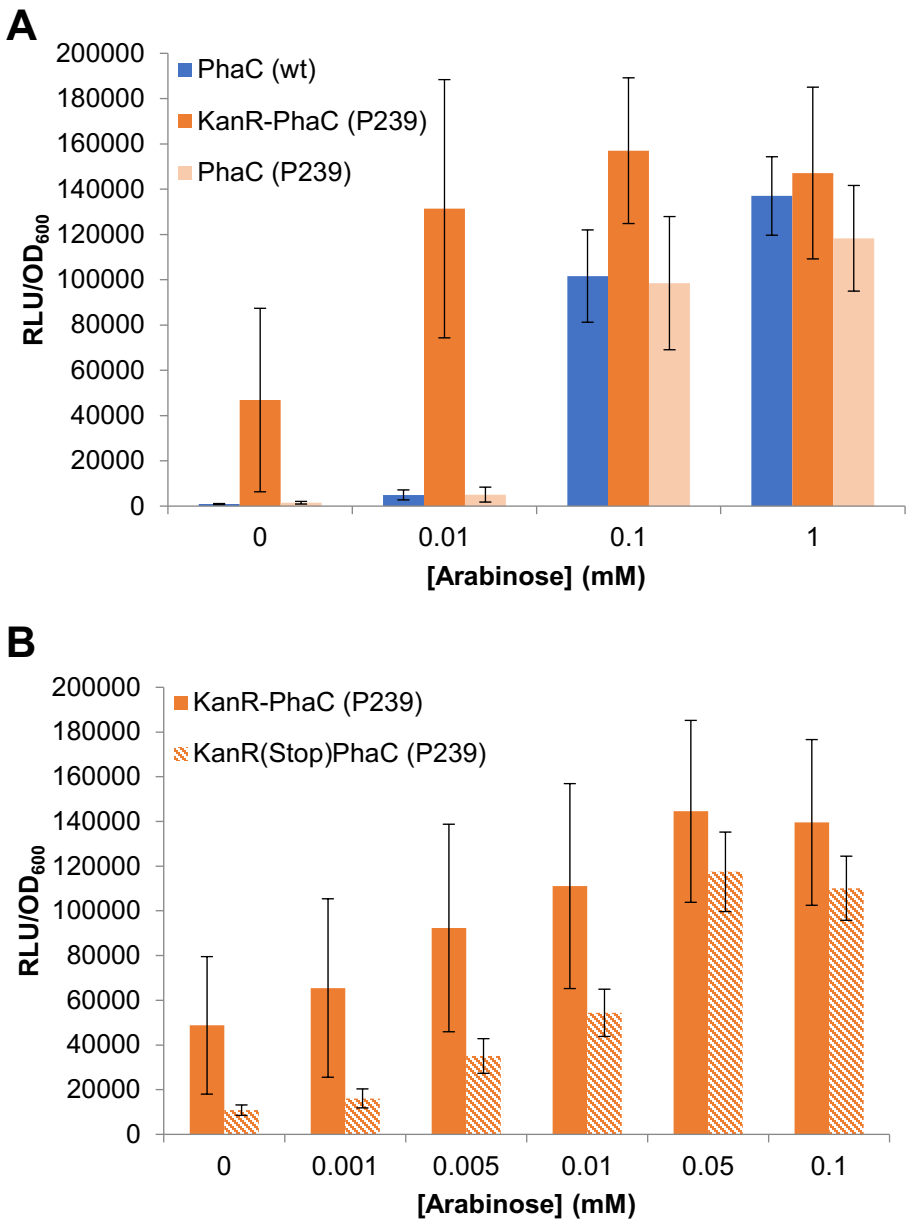


Figure 13: Luciferase Reporter Assays of Evolved KanR-PhaC Fusion.
A. A plasmid containing the full *phaABC* cassette was cloned with either *phaC* containing a single consensus silent mutation or with the *kanR-phaC* fusion gene from the PACE evolved SP4 and compared to the wild-type pathway. Bars are averages and error bars are standard deviations across 12 biological replicates. **B.** To better test the effects of the *kanR* gene, the stop codon was re-introduced between the *kanR* and *phaC* genes as shown with the shaded bars. Bars are averages and error bars are standard deviations across 8 biological replicates. RLU, Relative Luminescence Units. OD_{600}, Optical Density at 600nm. P239 represents the conserved silent mutation evolved in the individual PhaC mutant tested.

Section 1.6: Simultaneous Evolution of multiple PHA pathway enzymes

Synthetic biologists are constantly developing new tools to allow for pathway improvements via dynamic changes in gene expression, enzyme co-localization, or other processes based largely on known biological phenomena[4,56,77,78]. For PHB synthesis, for example, the order of genes on the operon itself can greatly impact observed PHB production in *E. coli*[79]. However, it is not always easy to predict the best optimization routes for all pathways. An unbiased selection for pathway optimization involving multiple genes could theoretically result in a range of mutations – such as the RBS/promoter changes and enzyme fusions observed in my initial PhaC evolutions – which in turn could help inform the best routes toward engineering synthetic pathways. Additionally, such a selection would not require the researcher to choose one enzyme under the assumption that it will have a larger impact on pathway kinetics than the others. In theory, evolving an entire biosynthetic pathway in PACE accesses these mutations that would otherwise be overlooked or underrepresented in traditional selection techniques. I thus wanted to test the limits of our phage-assisted techniques by evolving the entire PHB production pathway simultaneously on a single SP transcript.

As mentioned previously, cloning a large region of interest onto the SP can be difficult. In my attempts to isolate SP2, which contains the full *phaABCP* cassette from my original PHB production plasmid (**Figure 13A**), I observed frequent truncation of the pathway during activity-independent propagation in our S1030 pJC175e cloning strategy[8]. Presumably, this is because viable phage with significantly smaller genomes will replicate faster and more efficiently, sweeping the population before larger phage have a chance to reach higher titers. This is corroborated by plaque phenotypes, as plaque diameters for phage with larger genomes tend to be smaller than those for phage with smaller genomes. Typically, attempts to clone the SP containing a full *phaABC* cassette or equivalent using activity-independent propagation strains fail because of the selective propagation of phage expressing truncated pathway cassettes. Ultimately, I had to use activity-

dependent propagation based on the PhaR AP system to properly isolate SP2. Likely due to its size and lack of pathway optimization, SP2 is incapable of forming visible plaques on these host strains, much less persisting under the relatively stringent continuous flow environments of PACE. Infection of lagoons containing the PhaR-AP results in full washout within 40 hours at the lower limit of lagoon flow rates. However, that I could successfully clone and isolate a population of SP2 using my PhaR-AP system was enough to demonstrate that PHB-producing phage *could* be enriched in overnight in this host. To capitalize on this, I tried adding a mutagenesis plasmid to overnight propagations in a manner similar to PACE to see if several passages of overnight phage infection into fresh host cultures could evolve SP2, albeit on a slower time scale than PACE itself. The end goal was to isolate a mutated SP2 that would be fit enough to survive in a continuous flow environment.

In order to pre-mutagenize my phage population and speed up the process of finding enabling mutations, I switched from the MP system to the Drift Plasmid (DP) system – a functionally identical plasmid with regard to mutagenic capabilities, but with an additional copy of gene-III coupled to a hybrid P_{tet}/Phage Shock Promoter (PSP) regulatory circuit (**Figure 14A**)[8,67]. In this context, supplying anhydrotetracycline (ATc) to a host culture lowers the stringency of the selection by providing additional gene III for all phage regardless of their ability to de-repress PhaR from the AP. Selection stringency can thus be altered on the fly based on the concentration of ATc added to the evolving culture. This essentially allows populations to "drift" during the evolution, picking up mutations in the absence of aggressive selection. My initial fear in applying this to PACE was that even a few hours of such drift during a continuous flow experiment would be sufficient to seed the population with a large fraction of phage encoding truncated, non-functional pathway variants. However, by drifting in discreet, overnight passages, I could monitor this phenotype more carefully and tune drift between passages as needed. Given its analogy to PACE, I will refer to this

iterative, overnight phage passaging selection as Phage Assisted Non-Continuous Evolution, or PANCE (**Figure 14B**).

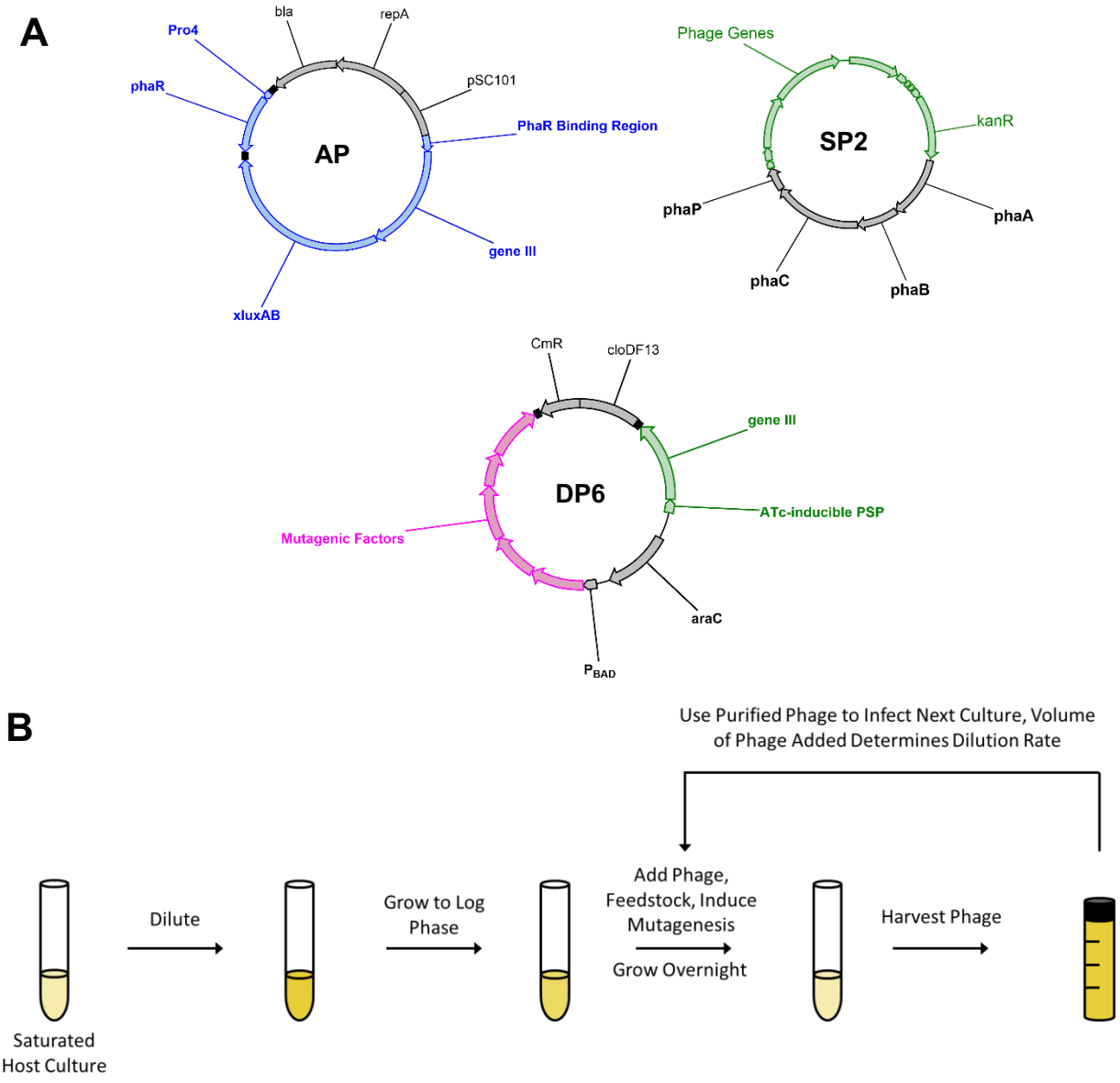


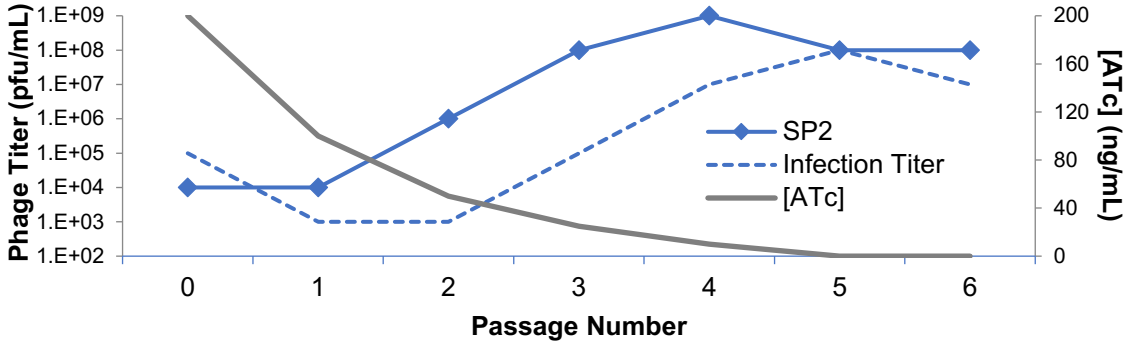
Figure 14: Phage Assisted Non-Continuous Evolution Protocol for evolving SP2.
A. S1030 cells containing the Accessory Plasmid (AP) and Drift Plasmid 6 (DP6)[67] were infected with Selection Phage encoding *phaABCP* (SP2). bla and CmR, antibiotic resistance markers; pSC101/repA and cloDF13, origins of replication. **B.** A general outline of the Phage Assisted Non-Continuous Evolution (PANCE) protocol. Fresh, saturated cultures are prepared daily by inoculating a colony from a plate into desired media overnight and back-diluting into fresh-media the next day. Log-phase cultures are infected with selection phage, any desired feedstocks, and arabinose and ATc to induce mutagenesis and drift to the desired level. Phage are isolated after overnight growth and isolated phage from one passage are used to infect subsequent passages.

Following the general PANCE protocol outlined in **Figure 14B**, I passaged SP2 at a range of dilution stringencies to prevent the chance of accidentally diluting out too much of the phage

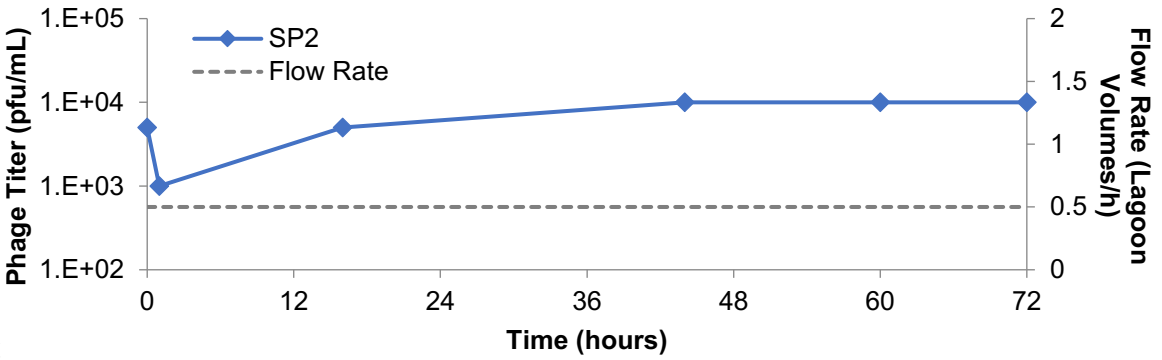
population between rounds. It was clear that dilution above 100-fold between rounds would very quickly "wash out" the phage populations, and the only dilution series to survive the 6 rounds tested with a robust population size was the one diluted only 10-fold between rounds. Crucially, this 10-fold dilution PANCE selection schedule, as shown in **Figure 15A,** eventually yielded a mixed population forming two clear activity-dependent plaque phenotypes—when previously SP2 had been incapable of forming activity-dependent plaques at all. One phenotype showed typical gene-III recombinant cheaters, a common fail-state for an evolution and essentially a useless result. However, another plaque phenotype was seen at lower titers within this population. I picked one of these small plaques, amplified and re-isolated it, and used it to infect a PACE experiment containing the same host strain used for PANCE.

Initially, I had attempted to add entire evolving PANCE population to PACE, but these washed out almost immediately. The single, activity-dependent plaque isolate, however, managed to survive in continuous flow as shown in **Figure 15B.** While the titers maintained were relatively low, they persisted above background and consistently showed retention of the full PHB pathway cassette. Sequencing of these phage over time failed to show convergence on any new mutations in the population, but the input phage themselves contained several mutations that were strongly fixed in the evolving population at the start of the experiment. I sub-cloned this PACE-stable pathway onto a separate plasmid and compared it to the equivalent wild-type pathway in both luciferase assays and PHB chemical analysis. For the chemical analysis, I grew the producing cultures overnight rather than testing them at the same timepoint as the luminescence reading to better test for significant changes in yield rather than rate of PHB production. This longer growth period likely explains the lower correlation of luciferase reporter results with chemical analysis, but both assays showed improved activity for the evolved pathway compared to the wild-type control, corroborating the PACE results.

A



B



C

Gene	*phaA*	*phaC*					*phaP*
Amino Acid	207	4	108	325	372	423	17
pTR39	R	G	R	V	A	L	E
pTR39m1	**C**	**D**	R	**I**	A	L	**Q**

D

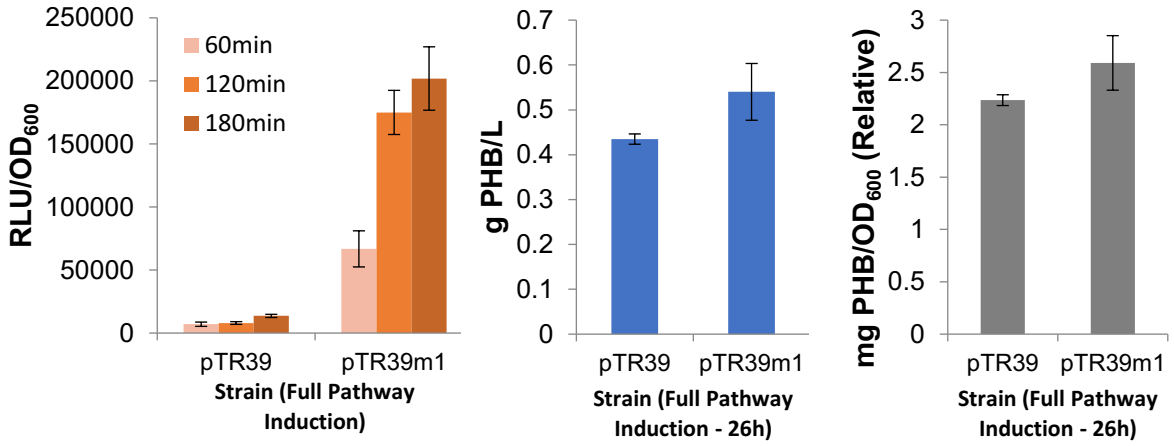


Figure 15: PANCE of SP2 enables continuous propagation of SP2.
A. Summary of PANCE conditions and titers. Infection titers are calculated titers based on the previous passage phage titer and the dilution rate for a given timepoint. pfu, plaque-forming units. **B.** SP2 titers in a PACE experiment seeded with a clonal isolate of the PANCE experiment in A. **C.** Summary of mutations in the isolated SP2 gene cassette subcloned onto an arabinose-inducible plasmid pTR39. Amino acids with the same identity as wild-type represent silent mutations at that position. **D.** Luciferase and PHB production data for wild-type and evolved pathways. RLU, Relative Luminescence Units. OD_{600} optical density at 600nm. Bars are averages and error bars are standard deviations across 3 biological replicates.

While improvements in PHB titer after 26 hours of growth were modest, I was heartened to see the evolution of a G4D mutation from my initial PANCE experiment. The G4 position has been previously identified from plate-based PHB screens and extensively characterized[69], helping validate my PhaR-based selection. The G4X mutations from this previous work show a greater difference in *in vivo* PHB production, but importantly the strains from this study were grown in supplemented LB and only induced for 14 hours rather than 26 hours, supporting the discrepancies I observed between my short-timescale luciferase data and my overnight PHB production data. Regardless, this result showed that by using PANCE, I could recapitulate previously published PhaC evolution results by selectively evolving a full pathway transcript without any synthetic library construction.

Section 1.7: PANCE of the NphT7-dependent PHB Pathway

After demonstrating the feasibility of full-pathway evolution, I wanted to reattempt the process using everything I had gleaned from these three separate attempts at PHB PACE. Based on my observations from the PhaC evolution, I decided to remove the *kanR* resistance gene (which initially was kept as a convenience for infectibility assays rather than any benefit during selection) from the next iteration of selection phage constructs given its clear potential to provide uninteresting solutions during PACE. Next, since the PHB production phenotype could be improved by evolving NphT7 on its own in the absence of *phaP* expression, I cloned a new SP containing the *nphT7 phaB phaC* cassette without a phasin gene. The lack of phasins, as discussed previously, helps to make any mutations within this cassette easier to interpret – but it also reduced the phage genome size slightly, theoretically improving its overall fitness. To further standardize this new PHB production cassette, I used the same RBS for each gene within the transcript and separated them by 15bp spacer sequences such that, should a stop codon mutate during evolution, it would encode a somewhat flexible linker region. Cloning this phage proved to be a greater challenge than SP2 cloning, requiring careful monitoring of the phage outgrowth after transformation into a

propagation strain containing the PSP-gene III circuit. While clonal phage was initially isolated and sequence verified, PCR analysis of the stored phage population showed severe degradation, with PCR amplicons indicative of loss of the pathway genes in the stock solution itself**.** Because of this, I decided to reapply the iterative PANCE approach to evolve this pathway rather than attempt PACE.

Despite issues with cloning and stock stability, SP20 was highly capable of propagating overnight in cultures of the standard PACE/PANCE growth medium I had been using for all selections up to this point, Davis Rich Media (DRM). Using DRM, I never needed to drift the evolving population to increase phage titers. I thus continued PANCE as shown in **Figure 16B**, with occasional monitoring of population titers by plaque assays and PCR amplifications of the SP20 region of interest to ensure the full production cassette was being maintained during the evolution. Outside of one adjustment to dilution stringency at passage 5 due to a relatively low observed titer ($\sim 10^6$ pfu/mL), the PANCE dilution schedule was kept as consistent as possible in an effort not to bias the evolution. Sequence data was collected every 10 passages and stringency was increased by diluting each subsequent passage by an order of magnitude more than the previous 10 passages, eventually allowing me to dilute each passage 100000-fold between rounds in DRM without any signs of washout. At this stage, these populations could easily be moved to PACE, but I decided to test a separate question: can PANCE allow phage selection in less rich and/or more defined growth media? The use of defined media would permit me to run metabolic selections in growth conditions much closer to those relevant to industry, which typically are always defined to maximize consistency and yield. However, DRM is used for PACE and PANCE selections because of its ability to support robust phage propagation. Ideally, evolved phage populations would propagate sufficiently in less rich media such that selections could continue despite generally lower propagation rates. I prepared M9 media supplemented with trace vitamins and minerals and a low concentration of casamino acids (0.1% w/v, necessary for the auxotrophic S1030 strain used for PACE and PANCE), with glucose (0.2% w/v) as a feedstock for PHB production. Using this “M9+”

media, I ran 10 more passages of PANCE at a less aggressive dilution rate. Seeing maintenance of healthy phage titers (>10^6 pfu/mL) after these passages, I increased the dilution rate as I had previously and ran an additional 10 passages, pausing the experiment at 50 total passages, 30 of which were in DRM and 20 of which were in supplemented M9+ media.

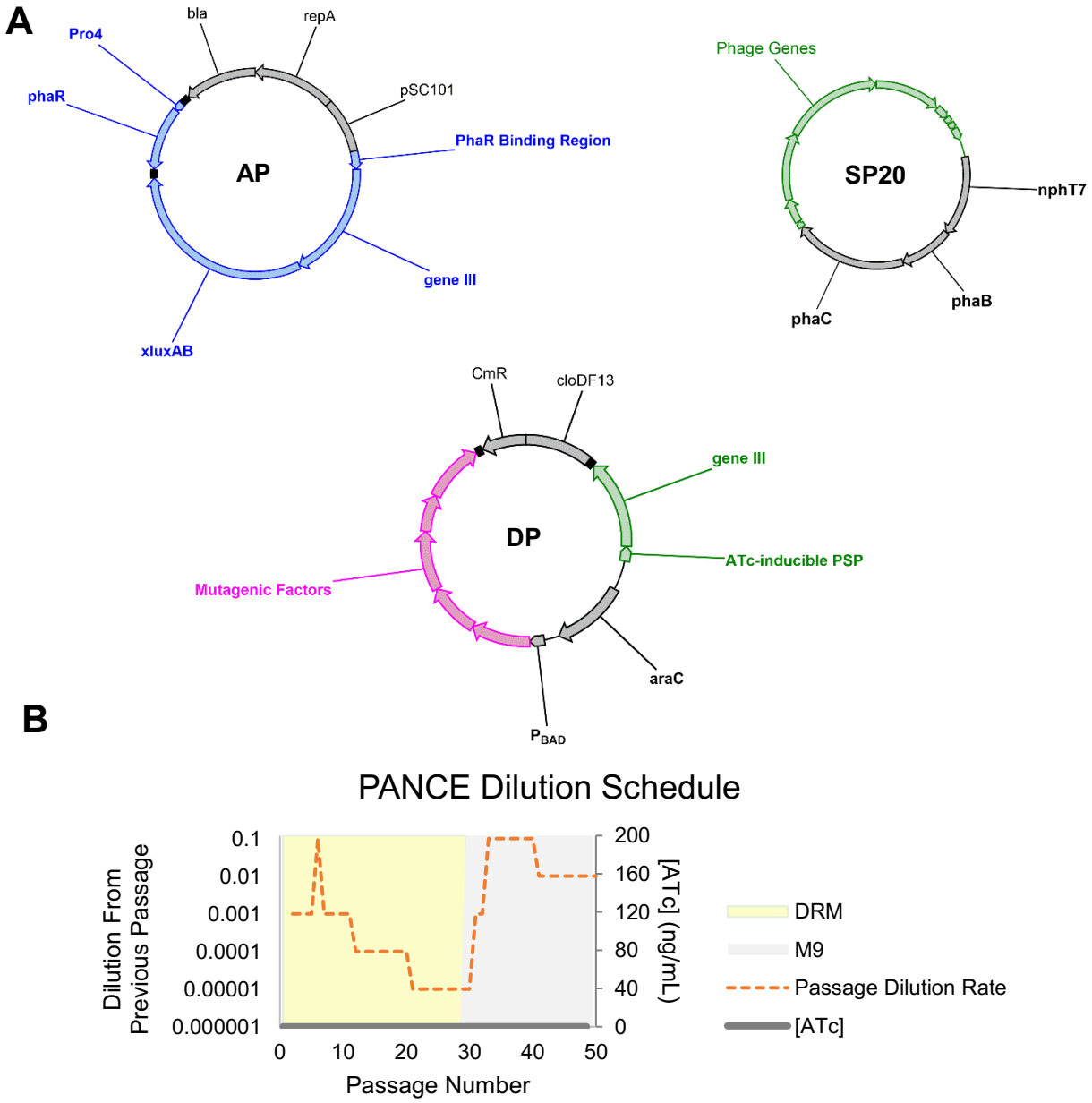


Figure 16: PANCE of an NphT7-Dependent PHB Pathway.
A. S1030 cells containing the Accessory Plasmid (AP) and Drift Plasmid 6 (DP6) were infected with Selection Phage encoding a single transcript with *nphT7* and *phaBC* (SP20). bla and CmR, antibiotic resistance markers; pSC101/repA and cloDF13, origins of replication. **B.** Summary of PANCE conditions for this experiment. Three separate cultures were maintained using the same conditions shown to test divergence of isolated populations. DRM, Davis Rich Media. M9, a minimal medium supplemented with glucose, casamino acids, and trace vitamins and minerals to support growth of the auxotrophic S1030 host.

Sequencing of every 10th passage shows early divergence of the three replicate populations tested (**Figure 17**). Cross-referencing sequencing data with earlier characterizations of mutations in PhaC reveals that no conserved or known essential positions have been mutated in the population consensus. This is possibly due to the final activity-dependent plaque screen performed on the population to ensure proper sequencing and subcloning of each pathway, but also helped assure me that no obvious cheating mechanism—for example direct binding of PhaR by PhaC—had been evolved at the expense of known polymerization activity**.** Two commonly mutated positions in the PhaC region—G4 and F420—have have been previously characterized through site-saturation mutagenesis and shown to impact *in vitro* and *in vivo* activity[69,70]. This F420 position is particularly interesting given its implication in lag-phase reduction for *in vitro* polymer production, though the full extent of how this affects *in vivo* polymer production is unknown. While it did not fix in the population, I observed one instance of a stop codon mutation creating an in-frame fusion of NphT7 to PhaB in population 1 at passage 20. It's possible these mutations are less common in my sequencing results given the tendency of the defined RBS sequences I placed between these genes to delete bases from their spacer regions, which would shift these fusions out of frame without further insertions or deletions to this region. However, I was pleased to confirm my initial hypothesis that fusion constructs during full-pathway evolution were being sampled and tested for activity during PANCE.

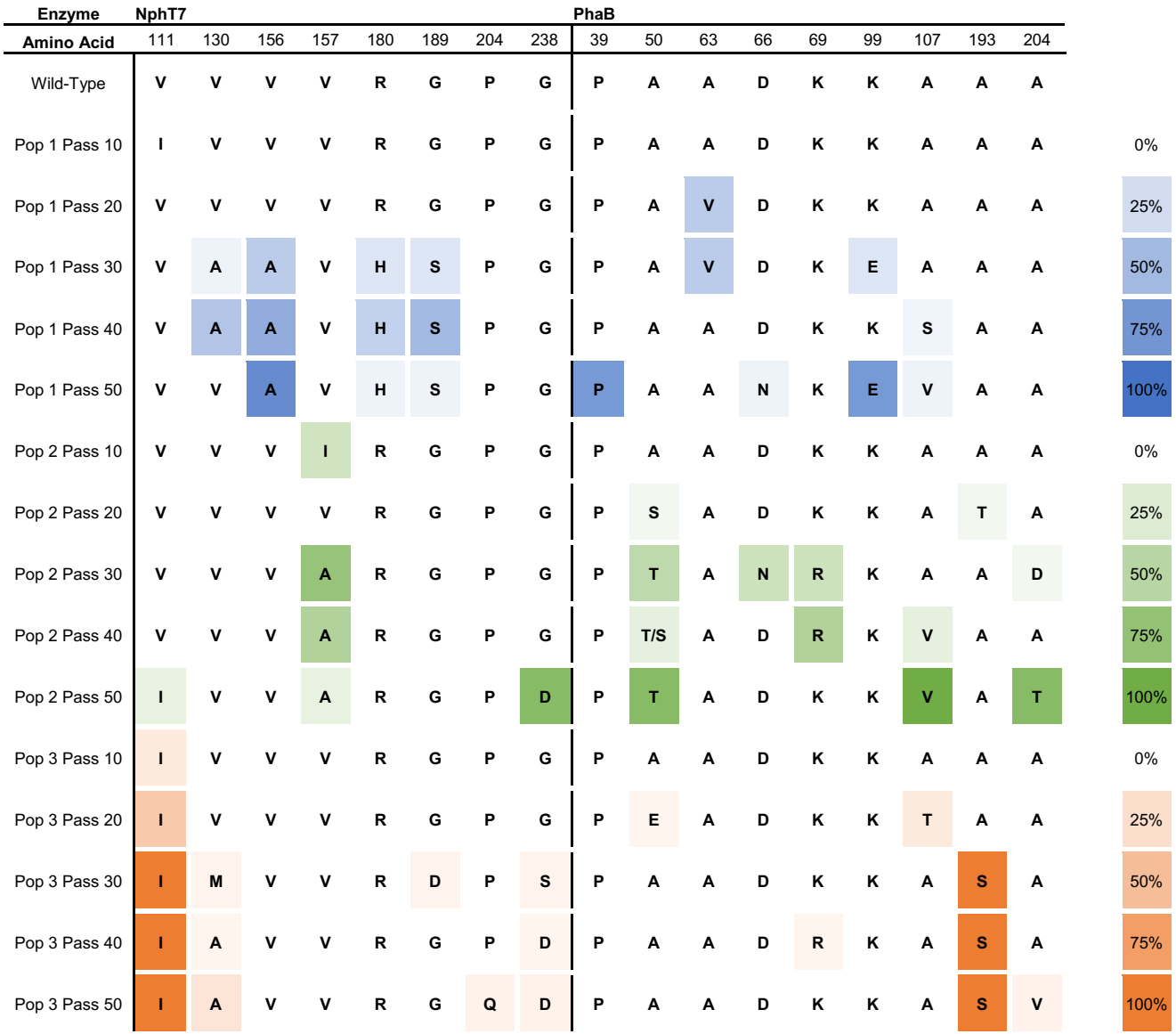


Figure 17: Sequencing summary for SP20 PANCE.
The degree of shading indicates the percentage of convergence for a given mutation at each passage. Pop 1, 2 and 3 are the three replicate populations evolved.

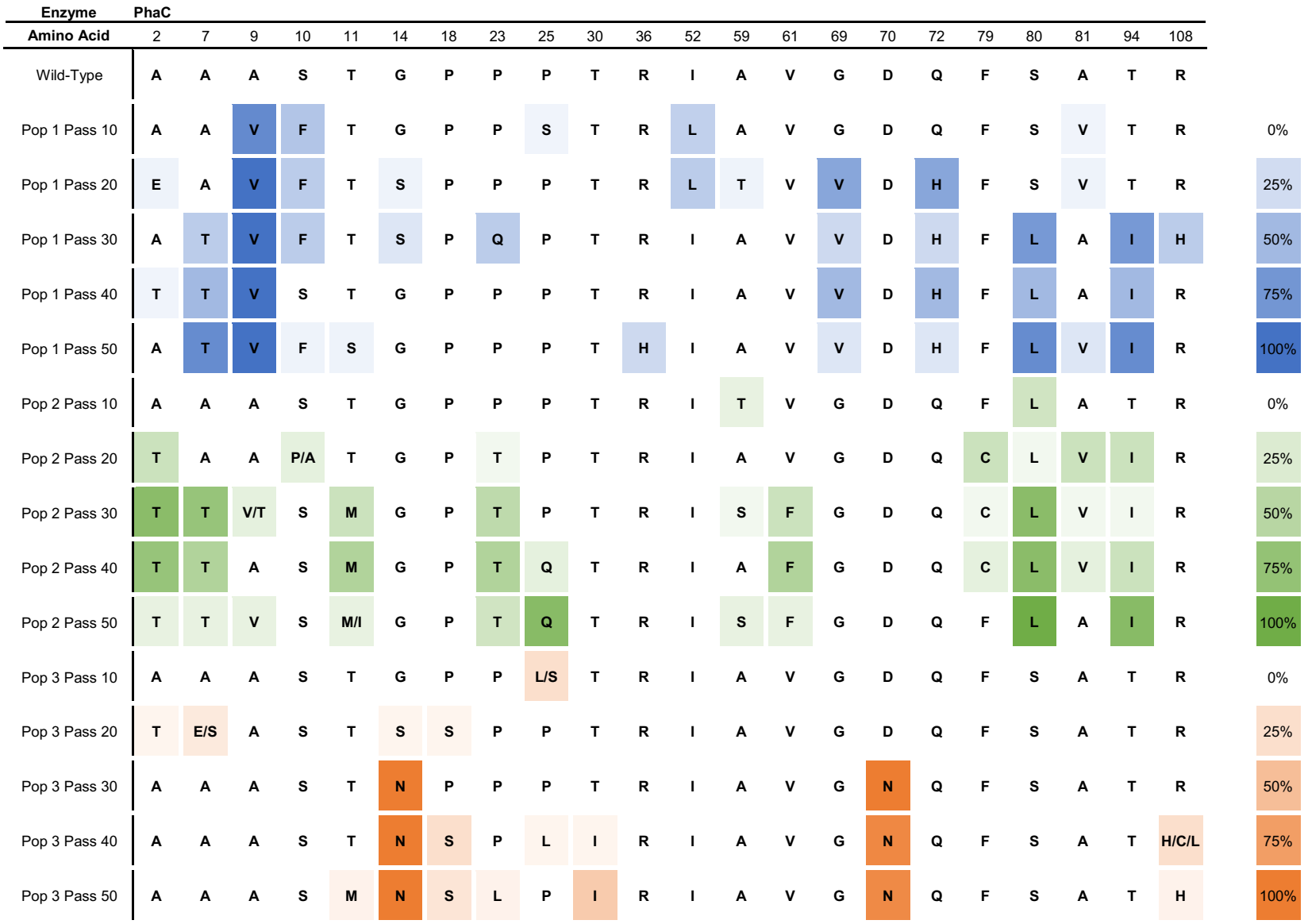


Figure 17 (continued): The degree of shading indicates the percentage of convergence for a given mutation at each passage. Pop 1, 2 and 3 are the three replicate populations evolved

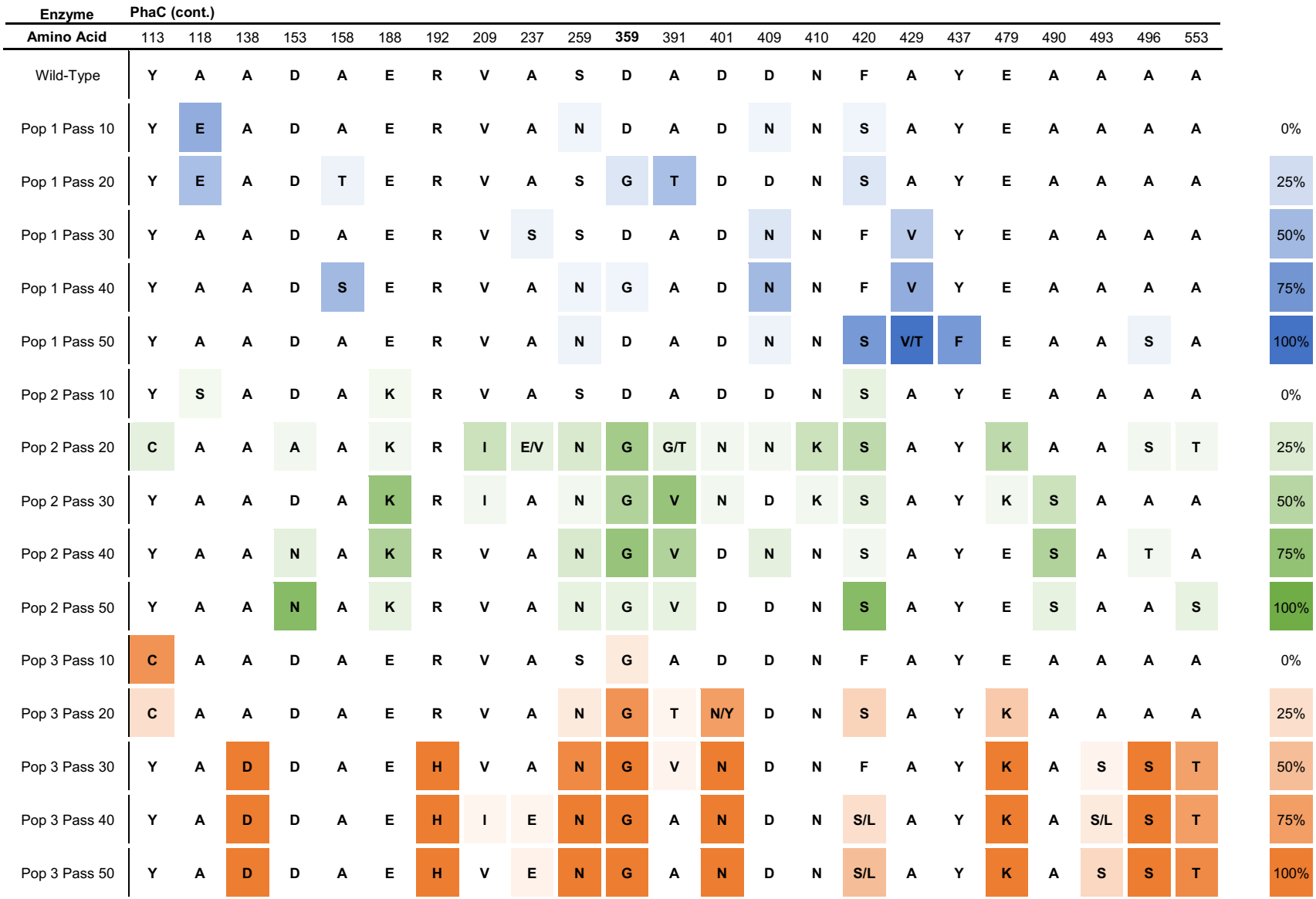


Figure 17 (continued): The degree of shading indicates the percentage of convergence for a given mutation at each passage. Pop 1, 2 and 3 are the three replicate populations evolved

Characterization of these pathway variants is still ongoing, but several variants from later passages have been isolated, subcloned onto a separate vector, and assayed for apparent *in vivo* activity. From these data, significant increases are apparent in all variants tested against the starting, unmutated cassette. High background signal was seen in the absence of arabinose-controlled induction, suggesting a similar embedded-promoter phenotype hypothesized previously for the *kanR-phaC* fusion cassette. Fortunately, this background signal stops increasing relatively early on when monitored continuously and is dwarfed by the signal produced by pathway induction at very low arabinose concentrations for all variants except the population 3 isolate from passage 50. From the data in **Figure 18**, three key improvements in apparent activity have been evolved in these mutant pathways. First, the maximum signal reached within a 4-hour period is far greater for isolates from population 3 of the evolution. Second, all isolated mutants increase in signal significantly faster than the wild-type control. Finally, while the wild-type pathway shows stronger signals at higher doses of arabinose, all mutant pathways show greater signal at far lower levels of pathway induction, suggesting a much more efficient ratio of pathway biomass to product output. One major concern of the sharp increase in background luminescence for these pathways is the potential for direct binding of one or more pathway enzymes to the regulatory PhaR repressor protein controlling reporter transcription. This would allow for robust signal from the PhaR promoter even in the absence of PHB production, effectively "cheating" the selection circuit. I plated a subset of the populations tested in luciferase assays onto agar plates containing Nile Red and either arabinose (to induce PHB production) or glucose (to repress transcription from the P_{BAD} promoter). I compared serial dilutions of replicate cultures on the glucose plates to the arabinose plates and found Nile Red accumulation to be much greater for most populations on the arabinose plates than on the glucose plates, suggesting a strong PHB production phenotype and helping alleviate any fear of cheaters. However, the passage 50 variant from population 3 showed no Nile Red staining, suggesting it might be de-repressing PhaR through a currently unknown mechanism

unrelated to PHB production. More characterization will be necessary to determine if the mutations in any of the individual mutants tested contribute to improved PHB production, but the retention of PHB production on plates for most tested variants and the sharp increase in apparent *in vivo* activity for all tested variants are promising signs of progress.

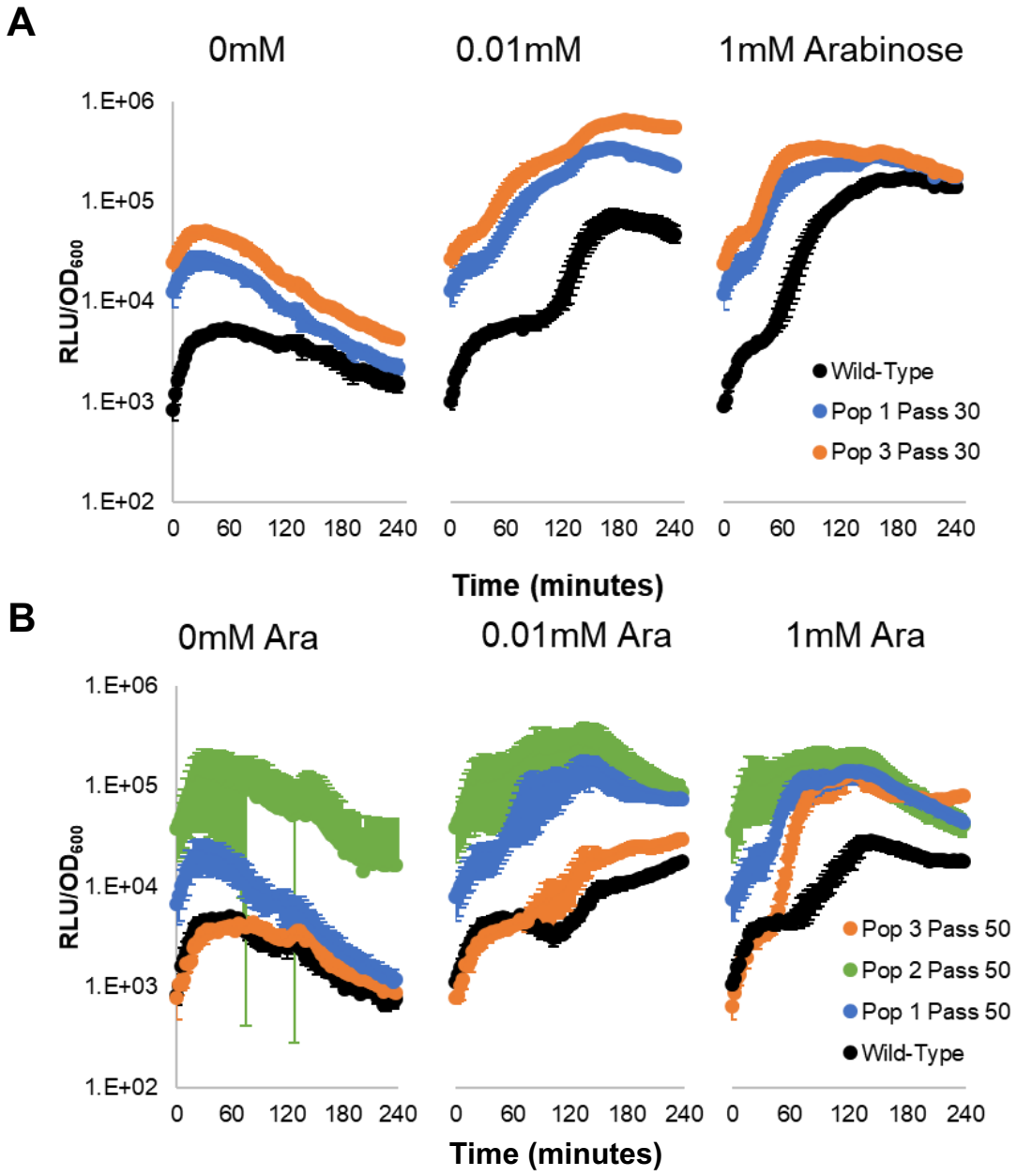


Figure 18: Apparent Activity of individual pathway variants isolated at passage 30 and 50.
Luminescence was monitored every 2.5 minutes over 4 hours after induction of pathway expression using the indicated concentration of arabinose. Points represent the average and errors bars the standard deviation across 6 biological replicates. **A.** Isolated variants at passage 30 tested for activity in DRM. **B.** Isolated variants at passage 50 tested for apparent activity in M9+ Media.

Section 1.8: Progress toward Phage-Assisted Evolution of Novel PHB Derivatives

While PHB is a good model polymer for selection design, validation, and optimization, it is not on its own a particularly high-value target. Of the hundreds of monomers found in naturally occurring PHAs[80], the ultimate goal of my PhaR-based PACE selection is the either the evolution of existing, low-activity pathways known to incorporate novel monomers toward improved native function or the evolution of altered specificity for high-activity pathway components not known to produce novel PHA materials. All known PHA synthases are *R*-specific given that *S*-PHAs have never been isolated despite a vast array of other monomer properties observed in nature[80]. Stereochemical changes to polymer chains can impact a wide variety of industrially relevant PHA polymer properties[81], making the *R*-specificity of these enzymes a potential limitation in producing PHAs with useful properties. To break the stereospecificity of PHA synthases, I sought to take advantage of ketoreductases known to produce an excess of *S*-3-hydroxybutyryl-CoA, limiting the total amount of PHB that can be produced from *R*-3-hydroxybutyrate pools. Current efforts use a pathway with Hydroxyacyl-CoA Dehydrogenase 2 (Hadh2) from *Ascaris suum*, which produces *S*-3-hydroxybutyryl-CoA at over ten times the rate of *R*-hydroxybutyryl-CoA[82]. By constitutively expressing this gene in the host along with *nph*T7 and placing a PHB synthase gene on the selection phage, a selective environment is created that theoretically limits *R*-PHB production through overproduction of *S*-monomers in the cytosol. Similarly, the same architecture can be applied using an *S*-specific Hydroxybutyrate Dehydrogenase (Hbd) from *Clostridium acetobutylicum*[63], which should eliminate *R*-PHB production outside of background metabolites from *E. coli*. Initial tests of these various background strains suggest that these *S*-selective and specific ketoreductases limit apparent overall PHB accumulation phenotypes, and Nile Red plate-based screens show clear polymer accumulation for strains expressing PhaB and Hadh2, but not Hbd. Together, these data suggest the basis for a selection for *S*-PHB production using PACE as illustrated in **Figure 19**.

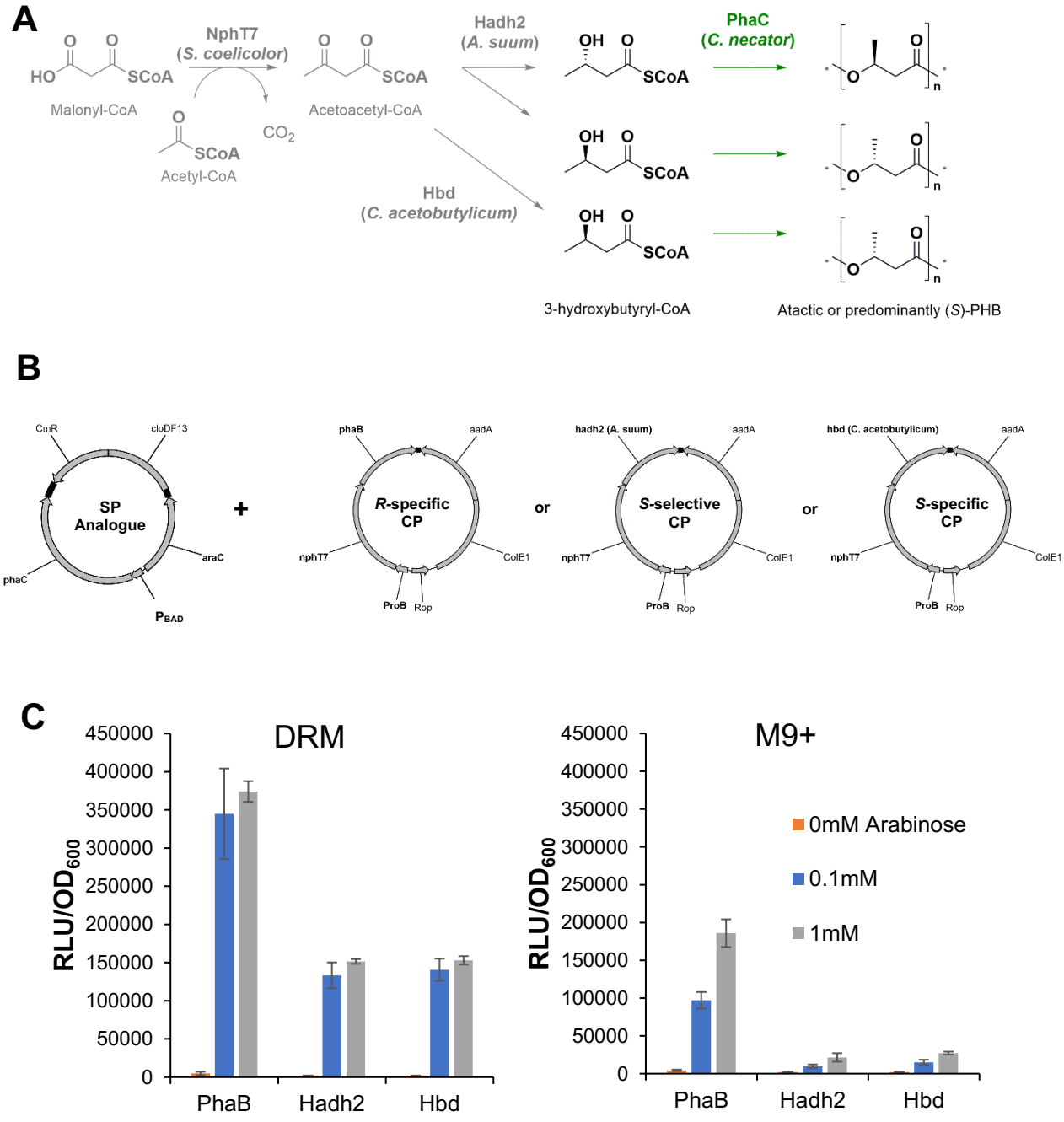


Figure 19: Selection for evolving *S*-3-hydroxybutyryl-CoA incorporation by PhaC.
A. Summary of the complementary pathways (gray) used to evolve the PhaC reaction (green) toward lower stereospecificity. Two keto-reductases, one stereoselective and the other stereospecific, are shown providing the (*S*)-monomer not known to be incorporated during polymerization. **B.** List of plasmids used to validate the selection conditions in mock host strains. All strains additionally had the PhaR-AP to generate luciferase reporter readout. aadA and CmR, antibiotic resistance markers; ColE1/Rop and cloDF13, origins of replication. **C.** Luciferase reporter data for Davis Rich Media (DRM) and M9+ media. Bars show the average of maximum signals obtained during continuous monitoring for 4 hours after induction of PhaC. Arabinose controls induction of the target PhaC from the P_{BAD} promoter. Error bars show standard deviation across 4 biological replicates. RLU, Relative Luminescence Units. OD_{600}, Optical Density at 600nm.

Initial applications of this evolution system to PACE mostly failed – All *phaC*-containing SPs used developed gene III recombinant cheaters early in the experiment. However, I also included SP20 as a positive control in a separate lagoon to test the hypothesis that *phaB* encoded on the phage backbone itself could make up for the lower concentrations of *R*-hydroxybutyryl-CoA theoretically provided by the CP in the host expressing Hadh2. For most of the experiment, the entire *nphT7-phaB-phaC* cassette could be detected via PCR in the evolving SP20 phage pool, supporting my hypothesis that expression of PhaB from the SP would help make up for the lower (*R*)-3-hydroxybutyryl-CoA pools in the host complement pathway. However, toward the end of the experiment, a notable shift in the size of this PCR product was seen, typically corresponding to a deletion within the region of interest or recombination with the AP to form a gene III recombinant cheater. When I tested these phage for wild-type-like activity, however, none was observed. Sequencing identified these phage as encoding a fusion of the first 18 amino acids of NphT7 with the latter 550 amino acids of PhaC. Two notable mutations were present in the PhaC region of this gene outside of this fusion (T73I and D506N in the fusion, corresponding to T94 and D527 in the original *phaC* gene). The N-terminus of PhaC was highly mutated in my previous evolution of a full PHB production pathway, and it has been previously reported that the N-terminus of PhaC is not essential for PHB synthase activity[74]. Interestingly, the N-terminal region has been proposed as a modifier of molecular-weight distribution and product specificity for *C. necator* PhaC[83,84], so I decided to test these variants further.

I sub-cloned one of the NphT7-PhaC fusion genes from an isolated phage from the lagoon and tested it using the same assay described in **Figure 19**. I decided to test in M9+ media to best assess any improvements in incorporation of non-wild-type substrates. The data are shown in **Figure 20**. The high background is very reminiscent of my original KanR-PhaC fusion protein results, and similarly suggest that an embedded promoter in this new sequence is somehow enabling formation of an N-terminally modified PhaC protein even before addition of arabinose.

This early induction would help explain the increase in signal for all strains tested, especially given the disproportionately large background signal seen specifically for the PhaB complement. Were this background due to a non-polymerization phenotype, I would expect this large signal to appear regardless of the pathway complement provided.

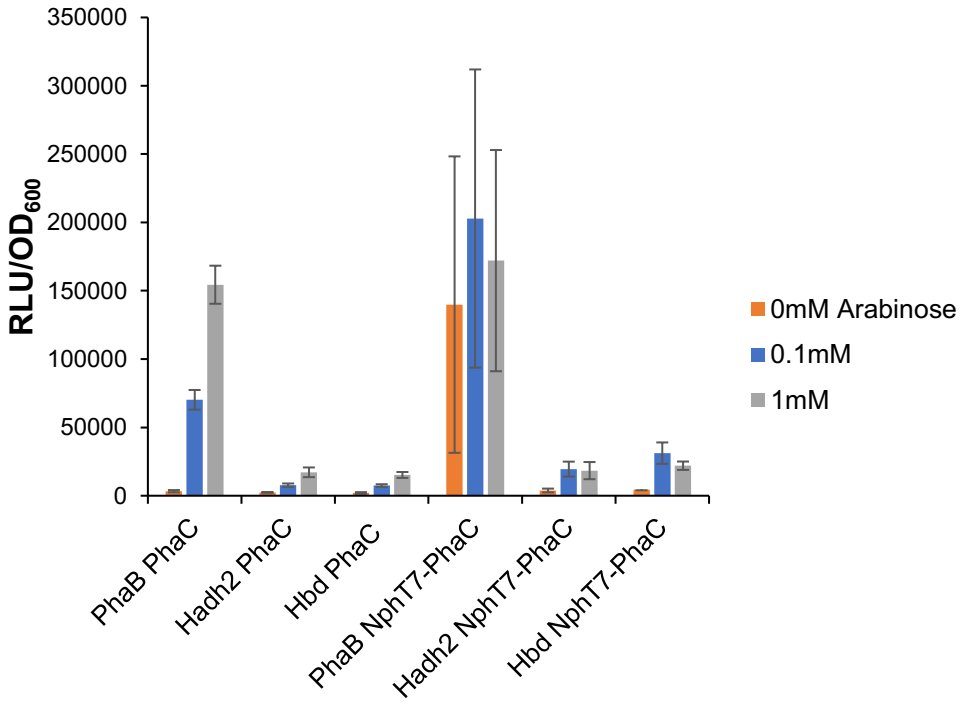


Figure 20: PACE-evolved NphT7-PhaC fusion shows significantly increased background activity.
Bars show average maximum luciferase values obtained during continuous monitoring for 4 hours after induction of PhaC. Errors bars show the standard deviation for 4 biological replicates. RLU, relative luminescence units. OD_{600} optical density measured at 600nm. NphT7-PhaC denotes the fusion of the first 18 amino acids of NphT7 to the latter 550 of PhaC.

Given the quick convergence on cheaters from PACE, I also sought to evolve PhaC using either Hadh2 or Hbd in PANCE. I decided to try the same approach as my full pathway evolution, in which the initial set of passages would be performed in DRM and the latter set switched to the more stringent supplemented M9+ minimal media conditions. I started with 2 sets of 3 populations of SP21, one set evolved in the Hadh2 complement strain and another in the Hbd complement strain as illustrated in **Figure 21A.** Currently, I have run 30 passages in DRM and seen most populations evolve cheaters as the PACE experiment did. However, two of the Hadh2 and one of the Hbd

populations did not evolve cheaters by passage 20 – these were sequenced, and the Hadh2 populations were split into two lineages, one passaged further in Hadh2 host and one passaged in Hbd host to see if pre-evolution with the more permissive ketoreductase would change the ability of PhaC to evolve with Hbd. Comparing sequencing data across these 30 passages to the sequences evolved for the full NphT7-PhaB-PhaC pathway suggests one or two divergent mutations and a few key similarities, most notably S80L (**Figure 21C**). S80P mutations have been previously reported to lower, but not eliminate, PHB production[49], but it's unclear what the difference between a proline and leucine residue might be for this site. Further evolution of these lineages in M9+ media will hopefully force more drastic changes, but the threat of wild-type recombinant cheaters should not be ignored. My recommendation for future work is to re-code the recombination-prone portions of these selection phage and attempt the selection in M9+ media directly. Characterization of these DRM-evolved PhaC variants might offer better starting activities to further enable this M9+ evolution. Ideally, this selection will demonstrate that PANCE can be a valuable approach for evolving metabolic pathways in less noisy growth conditions, providing better control than typical, rich-media selections like PACE.

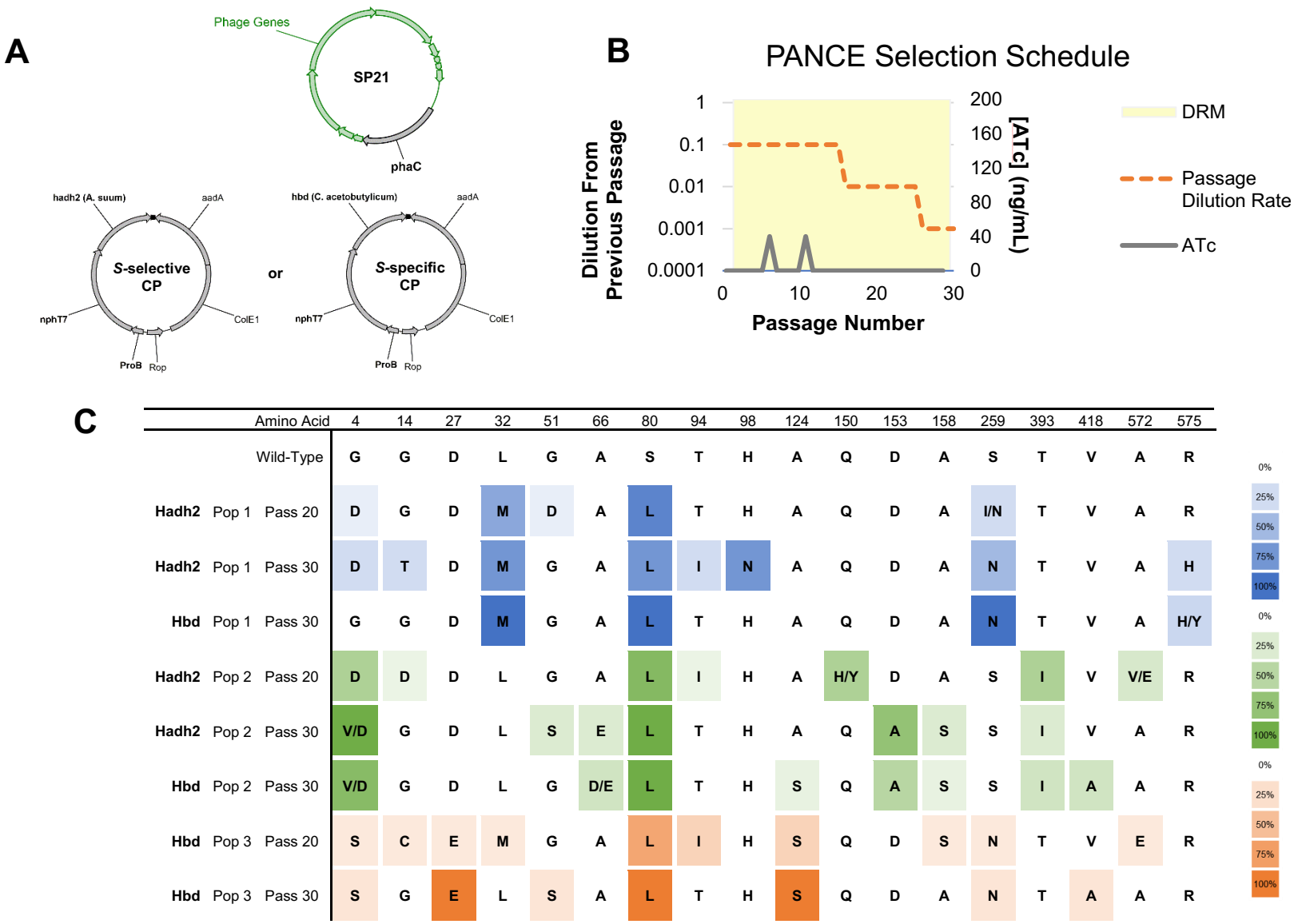


Figure 21: PANCE of PhaC with *S*-selective or *S*-specific complements.
A. Plasmids used for PANCE of *S*-PHB. aadA, antibiotic resistance marker; ColE1/Rop, origin of replication. **B.** Summary of PANCE conditions for the experiment **C.** Sequencing summary for passages 20 and 30 for non-cheating populations. The degree of shading indicates the percentage of convergence for a given mutation at each passage. Pop 1, 2 and 3 are the three replicate populations evolved.

Section 1.9: Discussion

Previously published evolutions of PHA production phenotypes have all relied on synthetic library construction or unreliable dye-based screens. My work shows the potential of *in vivo* reporter-based approaches as an alternative to these technologies. While this selection has yet to demonstrate a ground-breaking result for novel PHA production as I had initially desired, I observed the evolution of previously characterized, beneficial mutations in the *C. necator* PHB Synthase PhaC[69,70] in the context of targeted, phage-assisted evolution of full PHB production pathways. Further characterization of mutations in this synthase, as well as characterization of mutations seen in the NphT7 and PhaB enzymes encoded upstream will hopefully elucidate the mechanisms of my selection further. The capacity of PHB to act as a metabolic sink additionally allowed me to evolve NphT7 on its own, suggesting this selection as a great potential tool for driving improved flux through upstream components of PHA production pathways provided that an appropriate PHA synthase gene can be identified and encoded in the selection host. Such systems could be applied toward evolving fatty acid synthase machinery or other pathways of great interest to value-added small molecule production in *E. coli*.

Another important result from my work on PHB evolution is the validation of M9+ as a PANCE-compatible selection medium. As I show in my validation of the *S*-PHB selection strains, DRM can significantly limit the stringency of metabolic selections by significantly increasing background activity in strains that a metabolic engineer would typically want to be resource-limited to minimize cross-talk and unpredictable metabolite concentrations during selection. Further work with different media conditions in PANCE might be able to identify fully-defined media for selections that would allow researchers to have even more fine-tuned control over their selection environments, further enabling this system to be applied to complex and impactful metabolic pathway evolutions. Together, I hope that my attempted evolutions of various PHB components will inspire future PACE work to focus not just on increased flow-rates or intricate

synthetic biology circuit design, but also to look at host engineering and growth conditions as a viable strategy, especially when coupled to easily parallelized procedures such as PANCE.

Section 1.10: Materials and methods

Reagents

Unless specified, all chemical reagents were purchased in the highest grade available from Sigma–Aldrich. For luciferase reporter assays, M9 salts and LB, 2xYT, Agar, and casamino acids were purchased from US Biological Life Sciences. Trace Elements (MD-TMS) and Vitamin Solution(MD-VS) were purchased from ATCC. Antibiotics and Arabinose were purchased from GoldBioTechnology Inc.

Strains and plasmids

Invitrogen Mach 1 T1R (Invitrogen) or NEB Turbo (NEB) chemically competent *E. coli* strains were used as cloning hosts. Luciferase reporter assays, phage-based assays, and all evolutions were carried out using E. coli S1030[8]. The original plasmid containing the *C. necator phaABCP* cassette (pMC001579) as well as plasmids containing the *nphT7* and *A. suum hadh2* genes were gifted by Michelle Chang's Research group at University of California Berkeley. The pMW218 plasmid encoding the *P. denitrificans phaR* gene was a gift of the Taguchi Lab at the School of Engineering of Hokkaido University, Japan.

Cloning

Plasmids and selection phage were constructed using USER cloning or KLD Enzyme Mix (NEB). DNA fragments were generated by PCR using Pfu Turbo Cx Hotstart DNA Polymerase (Agilent), VeraSeq 2.0 High Fidelity DNA Polymerase (Enzymatics), or Phusion U Hot Start DNA

Polymerase (Thermo Fisher Scientific). All amplicons were purified using kits from Qiagen and digested with DpnI during PCR fragment assembly for USER cloning. All restriction endonucleases and USER enzyme were purchased from NEB. Assembled vectors were transformed into chemically competent *E. coli* of various strains and verified by Sanger sequencing after amplification from individual colonies using illustra TempliPhi DNA Amplification kits (GE Healthcare). Cell growth for cloning purposes was carried out using 2xYT media supplemented with appropriate antibiotics (Kanamycin, 50μgml^{-1}; Carbenicillin, 50μgml^{-1}; Spectinomycin, 100μgml^{-1}). For phage cloning and replication, phage were transformed directly into S1059, S1381 or an equivalent phage containing a PSP-gene III plasmid (e.g. pJC175e or similar)[8] and grown overnight. For cloning phage containing a full PHB production pathway (SP2, SP20), cloning was done directly into S1030 pTR26p4pD cells or outgrowth from S1381 cultures was limited to 6 to 8 hours instead of overnight. All transformed phage cultures were then plated on lawns of this same strain and individual plaques were picked into fresh media, grown for a minimum of 6 hours, and verified by Sanger sequencing after amplification using illustra TempliPhi DNA Amplification kits (GE Healthcare).

Luciferase-based reporter assays

For all reporter assays, the desired number of individual colonies from transformation plates were grown overnight at 37°C in Davis Rich Media (DRM) or M9+ medium (M9 supplemented with 0.2% glucose, 0.1% casamino acids, and 1% (v/v) each of trace mineral and vitamin solution). Cultures were diluted 100 to 1000-fold into fresh medium supplemented with anhydrotetracycline (ATc) if needed for PhaR expression and grown until early exponential phase (OD approximately 0.4). At this stage, arabinose was added to induce expression from P_{BAD}. For continuous monitoring of luminescence, 200μl of sample were placed into a black, Corning black clear bottom 96 well plate and analyzed for optical density at 600nm and luciferase activity at 37°C on an Infinite Pro M1000 plate reader (Tecan). For discrete timepoints, 150uL of culture was

transferred to the same plates at given times post-induction and measured the same way. All liquid cultures and continuous cultures during plate-reading were grown with regular shaking or stirring.

Plaque Assays for Phage Titer Quantification

For plaque assays used to clone and titer phage, the desired host strain for plaquing was grown at 37°C in 2xYT to late log-phase (OD_{600} 0.6-1.2). 100-200μL of host culture was infected with 10μL of a phage dilution series and diluted into 1mL of molten 0.75% (w/v) agar in 2xYT and immediately plated on 1.5% (w/v) agar in 2xYT media. Agar was cooled and set before plates were inverted and grown overnight at 37°C.

Phage Infectibility Tests

For the kanamycin resistance assay for phage infectibility, target host strains were grown to log phase and infected with set quantities of phage particles before plating on 2xYT agar containing Kanamycin (50μgml^{-1}). The total colony-forming units obtained indicates the number of KanR-expressing phage that were able to infect the host cells.

For determining infectibility through plaque assays, the same general procedure was repeated, but wild-type phage encoding their own copy of gene III were added to host cells and mixed with molten 0.75% (w/v) agar in 2xYT and plated on 1.5% (w/v) agar in 2xYT, neither of which had antibiotic added. If host is infectible, these phage will form visible plaques within 12 hours at 37°C.

Overnight Phage Expansion Assays

Desired host strains for propagating phage were grown at 37°C in either 2xYT or DRM to late log-phase (OD_{600} 0.6-1.2). This culture was split into the desired number of 2mL cultures and infected with a known quantity of phage and infected cultures were grown for a set time at 37°C.

Cultures were centrifuged at max RPM for 2 minutes to separate phage in the media from host cells and this supernatant phage population was tittered using plaque assays to determine the degree of population expansion in liquid culture.

PACE Experiments

PACE experiments were performed as previously described[7–12,58]. All chemostat and lagoon systems were maintained using Masterflex Digital Pump (Cole-Parmer) systems at fixed RPM values manually calculated to provide a desired flow-rate for the tube diameter used during the experiment. TSS chemically competent *E. coli*[85] S1030 were transformed with desired AP, CP, or MP and plated on 2xYT agar containing 0.5-2% glucose (w/v). A single colony was grown to saturation overnight at 37°C in DRM containing appropriate antibiotics and diluted the next day 100- to 1000-fold into a chemostat at 37°C containing 50-100mL of Davis Rich Media supplemented with appropriate antibiotics for the AP/CP/MP used (Carbenicillin, 50μgml^{-1}; Chloramphenicol, 40μgml$^{-1;}$ Spectinomycin, 50μgml^{-1}). Once the chemostat reached an OD_{600} of ~0.8-1.2, dilution was started and adjusted in order to best maintain this OD_{600} range, which varied by host but typically fell in the range of 0.5-1.2 chemostat volumes per hour. Chemostat media was flowed into lagoons at a desired flow-rate. Lagoons were treated with 1M arabinose solution pumped from a syringe pump (New Era Pump Systems) at a rate adjusted to maintain a concentration of 10mM in the lagoon. Phage were injected into lagoons to start the evolution and collected from lagoon waste needles or waste lines at desired timepoints. Phage titers were determined by plaquing onto pJC175e-containing S1030 derivatives unless otherwise indicated.

Sequencing data was collected by picking individual plaques into fresh media and growing overnight. Overnight cultures were spun down and the supernatant used as template material for rolling circle amplification using illustra TempliPhi DNA Amplification kits (GE Healthcare). Sequences were determined by Sanger sequencing and results were aligned using SeqMan

alignment software (DNAStar) and manually analyzed and recorded. All liquid cultures and were grown with continuous shaking or stirring.

PANCE Experiments

TSS chemically competent *E. coli*[85] S1030 were transformed with desired AP, CP, or MP and plated on 2xYT agar containing 0.5-2% glucose (w/v). A single colony was grown to saturation overnight at 37°C in DRM containing appropriate antibiotics and diluted the next day 100- to 1000-fold into fresh DRM. Cultures were grown to log-phase (OD 0.3-0.6), treated with 10mM arabinose to induce mutagenesis, the desired amount of anhydrotetracycline for a given passage (typically 0 or 40ng/mL), split into the desired number of either 2mL cultures in single culture tubes or 500uL cultures in a 96-well plate, and infected with selection phage. Cultures were grown overnight at 37°C and harvested the next day via centrifugation (max RCF for 2 min). Supernatant containing evolved phage was isolated with optional filtration through a 0.2μm Costar spin filter (Corning) and stored at 4°C. These phage were then used to infect the next passage and the process repeated for however many passages were desired for the selection. Phage were diluted passage to passage a maximum of 10-fold and a minimum of 100000-fold. Phage titers were determined by plaquing onto pJC175e-containing S1030 derivatives unless otherwise indicated. Note that for all SP20 and SP21 PANCE experiments, chloramphenicol stocks were dissolved in DMSO rather than ethanol to prevent ethanol from affecting polymer accumulation during selection[86].

Sequencing data was collected by picking individual plaques into fresh media and growing overnight. Overnight cultures were spun down and the supernatant used as template material for rolling circle amplification using illustra TempliPhi DNA Amplification kits (GE Healthcare). Sequences were determined by Sanger sequencing and results were aligned using SeqMan alignment software (DNAStar) and manually analyzed and recorded. All liquid cultures and were grown with continuous shaking or stirring.

HPLC Quantification of PHB

TSS chemically competent *E. coli*[85] S1030 were transformed with a set of plasmids with PHB production genes partially or fully controlled by the P_{BAD} Arabinose-inducible promoter system and plated on 2xYT Agar containing 0.5-2% glucose (w/v). Colonies were picked into DRM with appropriate antibiotic and grown overnight at 37°C to saturation and back-diluted 100- to 1000-fold into fresh media. At early-to-mid log phase, PHB production was induced via addition of indicated concentrations of arabinose and cultures were grown further at 37°C. All liquid cultures and were grown with continuous shaking. After desired growth times, OD_{600} measurements and luciferase assay measurements were taken as needed. 1 to 2mL samples of each culture for PHB analysis were harvested by centrifugation (max RCF for 2 minutes) and the supernatant was discarded. Cell pellets were frozen at -80°C and dried overnight via lyophilization. Dry cell pellets were transferred to 4mL glass vials. Additionally, a small (1 to 10mg) sample of crystalline PHB was placed in a separate vial as a positive control for digestion efficiency. 1mL of concentrated sulfuric acid was added to each vial and cells were capped and heated for a minimum of 1 hour at 90 to 95°C with periodic shaking to ensure mixing. Vials were cooled on ice and 100uL of each digested pellet was added to 400uL of 0.028M sulfuric acid solution and filtered through a PVDF spin filter column (MilliporeSigma) at 2000g for 2 minutes. Samples were frozen at -20°C prior to analysis.

For HPLC analysis, 50uL of each thawed sample was added to 450uL of 0.8mg/mL adipic acid solution (aqueous, serving as an internal standard). A standard curve of crotonic acid was prepared from 5mg/mL serially diluted 2-fold a minimum of 6 times 50uL of each dilution was added to 450uL of the adipic acid solution. 100uL samples were injected into 0.028M sulfuric acid mobile phase through either an Aminex HPX-87H Column (BioRad) or a Fast Acid Analysis Column (BioRad) at 0.7ml/min at 60°C for 20 minutes per sample. Crotonic acid was quantified by measuring absorbance at 210nm. Crotonic acid peaks were manually integrated and normalized to the internal standard adipic acid peak. Crotonic acid concentrations were determined from the

standard curve, with necessary volumetric adjustments made to account for all dilutions made up to that point. Final PHB content was calculated by comparing the measured PHB internal digestion control to the weighed quantity at the start of the chemical digestion, and all mg PHB for each sample adjusted by this conversion percentage. Given the incomplete nature of the chemical digestion, this method was used for comparative PHB quantification, and results do not necessarily reflect differences in PHB yield if strains were not assayed in parallel.

Nile Red Plate Screen for PHB Production

1mg/mL stock solutions of Nile Red were prepared in DMSO and stored at -20°C. This was diluted in molten 2xYT agar to a final concentration of 0.5ug/mL with antibiotics and either 2% (w/v) glucose or 10mM Arabinose. A sample of each PHB producing strain to be screened was plated on 10mM arabinose (to induce PHB production) and 2% (v/v) glucose (to repress PHB pathway induction) and grown overnight at 37°C. Plates were visually inspected for red pigmentation the following day or checked for fluorescence using a gel imager with a mid-range UV light source (typically used for ethidium-bromide-stained agarose gel imaging).

NphT7 Purification

pET vectors encoding 10x-His-NphT7 genes were transformed into BL21* (de3) Chemically Competent *E. coli* (Invitrogen) and plated on 2xYT Agar with 50µg/mL carbenicillin at 37°C. A single colony of each variant was inoculated into 10mL of 2xYT media with 50µg/mL carbenicillin and grown overnight at 37°C. These 10mL cultures were transferred to 1L of LB with 50µg/mL carbenicillin and grown to OD 0.5-0.7, cooled on ice for 20min, and induced with 1mM IPTG overnight at 20°C (220 RPM). Cultures were spun down at 4000g for 15 minutes and resuspended in lysis buffer (Protease inhibitor (Roche cOmplete, Mini, EDTA-free), 100 mM Tris-HCl (pH 8.0), 500 mM NaCl, 1% (v/v) Tween 20, 1mM DTT, and 20% (v/v) glycerol.), placed on ice, and lysed via

sonication (1 sec on, 1 sec off for 10 minutes at medium power). Lysed cells were spun down for 20 minutes at 10000g. Soluble lysate was decanted and re-spun under the same conditions. The soluble lysate was incubated directly with 2.5mL of Ni-NTA resin (5mL of a 50% solution) at 4 degrees for 30 minutes on an automated rotating platform. Resin was poured into a column and washed with 15mL of wash buffer (100 mM Tris-HCl (pH 8.0), 500 mM NaCl, 1% (v/v) Tween 20, 1mM DTT, and 20% (v/v) glycerol) at 0mM imidazole. The column was then washed with 15mL of 10mM imidazole in wash buffer, followed by 15mL of 50mM imidazole wash buffer. The enzymes were eluted from the column using 300mM imidazole in wash buffer and transferred to Amicon 10kDa CO Spin Columns (MilliporeSigma) and concentrated. A 1/100 dilution of this concentrate as well as intermediate samples from the purification process were loaded was analyzed via SDS-PAGE using a Bolt 4-12% Bis-Tris Plus Gel (ThermoFisher) and visualized with Coomassie stain. This purified concentrate was then loaded onto a Hi-Trap HP SP (GE Healthcare) cation exchange column ÄKTA Pure Fast Performance Liquid Chromatograph (GE Healthcare) and exchanged into storage buffer (100mM HEPES, 100mM NaCl, 20% (v/v) glycerol, 1mM DTT in ddH20, pH 7.5). The final yield was 10-20 mg of enzyme as measured using a NanoDrop 1000 spectrophotometer (ThermoFisher). Aliquots of enzyme were stored at -80°C prior to analysis.

NphT7 Kinetics

NphT7 activity was assayed largely as previously described[62]. Total reaction volumes were performed at 0.5mL in a 1mL quartz cuvette. Substrates were added to reaction buffer (100mM Tris-HCl pH 8.0, 5mM $MgCl_2$, 1mM DTT) using either 100μM malonyl-CoA for variable acetyl-CoA assays or 200 μM acetyl-CoA for variable malonyl-CoA assays. This mixture was incubated at 30°C for 1 minute before the reaction was initiated by addition of enzyme (5ug total per reaction). Reaction progress was monitored by measuring formation of the magnesium-enolate complex via increase in absorbance at 303nm in a Beckman Coulter DU-800 spectrophotometer equipped with a

temp-controlled cell holder set to 30°C. Initial velocities were determined from the slope of a plot of the enolate complex formation vs. incubation time for the linear portion of this increase. The molar extinction coefficient (ε) of the enolate complex at 303 nm was assumed to be 8,300. Steady-state kinetic parameters were calculated by fitting the data to the Michaelis-Menten equation using gnuplot software.

Chapter 2: Development of a Formaldehyde Biosensor with Application to Synthetic Methylotrophy

Adapted from Woolston Benjamin M., Roth Timothy B., Kohale Ishwar, Liu David R. & Stephanopoulos Gregory. Development of a formaldehyde biosensor with application to synthetic methylotrophy. *Biotechnology and Bioengineering* **115**, 206–215 (2017)

Section 2.1: Introduction

Formaldehyde is a highly toxic chemical and is classified as a human carcinogen by the International Agency for Research of Cancer[87]. As a potent electrophile, its toxicity stems from its ability to react rapidly with nucleophilic components of DNA, RNA, and proteins, leading to protein and DNA damage in the form of crosslinking[88]. Somewhat paradoxically, formaldehyde is also a ubiquitous intermediate in one-carbon metabolism across the tree of life. In methanotrophs, methane is oxidized to methanol by methane monooxygenase (Mmo), and then to formaldehyde by a PQQ-dependent methanol dehydrogenase (Mdh)[89]. Methylotrophic yeasts such as *Pichia pastoris* convert methanol to formaldehyde using an FAD-linked alcohol oxidase (AOX)[90], and Gram-positive methylotrophs typified by *Bacillus methanolicus* perform the same conversion using an NAD-linked Mdh[91]. In all these organisms, formaldehyde acts as a branch point between further oxidization to CO_2 for energy conservation, and incorporation into biomass via the serine cycle, ribulose monophosphate (RuMP) pathway, or the xylulose-5-phosphate (Xu5P) pathway[89,91,92]. These pathways are of growing interest in the field of metabolic engineering, where researchers seek to convert relatively cheap methanol feedstocks into higher value commodity chemicals with either native or "synthetic" methylotrophs[93,94]. Besides methylotrophs, formaldehyde is present at low levels in all organisms as a result of demethylation reactions[95]. Because of its cytotoxicity, the intracellular formaldehyde concentration must be tightly controlled, which has led to the evolution of a variety of highly coordinated metabolic strategies for detoxifying formaldehyde[96]. The need to keep the concentration of formaldehyde low while supporting high flux places an even more stringent burden on methylotrophs that rely on formaldehyde metabolism for growth. The ability to easily measure intracellular formaldehyde could, therefore, provide basic insights into the regulation of formaldehyde metabolism in native methylotrophs, as well as aid in the development of synthetic methylotrophy. However, typical methods are limited by low sensitivity, cumbersome workflows, or the requirement for expensive HPLC instrumentation. In the work conducted thus far

in *E. coli* on synthetic methylotrophy, formaldehyde has been measured in culture supernatants using the Nash assay[97], taking advantage of the fact that formaldehyde can diffuse rapidly across the cell membrane[98–100]. Due to the low assay sensitivity (Limit of detection, LOD 1μM), evaluation of the synthetic methanol assimilation pathway activity required elimination of the native detoxification pathway. The assay is also limited to small numbers of samples in a kinetic experiment due to the need to separate cells from supernatant before analysis. The gold standard for environmental formaldehyde quantification involves derivatization with 2,4-dinitrophenyl hydrazine (2,4-DNPH), followed by HPLC to separate the various carbonyl derivatives before quantification via UV[88]. While highly sensitive (LOD 0.2μM)[101], this technique suffers from the same bottleneck of requiring cell separation before analysis, and the additional challenges of low throughput and high cost due to the requirement for HPLC separation. Recently, there has been tremendous interest in developing genetically encoded biosensors for monitoring the concentration of a multitude of different compounds[102–104]. These sensors offer several advantages compared to traditional methodologies: Since the signal is often a fluorescent protein such as GFP, sensor read-out can be determined easily using widely available instrumentation without the need for separating cells from their media, and is, therefore, adaptable for high-throughput sampling. In addition, because GFP is stable, the fluorescence signal represents an integral of the substrate concentration over time, allowing for significantly increased sensitivity compared to single time-point measurements. Biosensors are also powerful tools for directed evolution[6,105], thus the development of a formaldehyde sensor would aid ongoing efforts to evolve more active variants of Mdh[100]. Finally, biosensors are gaining interest in metabolic engineering for their ability to actuate a dynamic metabolic response to the presence of the target analyte[4,106]. Such a regulatory strategy could mitigate the toxicity of formaldehyde in an engineered methylotrophy. The native glutathione-dependent formaldehyde detoxification pathway in *E. coli* provides a convenient architecture for a formaldehyde biosensor. In this pathway, formaldehyde reacts spontaneously

with the nucleophilic cysteine residue of glutathione to form the hemiacetal S-(hydroxymethyl)-glutathione. This adduct is enzymatically oxidized to S-formylglutathione by FrmA. Hydrolysis of this species by FrmB liberates glutathione and produces formate, which is much less toxic than formaldehyde[107]. Expression of *frmA* and *frmB* is controlled by a repressor protein FrmR, expressed in the same operon. In the absence of formaldehdye, FrmR binds to the promoter region, preventing transcription. In the presence of formaldehyde, the nucleophilic Cys36 of the FrmR reacts with formaldehyde and causes a conformational change that results in dissociation from the promoter[108–110]. Addition of 0.25mM formaldehyde is sufficient to induce an approximately 100-fold increase in *frmA* transcript after 30min[111]. Previously, Tralau et al. employed a GFP-linked biosensor based on this regulatory system to detect formaldehyde produced during the oxidation of dimethylglycine in *E. coli*[112]. In a more recent publication, this biosensor was improved through a sort-seq approach to screen multiple mutations of the *frm* promoter region to improve dynamic range and elucidate the repressor bindings sites[113]. The authors further showed that the evolved stronger promoter could enhance the benefit of methanol on biomass formation during growth on yeast extract. Here we report further characterization of the biosensor, rational mutation of the promoter to increase sensitivity at low formaldehyde concentrations, and significant advances in the application of the sensor to the ongoing efforts to engineer synthetic methylotrophy. First, we compared the response of an autoregulated version of the sensor, where *frmR* was expressed from its own promoter in an operon with the signal, to a variant where *frmR* was expressed under the orthogonal P_{tet} promoter to remove the negative feedback of the original construct. Second, we tuned the reporter to physiologically relevant formaldehyde concentrations (1–40μM in wild-type cells)[98] by systematic mutation of the *frm* promoter region. To demonstrate the utility of the refined biosensor, we then showed that the reporter could detect formaldehyde production in engineered *E. coli* strains without deletion of *frmA*. To examine the potential of the assay for Mdh evolution, we verified that our reporter could distinguish between the varying activities of various Mdh genes,

and then developed methods to distinguish between individuals within a mixed population by the addition of glutathione to minimize intercellular formaldehyde diffusion. Finally, as a proof-of-concept and first step toward dynamic regulation we examined the ability of our biosensor to report formaldehyde levels in strains engineered for methanol assimilation, facilitating high-throughput optimization of induction conditions for Mdh and the formaldehyde assimilation enzymes Hexulose phosphate synthase (Hps) and Phosphohexulose isomerase (Phi).

Section 2.2: Engineering of the *E. coli* FrmR repression circuit

Auto-regulated negative-feedback promoters are a means of maintaining steady-state levels of gene-expression across varying conditions, providing greater stability to genetic networks[114]. However, this feedback regulation also attenuates the dynamic range of the reporter system, obscuring differences in signal across broad input ranges. Previous work with the *frm* promoter showed relatively limited signal gain in response to toxic concentrations of formaldehyde[112]. We hypothesized that we could achieve a stronger, dose-dependent response to formaldehyde by taking the promoter out of its auto-regulatory context. To test this, we placed *frmR* expression under inducible control of the P_{tet} promoter (construct pTR47). We then compared this construct to an analogous, auto-regulated synthetic plasmid (construct pTR47auto) (**Figure 22**). Both constructs express the luciferase reporter genes *luxAB* from the *frm* promoter, the latter of which was subcloned out of the *E. coli* genome. Interestingly, there was a SNP in this region in our strain, but it did not occur in a sequence predicted to be important in controlling DNA geometry or facilitating protein–DNA interactions[109]. With construct pTR47, in the absence of ATc induction, significant background signal was observed, which rose further upon addition of formaldehyde. The most likely explanation for the high background signal is that the amount of FrmR that arises from chromosomal expression is insufficient to fully repress the multiple-copy plasmid-borne *frm* promoter. That the signal increased further upon formaldehyde addition suggests that this low

level of *frmR* expression is still sufficient to partially repress the promoter. Addition of 40ngml^{-1} ATc to induce *frmR* expression lowered background signal by 100-fold without significantly lowering observed reporter values for the "on" state, enabling detection of formaldehyde between 1 and 100μM with three orders of magnitude difference in signal, and a dissociation constant (K_a) of 37±1.5μM. Comparing this de-coupled repressor to an analogous auto-regulated system (pTR47auto), we saw an order of magnitude improvement in signal-to-background gene expression (**Figure 22**). In addition, the response in the decoupled system showed significantly higher cooperativity, with Hill coefficient n=2.7±0.2 compared to n=1.2±0.1.

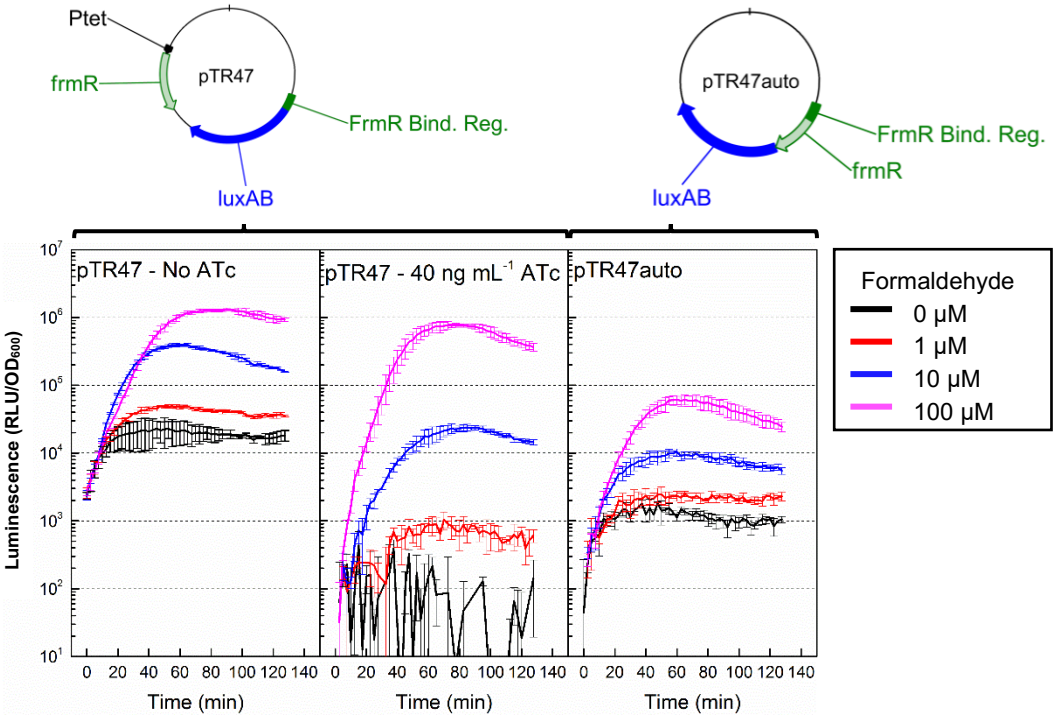


Figure 22: Reporter construct design and formaldehyde response.
Top: The de-coupled plasmid pTR47 controls *frm*R expression through the P_{tet} promoter, while the pTR47auto plasmid mirrors the native context of the *frm* operon of *E. coli.* **Bottom**: Comparison of the coupled (pTR47auto) and de-coupled (pTR47) reporter constructs, with or without anhydrotetracycline (ATc). Error bars show the standard deviation of three biological replicates. RLU, relative luminescence units; OD, Optical Density at 600nm. Breaks in the lines for the 40ngml−1 ATc graph at low concentrations of formaldehyde are due to negative values for luminescence at those time-points

Despite the improvement derived from removing auto-regulation, we sought better separation of signal response at lower formaldehyde concentrations more relevant in cells engineered for methanol assimilation (1–40μM in wild-type cells)[98]. Literature precedent in the RncR repressor protein of *E.coli* suggested that repeated cytosine and guanine tracts can induce A-form DNA geometry, likely playing an important role in DNA–protein interactions at binding sites at or around these tracts[115]. In the native P_{frm} sequence, two repeated stretches of guanine or cytosine nucleotides have been implicated as a contributing factor to the DNA geometry around the position that would typically bind to recruit the σ^{70} portion of the *E. coli* RNA polymerase to initiate transcription of downstream genes[109]. Changing two cytosine nucleotides in one tract (construct pTR47m4) allowed us to observe higher signal at lower formaldehyde doses compared to the wild-type binding sequence (construct pTR47), reducing the K_a from 37±1.5 to 11±0.9μM, and increasing the Hill coefficient from 2.7±0.2 to 5.0±2.5 (**Figure 23**). To ensure formaldehyde specificity, the reporter was tested with acetaldehyde, propionaldehyde, and methanol. None of these substrates elicited luciferase expression below mM concentration, and are, therefore, not significant activators of FrmR at physiological concentrations (**Figure 24**).

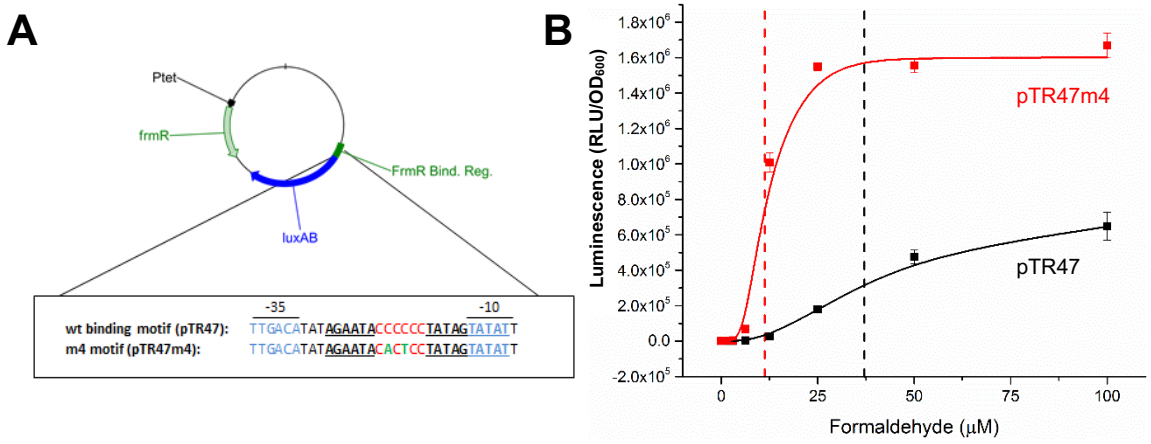


Figure 23: Rational promoter engineering increases sensitivity and dynamic range.
A. Schematic showing the specific differences in the pTR47 and pTR47m4 constructs. Nucleotides in the binding sequence that are underlined show predicted FrmR binding regions. Blue nucleotides are putative σ70 binding sequences. Nucleotides in purple represent the cytosine repeat tract, and red nucleotides show the mutations introduced in pTR47m4. **B.** Comparison of the formaldehyde-based response of cells containing the wildtype (pTR47, black) and mutant (pTR47m4, red) binding sites in the de-coupled circuit architecture. Error bars show the standard deviation across three biological replicates. RLU, relative luminescence units; OD_{600}, Optical Density at 600nm.

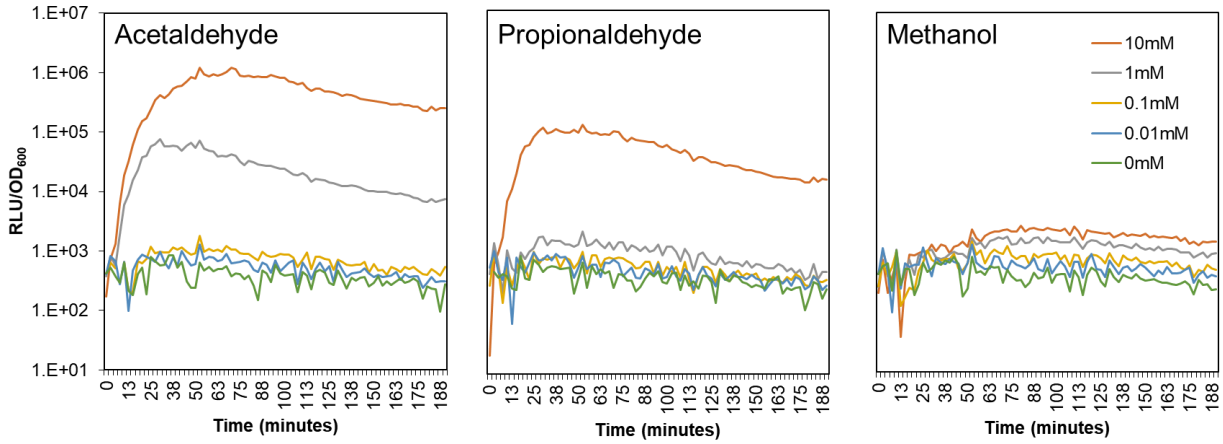


Figure 24: Comparison of signal from the mutant (pTR47m4) binding site based on dosing of various substrates. Lines show the average of three biological replicates. RLU, Relative Luminescence Units, OD600, Optical Density at 600nm.

The mutations tested in our m4 variant were examined individually by Rohlhill et al.[113]; however, none of the variants examined in that work had mutations at both positions within the C-tract without otherwise completely removing the tract. Our results suggest that this repeated C-tract may represent a sequence feature in which the type and combination of mutated bases can synergistically affect the responsiveness of the promoter/repressor system. To confirm this, we tested an additional range of single and double mutations to this C-tract as shown in **Figure 25.** From these data, we confirmed that our original m4 sequence was as good as or better than equivalent mutations to this region. Given that Rohlhill *et al*. originally assayed for promoter activities at 0.1mM formaldehyde[113], it's possible that this tract did not show significant importance in response. However, it seems clear that for lower doses of formaldehyde, changes to this tract can account for roughly 100-fold differences in relative luminescence.

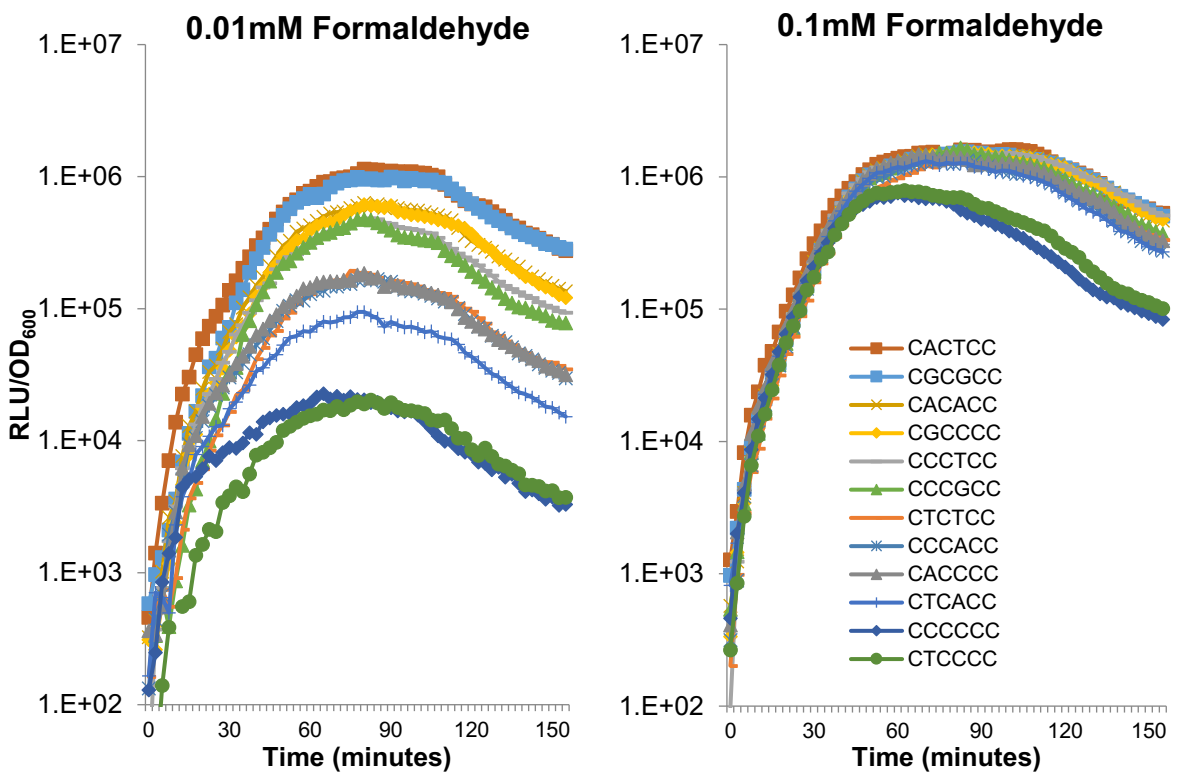


Figure 25: Comparison of mutations within the C-rich tract of the FrmR binding sequence.
RLU, relative luminescence units. OD_{600}, optical density measured at 600nm.

Because quantification based on fluorescent proteins is more widely used in biosensor applications, we additionally constructed a GFP version of the biosensor encoding the superfolder GFP[116] in place of luciferase in pTR47m4 (construct pTR47m4-gfp). This construct showed a similar response to formaldehyde, but with a lower signal-to-noise ratio and higher limit of detection (LOD) of 10μM (**Figure 26**). The Hill coefficient was reduced to 1.8±1.5, and the K_a increased to 21±11μM. We further assessed this construct in the context of a Δ*frmA* strain to see if removal of the detoxification system would improve sensitivity. In this strain, the reporter was much more sensitive at low formaldehyde concentration, with a Ka of 6.3±2.8μM, resulting in a reduced LOD of 5μM. Due to the higher baseline promoter activity in this strain, the overall dynamic range did not increase substantially. Interestingly, there was also considerably higher variability in the signal in the Δ*frmA* background, possibly because the absence of a detoxification system amplifies any initial noise in the formaldehyde spike in that strain. Prior work in synthetic methylotrophy has shown

the kinetic properties of Mdh to be much less favorable than those of the formaldehyde assimilation enzymes Hps and Phi, which has led to a search for catalytically more active homologs[99,100].

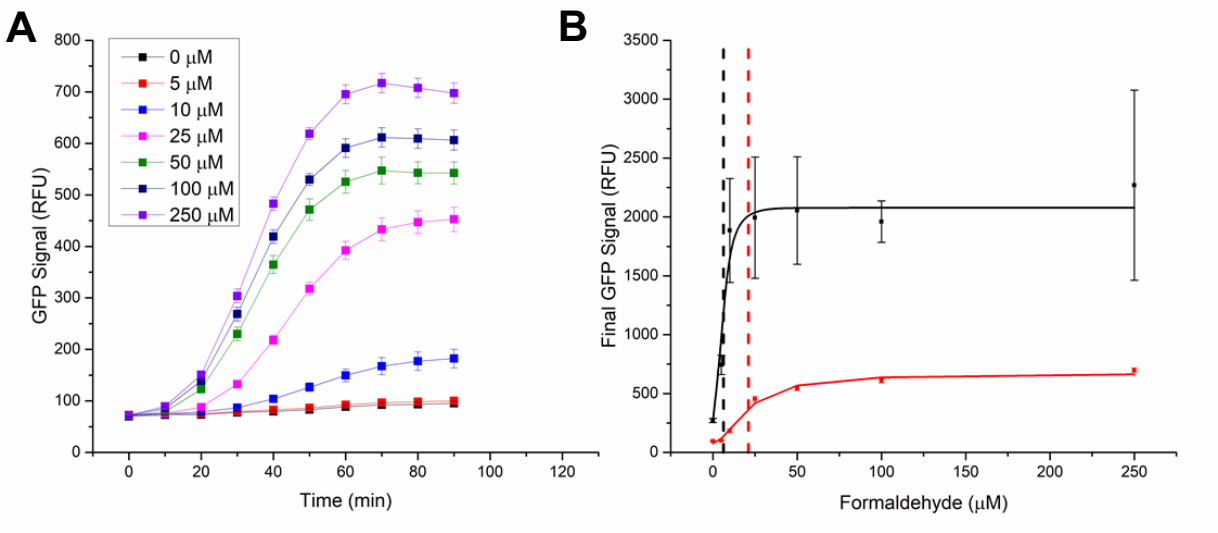


Figure 26: Evaluation of a GFP reporter variant in WT and Δ*frmA* Strains.
A: Time course of the dose-dependent response to formaldehyde in WT cells. **B:** Fitting of a Hill plot to the final GFP signal in WT (red) or Δ*frmA* (black) background. Dashed vertical line represents the Ka of the reporter (21 or 6.2μM). Error bars represent standard deviation of three biological replicates. RFU, raw fluorescence signal (AU).

Section 2.3: Monitoring Mdh activity using the FrmR reporter system

The *in vivo* comparison of different Mdh candidates has typically involved a labor-intensive Nash assay, which requires pelleting and washing cultures in order to chemically measure formaldehyde concentrations[98,99]. The experiment is conducted in a Δ*frmA* strain, otherwise detoxification lowers the formaldehyde concentration to virtually undetectable levels. This process is low-throughput and does not lend itself to the simultaneous comparison of many variants. In their efforts to evolve Mdh, Wu and coworkers developed a 96-well assay where the Nash reagent is added directly to the culture upon methanol addition. However, this procedure still required separation of the supernatant[100]. Rapid alternatives to assay *in vivo* Mdh activity with minimal user intervention would, therefore, be valuable. To test whether our formaldehyde reporter could meet this need, we cloned the gene encoding the evolved variant of Mdh2 from *Cupriavidus necator*[100] into pTR48 and transformed this plasmid into S1030 and S1030 Δ*frmA* cells containing pTR47m4-gfp. **Figure 27** shows the time course of GFP expression after simultaneous addition of methanol

and L-arabinose (to induce *mdh* expression). A clear methanol-dependent signal was observed in both strain backgrounds as early as 30min into the assay, and the response saturated after 2hr. These results clearly demonstrate the ability of the sensor to rapidly detect Mdh activity, even without deletion of endogenous formaldehyde detoxification pathways. The dynamic range, defined as the ratio of final GFP signal to initial signal, was higher in the WT cells (10.4±1.4 vs. 4.0±0.7), primarily due to the high background in the Δ*frmA* strain, therefore, we chose the WT background for further studies. In a parallel experiment with the same cells, we also measured formaldehyde concentration in the media using the Nash assay, to compare the sensitivity of the two methods (**Figure 27**). While methanol-dependent formaldehyde production was clearly measurable in both strains, there was only a 1.6-fold increase in concentration over the same time course, highlighting the superiority of the biosensor for *in vivo* measurement of Mdh activity. Given the similarity in LOD for both the Nash assay and our reporter with spiked formaldehyde, the increased dynamic range is most likely due to the amplification of the signal afforded by continual GFP expression in the presence of formaldehyde.

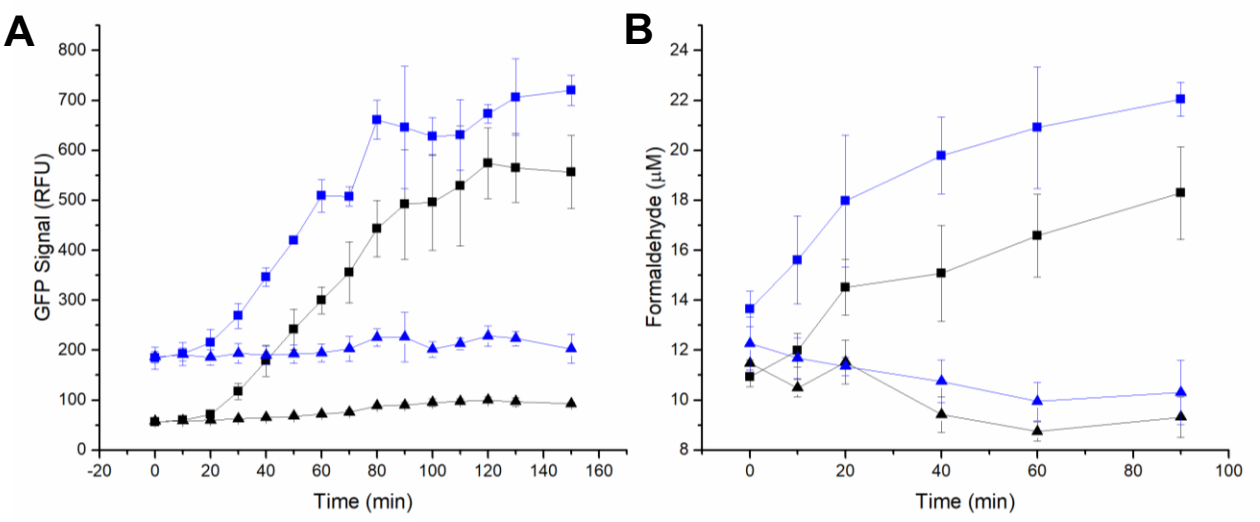


Figure 27: GFP biosensor enables sensitive Mdh activity measurements.
A. Timecourse of GFP expression from reporter pTR47m4-gfp in response to formaldehyde generated *in vivo* by cells expressing Mdh. S1030 WT (black) or Δ*frmA* (blue) cells expressing the evolved Mdh2 from *C. necator* (pTR48cnmdh4-1) were incubated with or without 500mM methanol (squares and triangles, respectively), and fluorescence was monitored for 150min. **B.** Time course of formaldehyde production in the supernatant of the same cells, measured by the Nash assay. Error bars represent standard deviation of three technical replicates**.** RFU, Raw Florescence Signal (AU).

With a functional assay in hand, we set about comparing several candidate Mdhs. The *mdh1* and *mdh2* genes from *B. methanolicus* were chosen given that the first is the major isoenzyme expressed during methylotrophic growth[117,118] and the second is the most active *B. methanolicus* Mdh based on expression in *E.coli*[98]. To assess the importance of the activator protein (Act), we generated a construct that co-expressed *mdh2* and *act*. This protein is a Nudix hydrolase which, at least *in vitro*, hydrolyzes the nicotinamide mononucleotide moiety of the NADH cofactor of Mdh, leading to a drastic reduction in the K_m for methanol[119–121]. We also included the alcohol dehydrogenase *adhA* from *Corynebacterium glutamicum*, which has a very low reported K_m[122] (Kotrbova-Kozak et al., 2007), as well as the *mdh* from *Geobacillus stearothermophilus*[99], and both the WT and evolved variants of *mdh2* from Cupriavidus necator[100]. After 2hr, signal was detected for all variants except Mdh1 (**Figure 28**). The strongest signal, from the evolved variant of *C. necator* Mdh2, showed over a 20-fold increase in OD-normalized GFP signal compared to an mCherry control. In general, the differences in fluorescence intensity between variants matched trends in previous reports: Mdh2 from *B. methanolicus* outperforms Mdh1 (30 vs. 1.7mUmg^{-1}), and co-expression of Act in *E. coli* does not change the *in vivo* activity[98]. This is reflected in our data by the nine-fold difference in signal between Mdh2 and Mdh1 at the highest methanol concentration, and the absence of any difference between Mdh2 and Mdh2+Act.

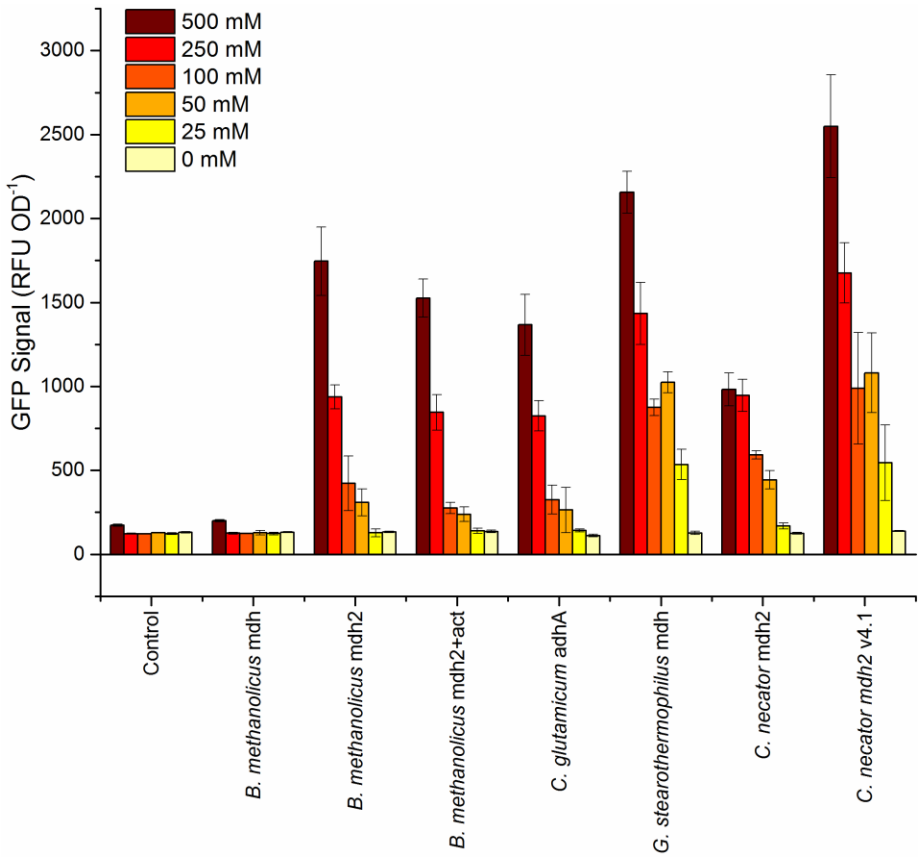


Figure 28: *In vivo* Comparison of Mdh variants using GFP formaldehyde biosensor.
GFP fluorescence (RFU) measured 2hr after induction of *mdh* expression and addition of various concentrations of methanol and normalized to culture density (OD, optical density measured at 600nm). Error bars represent standard deviations for three individual colonies. The control is cells carrying pTR48mCherry.

It should be noted that although the Mdh1 strain showed no GFP signal after 2hr, fluorescent signal above the control could be seen upon further incubation for a total of 8hr (**Figure 29**), long after signals from the other variants had reached saturation, reflecting the extremely slow but still detectable rate of methanol oxidation. Whitaker et al. showed that the Mdh from *G. stearothermophilus* outperforms the Mdh2 from *B. methanolicus* at low concentrations (60mM) of methanol[99]. This is also reflected in our data, where at 50mM methanol the former shows a fluorescent intensity 2.8-fold higher than the latter. As expected, the evolved variant of the C. necator enzyme outperformed the WT version. Given the reportedly low K_m of AdhA from *C. glutamicum*, we were surprised to see no improvement over Mdh2 from *B. methanolicus*, which supports more recently reported K_m data for this enzyme[100]. Our assay identified the evolved

variant from *C. necator* and the Mdh from *G. stearothermophilus* as the most promising candidates for engineering synthetic methylotrophy, but could not distinguish between the two.

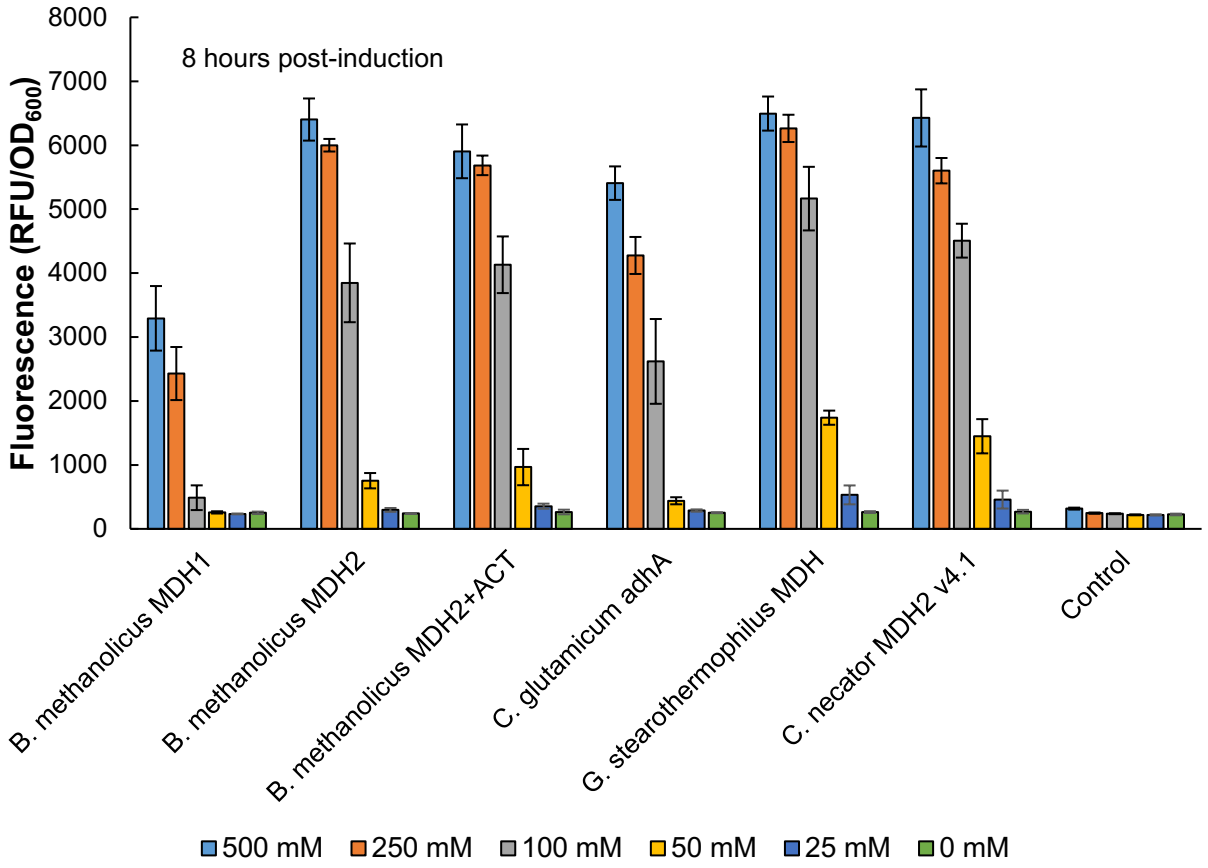


Figure 29: Mdh Comparison after Eight Hours.
Fluorescence was measured 8 hours after the addition of methanol in an experiment analogous to that presented in Figure 28. Error bars represent the standard deviation of three independent colonies.

Taken together, the data presented here show that our assay can be used to detect *in vivo* formaldehyde production by Mdh in a WT background and faithfully reconfirms trends reported in the previous literature. From inoculation to assay completion, the procedure requires a total of 5hr, and only one intervention to add the inducer and methanol. This a significant simplification over the current state of the art and should enable comparison of different Mdh candidates in a high-throughput manner. It should be pointed out that, since Mdh expression was induced at the same time methanol was added, the comparisons presented here between the different Mdh variants also include variations in their expression and folding kinetics.

Section 2.4: Application of the FrmR sensor to cell sorting

Having demonstrated that the formaldehyde biosensor could be used to discriminate between Mdh candidates, we were interested in evaluating the potential of the sensor for the directed evolution of novel variants using high-throughput technologies such as FACS or PACE[7]. In these approaches, the assay must be able to discriminate between candidates within a mixed population. Since formaldehyde diffuses rapidly across the cell membrane, we were concerned that the lack of spatial segregation of high- and low-activity mutants in a library could lead to the enrichment of cheaters, where the formaldehyde produced by a high activity variant could diffuse into a cell with a low-activity variant and activate the reporter. To assess this possibility, we co-inoculated a culture with two strains: one carrying pTR48mdh2, and one carrying pTR48mCherry, which expresses mCherry instead of *mdh2*. Both strains contained the reporter plasmid pTR47m4GFP, but only the one with *mdh2* should be able to produce formaldehyde from methanol and activate the reporter. If formaldehyde diffusion can activate the reporter in cells not expressing *mdh*, we would expect to see GFP signal in cells expressing mCherry after treatment with methanol. If not, the mCherry cells should show no GFP fluorescence. Cells were grown under the same conditions as before, and the population was analyzed by flow cytometry 2hr after induction and methanol addition. As shown in **Figure 30**, the mCherry+ cells showed mean GFP fluorescence similar to the mCherry−cells, indicating that formaldehyde produced in one cell was able to activate the reporter in another. To prevent this cross-talk, we devised two strategies to reduce the extracellular formaldehyde concentration: (1) addition of NAD-dependent formaldehyde dehydrogenase (FaDH) from *Pseudomonas sp.* and NAD+, to enzymatically oxidize the formaldehyde to formate and (2) addition of glutathione as a formaldehyde scavenger. Addition of the FaDH and NAD+ was unsuccessful, possibly due to the breakdown of the enzyme in the supernatant or unfavorable reaction conditions.

In contrast, addition of increasing concentrations of glutathione, from 2 to 10mM, increased the difference in GFP fluorescence between the mCherry+ and mCherry− cells (**Figure 30**). At 10mM, the mCherry+ cells showed only background levels of GFP, indicating complete sequestration of formaldehyde in the supernatant as the glutathione adduct. The addition of glutathione negatively impacted cell growth, with the OD at analysis of the 10mM culture roughly half that of the 0mM control. Despite the growth defect, the GFP fluorescence was roughly the same, indicating that the glutathione had no impact on the activation of the reporter or the activity of Mdh. These results clearly show the potential for cheaters in this assay and establish that glutathione should be added as a formaldehyde sink if this assay is employed for directed evolution of Mdh activity.

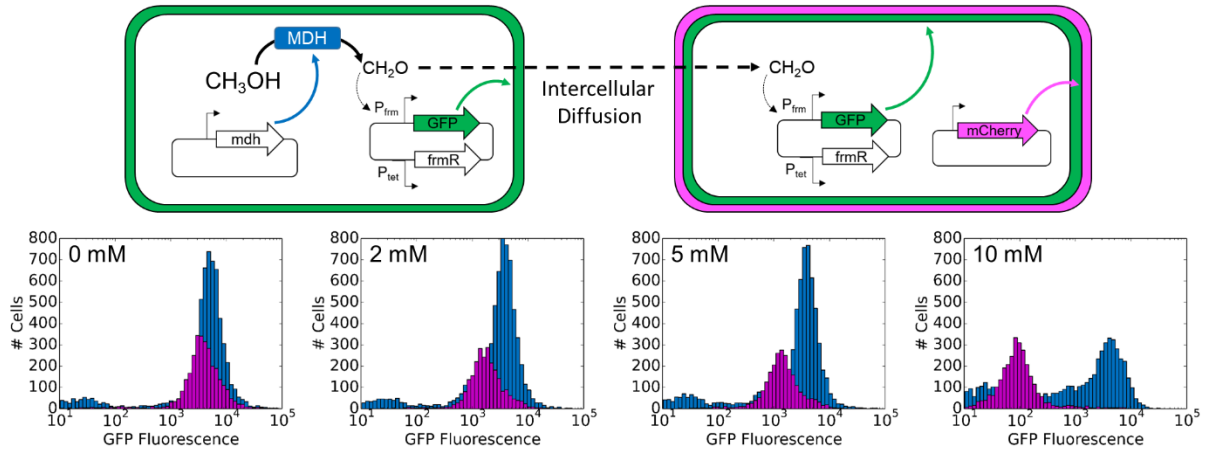


Figure 30: Analysis and elimination of cheaters in a mixed population.
Top: Schematic showing potential false activation of reporter in cells with no *mdh* (left) by diffusion from cells with active *mdh* (right) in a mixed culture. Cells that are only GFP+ are true formaldehyde producers, whereas cells that are both mCherry+ and GFP+ are "cheaters" activated by diffusion. **Bottom:** Analysis of GFP fluorescence in *mdh*-containing cells (blue bars) and mCherry-containing cells (pink bars) by FC with varying concentrations of glutathione added as an extracellular formaldehyde sink to prevent diffusion.

Section 2.5: Monitoring formaldehyde to optimize methanol assimilation in *E. coli*

As formaldehyde is highly toxic, strains engineered to metabolize methanol must carefully balance the production of formaldehyde by Mdh with its consumption by Hps and Phi. This balance can be engineered at the transcriptional level by varying the induction of the upstream and downstream pathways, and assessing the formaldehyde concentration. Here we used the

formaldehyde reporter to facilitate high-throughput testing of different induction concentrations. The evolved *mdh2* from *C. necator* was placed under the control of the T7 promoter, and *hps* and *phi* as an operon under control of the P_{tet} promoter (construct pETMEOH560). Formaldehyde was measured using the biosensor plasmid pTR59gfp, which is derived from pTR47m4 but with constitutive expression of *frmR* to avoid cross-talk with the tet promoter in the methanol plasmid. Cells were grown on xylose to provide non-limiting levels of Ru5P. Without induction of either *mdh* or *hps*/*phi*, only background GFP signal was detected (**Figure 31**). Upon induction with IPTG, the GFP signal rose to a maximum 14-fold increase at 50μM, before falling at higher induction levels. This fall in signal is likely attributable to the low solubility of Mdh at high induction levels, which leads to slower growth and reduced GFP synthesis (**Figure 32**). At 25μM IPTG, low-level induction of *hps*/*phi* with 10ngml^{-1} is sufficient to reduce the formaldehyde concentration to background levels. At 50μM IPTG, the higher Mdh activity requires an increase in Hps/Phi induction to compensate. At 100μM IPTG and above, even full induction of the downstream pathway with 200ngml^{-1} ATc was insufficient to fully consume formaldehyde. This insufficiency may arise because high level IPTG induction reduces the expression level from the P_{tet} promoter,as shown in experiments where IPTG and ATc controlled GFP and RFP respectively (**Figure 33**), such that the Hps/Phi activity is too low to balance formaldehyde generation. Taken together, the results here demonstrate the high sensitivity of the reporter through its ability to detect formaldehyde even in strains engineered for its consumption, and without deletion of the endogenous detoxification system. They also demonstrate how the reporter can be used to optimize the relative expression level of *mdh* and *hps*/*phi* to maximize flux while minimizing expression burden.

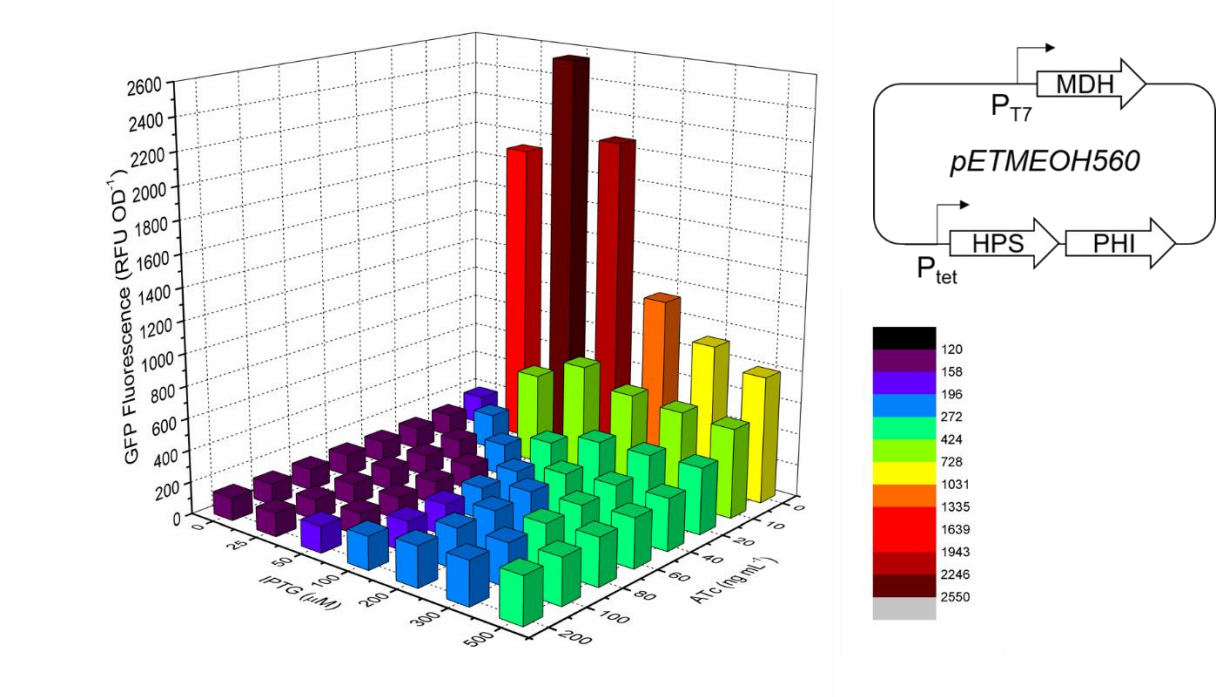


Figure 31: High-throughput pathway balancing monitored by the formaldehyde reporter.
Normalized GFP fluorescence in cells carrying pETMEOH560 2hr after induction with various concentrations of ATc and IPTG, and treatment with 500mM methanol. IPTG induction controls *mdh* expression, while ATc induction controls *hps* and *phi* (top right). Bars represent means of three biological replicates. RFU, raw fluorescence units; OD, optical density at 600nm.

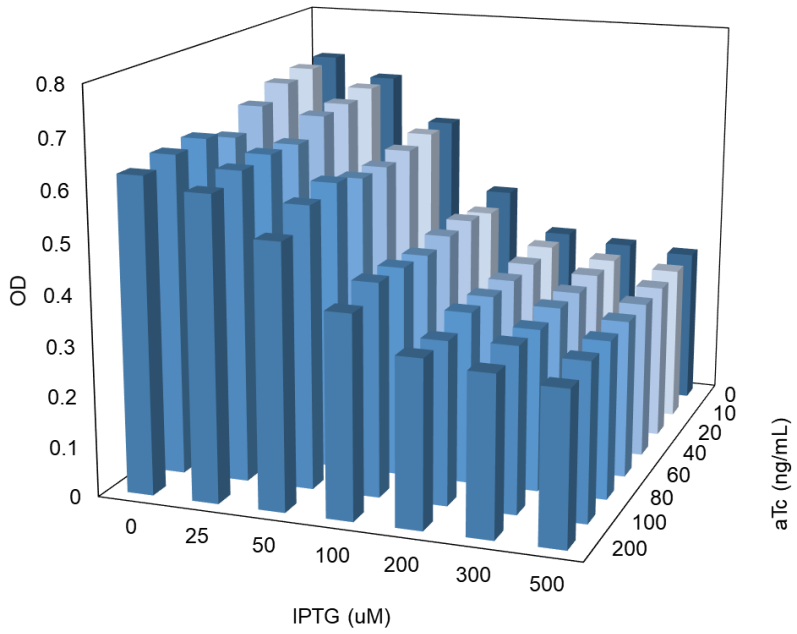


Figure 32: Effect of Induction on Growth.
Cells carrying pETMEOH560 were induced with varying concentrations of IPTG and ATc, and growth assessed after 2 hours by OD measurement. Bars represent the mean of three biological replicates.

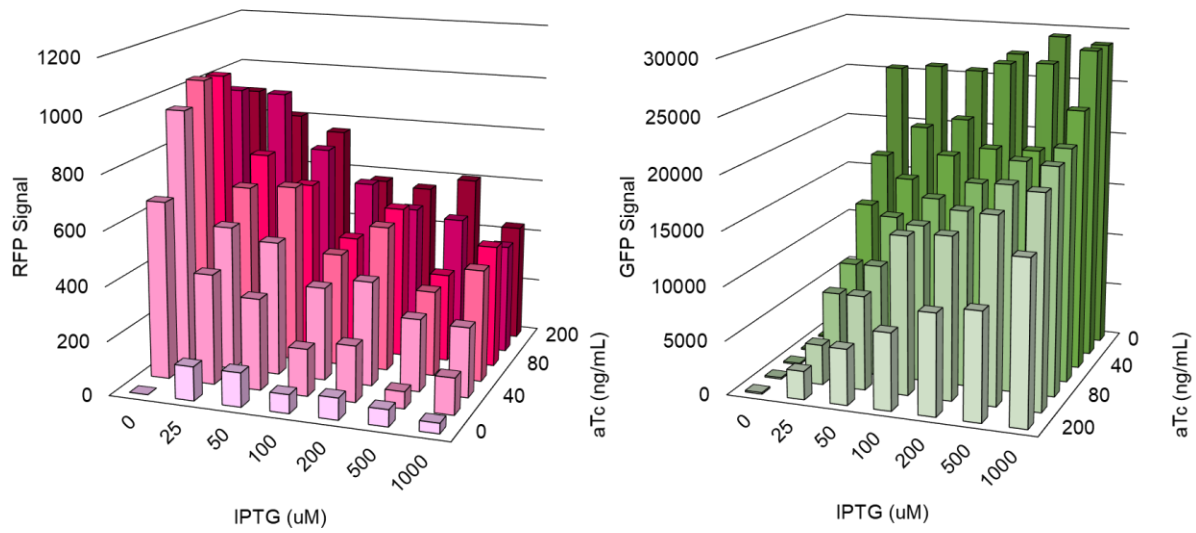


Figure 33: Inducer Cross-Talk in Cells Carrying pET-sGFP-tet-RFP.
Cells carrying pET-sGFP-tet-RFP were induced with varying concentrations of IPTG and ATc, and RFP fluorescence (left) and GFP fluorescence (right) measured after 2 hours. Bars represent the average of three biological replicates.

Section 2.6: Discussion

In this work, we made significant improvements to a formaldehyde biosensor, and demonstrated its utility in ongoing efforts to establish synthetic methylotrophy. Engineering of the promoter binding site and regulatory architecture led to dramatically improved sensitivity over the previous version. Our luciferase-based system could detect the addition of 1μM exogenous formaldehyde with a 10-fold increase in signal. That we were able to achieve this substantial improvement to sensitivity with a rational approach reinforces the continued utility of rational design in promoter engineering. We further showed our reporter could discriminate between Mdh candidates, and with the addition of glutathione as a formaldehyde sink could detect differences within a mixed population. In addition, we showed that the dynamic range of the sensor matches well with physiological formaldehyde concentrations in strains engineered for methanol assimilation, allowing high-throughput optimization of induction conditions for the upstream and downstream pathway. These proof-of-concept experiments demonstrate the utility of the biosensor, and pave the way for directed evolution of Mdh, as well as the implementation of

dynamic control strategies in synthetic methylotrophy. Finally, levels of formaldehyde in drinking water are routinely 50μgL^{-1} (1.7μM)[123], therefore, this assay has the additional potential beyond the synthetic biology applications to be used in environmental sensing and waste management.

Section 2.7: Materials and methods

Reagents

Unless specified, all chemical reagents were purchased in the highest grade available from Sigma–Aldrich. For luciferase reporter assays, M9 salts and LB, Agar, and casamino acids were purchased from US Biological Life Sciences. For all other assays, DIFCO M9 salts and LB, BACTO Agar, and casamino acids were purchased from Becton Dickinson. Methanol-free formaldehyde (16% v/v) was purchased as 1ml ampules from Thermo Fisher Scientific, and dilutions for cellular assays were prepared fresh daily. Trace Elements (MD-TMS) and Vitamin Solution(MD-VS) were purchased from ATCC.

Strains and plasmids

E. coli DH5α was used as a cloning host. Luciferase-based assays and Mdh comparisons were carried out using E. coli S1030[8] with or without *frm*A knocked out, as described in the text. Pathway optimization assays were conducted in *E. coli* MG1655(DE3)[124]. pBbS2k-RFP was a gift from Jay Keasling (Addgene plasmid #35330). pETM6-mCherry was a gift from Mattheos Koffas (Addgene plasmid #66534).

Cloning

Plasmids were constructed using USER cloning, Gibson Assembly Master Mix, or KLD Enzyme Mix (NEB). DNA fragments were generated by PCR with the primers listed in Supplemental

Table S2. Phusion U Hot Start DNA Polymerase (ThermoFisher) was used with all USER cloning primers, whereas Q5Polymerase (NEB) was used for all other primers. All amplicons were digested with DpnI before PCR purification for Gibson Assemblies and KLD ligations, or during PCR fragment assembly for USER cloning. All restriction endonucleases were purchased from NEB. Assembled vectors were transformed into either DH5α chemically competent cells (NEB) or Mach1 T1R chemically competent cells (ThermoFisher) and verified by Sanger sequencing. The *frmA* deletion in S1030 was made using the protocol described by Datsenko and Wanner[125]. Cell growth for cloning purposes was carried out using DIFCO LB media supplemented with appropriate antibiotics (Kanamycin, 50μgml^{-1}; Carbenicillin, 50μgml^{-1}; Spectinomycin, 50-100μgml^{-1}).

Luciferase-based reporter assays

For all reporter assays, a minimum of three individual colonies from transformation plates were grown overnight in M9+ medium (M9 supplemented with 0.2% glucose, 0.1% casamino acids, and 1% (v/v) each of trace mineral and vitamin solution). Cultures were diluted 100-fold into fresh medium and grown until early exponential phase (OD approximately 0.4) supplemented with anhydrotetracycline (ATc) to induce *frmR* expression. For direct testing of induction, formaldehyde was added to 200μl of sample in a black, Corning black clear bottom 96 well plate and analyzed for optical density at 600nm and luciferase activity at 37°C on an Infinite Pro M1000 plate reader (Tecan).

GFP-based formaldehyde reporter and MDH assays

GFP reporter assays with formaldehyde were conducted almost identically to the luciferase-based ones, but cell growth and fluorescence were measured on a SpectraMax M2e (Molecular Devices) spectrophotometer at 37°C with continuous shaking. For Mdh-linked assays, cells were induced with 10mM arabinose and immediately transferred to 96-well plates (200 μl per well),

where 50µl methanol was added to a final concentration of 0–500mM. Plates were covered with Breathe-Easy sealing membranes (Sigma), and growth continued at 37°C with maximum agitation on a Jitterbug 2.0 (Boekel Scientific). Pathway optimization experiments were conducted analogously, except that induction was mediated by IPTG and ATc. GFP fluorescence was assessed hourly to minimize temperature fluctuations that would affect growth and Mdh activity. Excitation and emission wavelengths were 488 and 525nm, respectively.

Formaldehyde measurement by Nash assay

Formaldehyde concentration in culture supernatants was assayed by a modification of the Nash reaction[97] for 96-well plate format. A total of 200µl of cells were centrifuged for 1min at 13,000 RPM, and 125µl supernatant transferred to the plate. A total of 125µl Nash reagent (5M ammonium acetate, 50mM acetylacetone) was added to each well, the plate was incubated at 37°C for 1hr, and absorbance was read at 412nm. A standard curve was prepared in culture medium in the range from 100 to 0µM. The plate was kept on ice until all samples had been collected.

Flow cytometry

Cells from GFP assays were diluted 500-fold into PBS buffer for FC analysis, which was performed using a BD FACS LSRII HTS-2. GFP fluorescence was measured using the 488nm laser and 530/30 filter, and mCherry fluorescence was measured using the 561nm laser and 575/26 bandpass filter. Cells were gated based on FSC-H and SSC-H, and10,000 events falling into this window were recorded. GFP versus mCherry plots were converted into tsv files using the freeware software Cyflogic and analyzed and plotted using in-house Python scripts.

Model regression

Response curves from both luciferase and GFP assays were fitted to the Hill equation, and confidence intervals around the model parameters (Smax, n, and Ka) were calculated using the MATLAB functions nlinfit() and nlparci(), respectively. The form of the Hill equation used was

$$\boldsymbol{S = S_{min} + (S_{max} - S_{min})(\frac{[F]^n}{K_a^n + [F]^n})}$$

where S is the measured signal, [F] is the concentration of formaldehyde, S_{max} the signal at saturation, S_{min} the signal at 0 formaldehyde, *n* the Hill coefficient, and K_a the dissociation constant.

Chapter 3: Phage-Assisted Evolution of Methanol Dehydrogenases

In Collaboration with Ben Woolston and Greg Stephanopoulos

Section 3.1: Introduction to methanol dehydrogenase applications and engineering

In ongoing efforts to better use non-renewable resources and reduce greenhouse gas emissions, methane gas from petrochemical production and manure decomposition has remained a constant problem[126,127]. Methane is expensive to transport and contributes significantly to global warming, making a cheap method for removing excess supply the wasteful combustion of natural gas into the less potent greenhouse gas carbon dioxide[128]. A preferable alternative would be to apply chemical and microbial engineering strategies to fix this single-carbon resource into higher molecular weight products through biorefining[129]. Specialized microbial cultures can oxidize methane to methanol, providing a feedstock for further fermentations, but engineering these conversions is a significant challenge[129–131]. Efficiently engineering microbes to ferment methanol is thus an attractive stepping-stone toward full bioconversion of methane into commodity chemicals[132]. A variety of methylotrophic microbes, so named for their ability to grow on methanol as their sole carbon source, typically assimilate methanol into central metabolism via oxidation of methanol into formaldehyde by one of a variety of methanol dehydrogenase (Mdh) enzymes. In gram-negative bacteria such as *Pseudomonas sp.* M27, oxidation of methanol is coupled to reduction of pyrroloquinoline quinone (PQQ) using cytochrome-dependent enzymes in the periplasm[133]. Methylotrophic yeasts such as *Pichia pastoris* can use molecular oxygen directly as an electron acceptor, generating peroxide as a byproduct[90]. In gram-positive bacteria such as *Bacillus methanolicus*, NAD-cofactors are used to oxidize methanol in the cytoplasm[121]. The resultant formaldehyde is then typically funneled into one of two assimilation pathways: the ribulose monophosphate (RuMP) pathway[134] or the serine cycle[135].

Methylotrophy in nature is typically inefficient and engineering these native methylotrophs directly is a significant challenge[91,136]. A promising alternative is to develop synthetic methylotrophy in model organisms not normally capable of growing on methanol, most notably *Escherichia coli*[94]. To engineer synthetic methylotrophy in *E. coli*, the RuMP pathway is typically the

pathway of choice given its better theoretical yield and relative compatibility with *E. coli* metabolism, requiring only three non-native enzymes to complete a methanol assimilation cycle[134]. Expressing an NAD-dependent methanol or alcohol dehydrogenase along with hexulose-6-phosphate synthase (Hps) and 6-phospho-3-hexuloisomerase (Phi) from native methylotrophs is sufficient to show incorporation of methanol feedstocks into *E. coli* central metabolism[98], but still fall far short of enabling *E. coli* to use methanol as an exclusive carbon source. A proposed reason for this inefficiency is the methanol dehydrogenases themselves, which are notoriously inefficient given their millimolar K_m values and preference for oxidizing other alcohols[99,100,121]. As a result, identifying and engineering improved NAD(P)-dependent Mdh variants has become a key component of developing synthetic methylotrophy in *E. coli*.

A variety of these genes have been isolated, characterized, and, in some cases, even evolved. Of particular note is the *mdh2* gene from *Cupriavidus necator*, whose activity was significantly improved using a combination of *in vitro* library generation, automated colony picking, and a plate-based screen for formaldehyde production[100]. Based on reported *in vitro* kinetic parameters, this evolved Mdh2 CT4-1 enzyme represents the state-of-the-art for methanol assimilation pathways in *E. coli*. Other researchers have taken a different approach by scaffolding downstream assimilation enzymes to *Bacillus methanolicus* Mdh1 decamers to improve the efficiency of formaldehyde flux from the Mdh through Hps and Phi[137]. Given the unpredictable behavior of these supramolecular assemblies for other Mdh variants, however, the most direct solution to the problem likely remains identifying Mdh enzymes that themselves exhibit enhance kinetic parameters. In the absence of naturally occurring highly-active Mdhs, improved enzyme evolution techniques represent the best approach toward achieving synthetic methylotrophy in *E. coli*. This chapter establishes the basis for an *in vivo* selection for improved Mdh activity using an optimized formaldehyde-sensor coupled to Phage Assisted Continuous Evolution and demonstrates significant improvements over the state-of-

the-art in terms of *in vivo* methanol assimilation. Based on the mutations we observe, we also propose a set of rational design options for future Mdh engineering efforts.

Section 3.2: Linking methanol oxidation to phage propagation

PACE requires variable gene expression based on the evolutionary target's activity of interest[7], which in this case is the oxidation of methanol into formaldehyde. We applied our previously engineered formaldehyde reporter system[138] to engineer an optimized Accessory Plasmid (AP) to regulate gene III expression based on formaldehyde production from evolving *mdh* libraries. To do this, we kept the same sensitized FrmR binding site developed for the pTR47m4 plasmid, but the P_{tet} promoter upstream of the *frmR* gene was swapped out for a constitutive promoter and gene III was placed upstream of the *xluxAB* luciferase reporter genes. In turn, we cloned two selection phage, replacing gene III with either the *mdh1* or the *mdh2* gene from *B. methanolicus* MGA3[121] (**Figure 34A**). Initial attempts to use PACE with this system resulted in either high background expression of gene III limiting cell infectibility or in high doses of methanol leading to non-specific phage propagation when added concurrent with phage to infectible cells. In early attempts to evolve selection phage encoding *B. methanolicus mdh* genes, we observed either washout (presumably from uninfectibility of host cells) or deletion of the *mdh* genes from the phage genome, showing the extent to which background FrmR-derepression from high methanol concentrations and individual, formaldehyde-producing SPs in the evolving population can negatively impact the overall experiment (**Figure 34B**). We addressed the uninfectibility issue first by reducing the strength of the ribosome binding sequence (RBS) driving gene III expression. We then prevented the second issue by including glutathione in the evolving population to act as an extracellular formaldehyde sink[138].

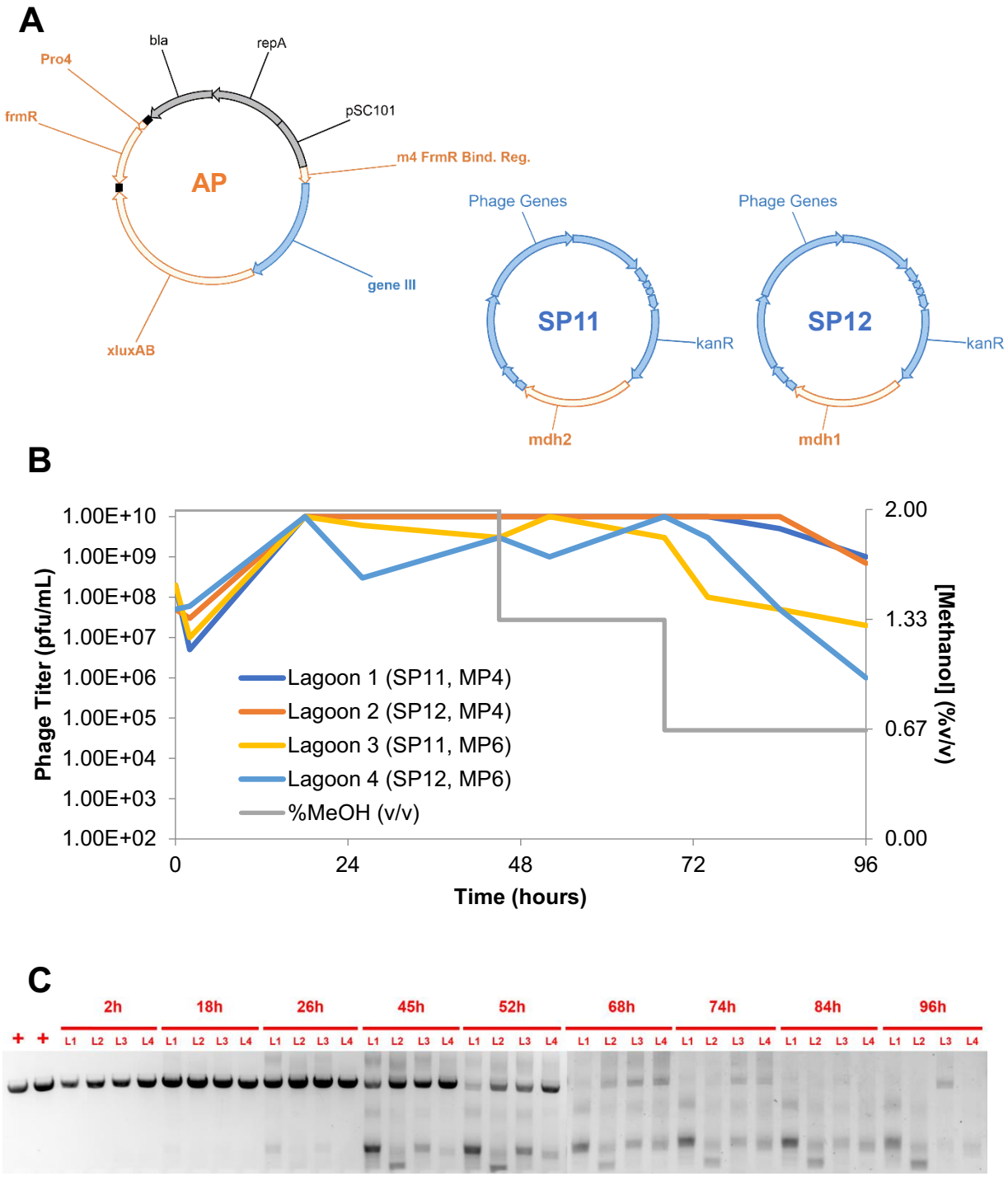


Figure 34: Propagation of Mdh SP in PACE without glutathione leads to loss of Mdh genes.
A. AP and SP combinations used in this experiment. Pro4 denotes a constitutive promoter characterized by Davis *et al.*[59] bla, antibiotic resistance marker; pSC101/repA, origin of replication. **B.** Titer data for four lagoons run at a constant flow rate of 1 lagoon volume per hour. MP4 and MP6 represent two strengths of mutagenesis plasmid as previously reported[67]. Methanol at 2% (v/v) is 500mM. pfu, plaque forming units. **C.** PCR using primers immediately upstream and downstream of the *mdh* gene on SP11 and SP12 generates DNA fragments shown in the two "+" lanes on the far left. For Lagoons 1 (L1), 2 (L2), 3 (L3) and 4 (L4), continuous evolution leads to loss of this fragment, implying gradual deletion of the *mdh* gene from the evolving phage population.

Section 3.3: Phage-Assisted Non-Continuous Evolution of *B. methanolicus* Mdh1 and Mdh2

Our AP optimizations and introduction of glutathione during phage propagation improved consistency in terms of host infectibility, but phage propagation rates were relatively low in the resultant conditions. This remained true even when we switched to S1030 host cells lacking the formaldehyde detoxification gene *frmA*, which should be sensitized to intracellular formaldehyde relative to strains expressing *frmA*. It was clear from these results that our optimized SP and AP combination was insufficient to permit evolution using continuous flow. Rather than spin our wheels trying to optimize these constructs further, we decided to alleviate this burden of propagation by using Phage Assisted Non-Continuous Evolution (PANCE), a technique which I had previously applied to the evolution of PHB production pathways. This system uses iterative rounds of overnight phage propagation in discrete cultures of host in place of a chemostat, allowing for more stable monitoring and maintenance of evolving phage populations at the cost of a much slower rate of evolution. To help maximize the rate of phage population expansion and subsequent generation of diversity through the mutagenesis plasmid (MP), we used the Drift Plasmid system[67] (DP6) to adjust selection stringency based on dosing of anhydrotetracycline (ATc) in the growth medium. The full set of selection interactions in the resultant AP/MP/SP combination is illustrated in **Figure 35**.

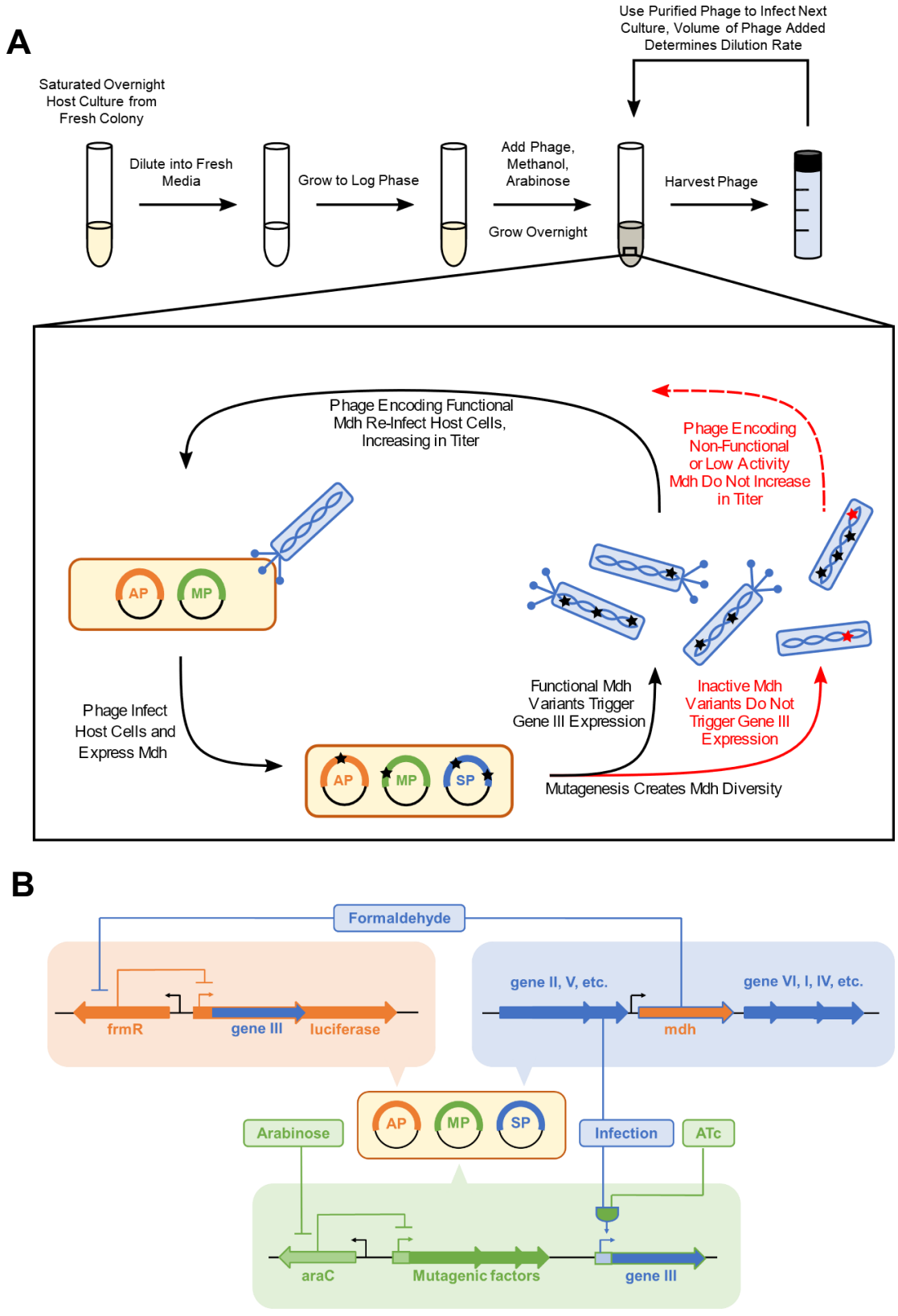


Figure 35: Phage-Assisted Non-Continuous Evolution of Methanol Dehydrogenases.
A. PANCE Selection Protocol for Mdh Evolution. **B.** Diagram of Host Cell showing the AP, MP, and SP characteristics for the selection.

Initially, we ran PANCE the way an equivalent PACE experiment would typically be run, with an extended period of drift at lower and lower stringencies (high to low doses of anhydrotetracycline (ATc)), followed by a period of selection without drift. Initial attempts to wean the selection off drift resulted in washout within a few passages, thus necessitating a selection regime in which we applied drift every other passage. As **Figure 36B** shows, this oscillation between selection and drift was sufficient to maintain functional methanol dehydrogenase-carrying phage populations while simultaneously failing to maintain a negative control phage carrying an enzyme with an unrelated function (*phaC* from *C. necator*[16]). We adjusted the volume of phage transferred between passages to mimic the flow-rate adjustments we would typically make during a PACE experiment, and we additionally adjusted stringency in later populations by lowering the concentration of methanol added to the evolving culture. During the early stages of this process, phage titers were monitored daily for each passage to ensure no washout occurred at each step. Unlike PACE, where washout typically requires that a continuous flow setup be re-constructed following technical failure and phage populations expanded in titer for subsequent re-infection, washout during a PANCE experiment simply requires a different volume of the most recent, high-titer passage be added to a fresh host culture. As a result, while PANCE is considerably slower than PACE, it allows far more room for error, which is particularly important when evolutionary targets have particularly low starting activities and potentially require many evolutions to improve their desired phenotype. Eventually, we settled into a consistent pattern of selection conditions for our evolving population and monitored phage titers every several passages. We continued this for 70 passages using a single evolving population as shown in **Figure 36B**.

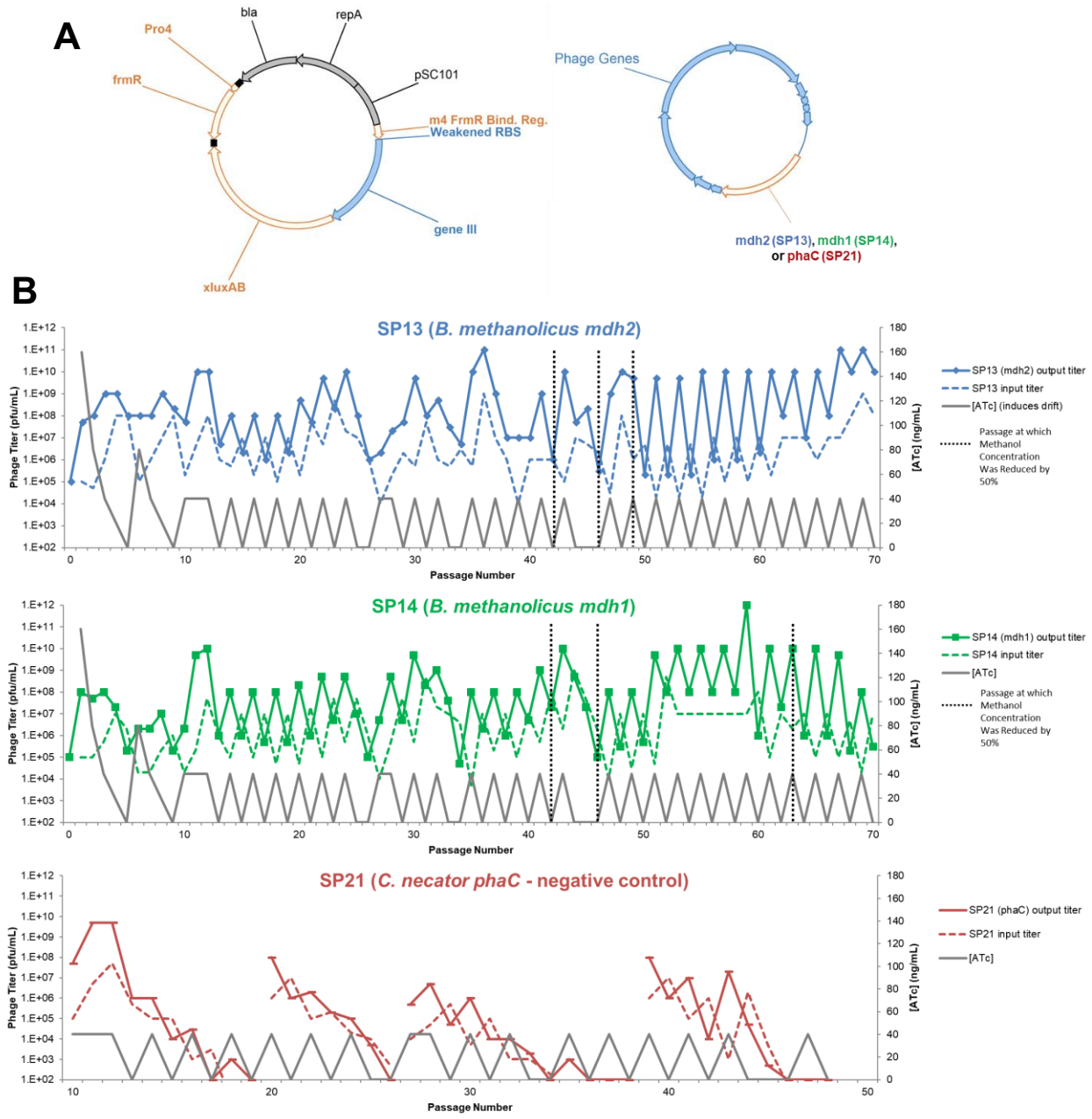


Figure 36: Selection schedules for PANCE of *B. methanolicus mdh1 and mdh2*.
A. Illustration of the AP and SP architecture used for the experiment. Host cells additionally contained DP6 as shown in Figure 35B. Apart from the weakened RBS, the AP shown on the left is identical to the one used in Figure 34A. **B.** Titers were taken for single cultures of each SP over 70 discrete passages. Input titers were calculated by multiplying the dilution rate for one passage by the previously calculated titer with the input titer for the first experiment determined by the titer of clonal starting SP used to infect the first passage. SP21 cultures were periodically evolved in the same host in parallel to SP13 and SP14, with breaks in the lines indicating re-infection with starting phage after confirmation of washout. All cultures were started at 2% (v/v) methanol, which is roughly 500mM. Methanol concentrations were halved starting with each indicated passage as marked by a dotted black line – by the end of the evolutions of SP13 and SP14, methanol concentrations were 0.25% (v/v) (62.5mM). pfu, plaque-forming units.

Sequencing data across this initial PANCE experiment showed convergence on several mutations for both target *Bm mdh* genes. For *Bm mdh2*, the most consistent observed mutation was M163V (**Figure 37**). This is adjacent to the A164 position homologous to the key mutated A169 residue in the evolution of *C. necator mdh2 CT4-1*[100,121]. For *Bm mdh1* three mutations clearly swept the population: the first two were N112H and G191S, which evolved almost immediately during the selection, and the third was P9Q (**Figure 37**). To make sure our selection results were having the desired effect on methanol oxidation phenotypes, we subcloned several genes from sequenced individuals at intermediate passages and assayed these variants using our luciferase reporter system (**Figure 38**). For *Bm mdh2*, it was apparent that M163V provided the bulk of increased apparent activity—consistent with its proximity to the key residue for homologous Mdh *C. necator* Mdh2. Results for *mdh1* were also promising; although none of its converged mutations had any known alignment with key residues in other characterized Mdhs, the apparent activity of the P9Q-containing triple mutant that swept the population by passage 45 was similar to activities we saw from the *Bm mdh2* population. Given that we had previously shown wild-type Bm Mdh2 to be significantly more active than Mdh1 *in vitro*, this was a significant improvement.

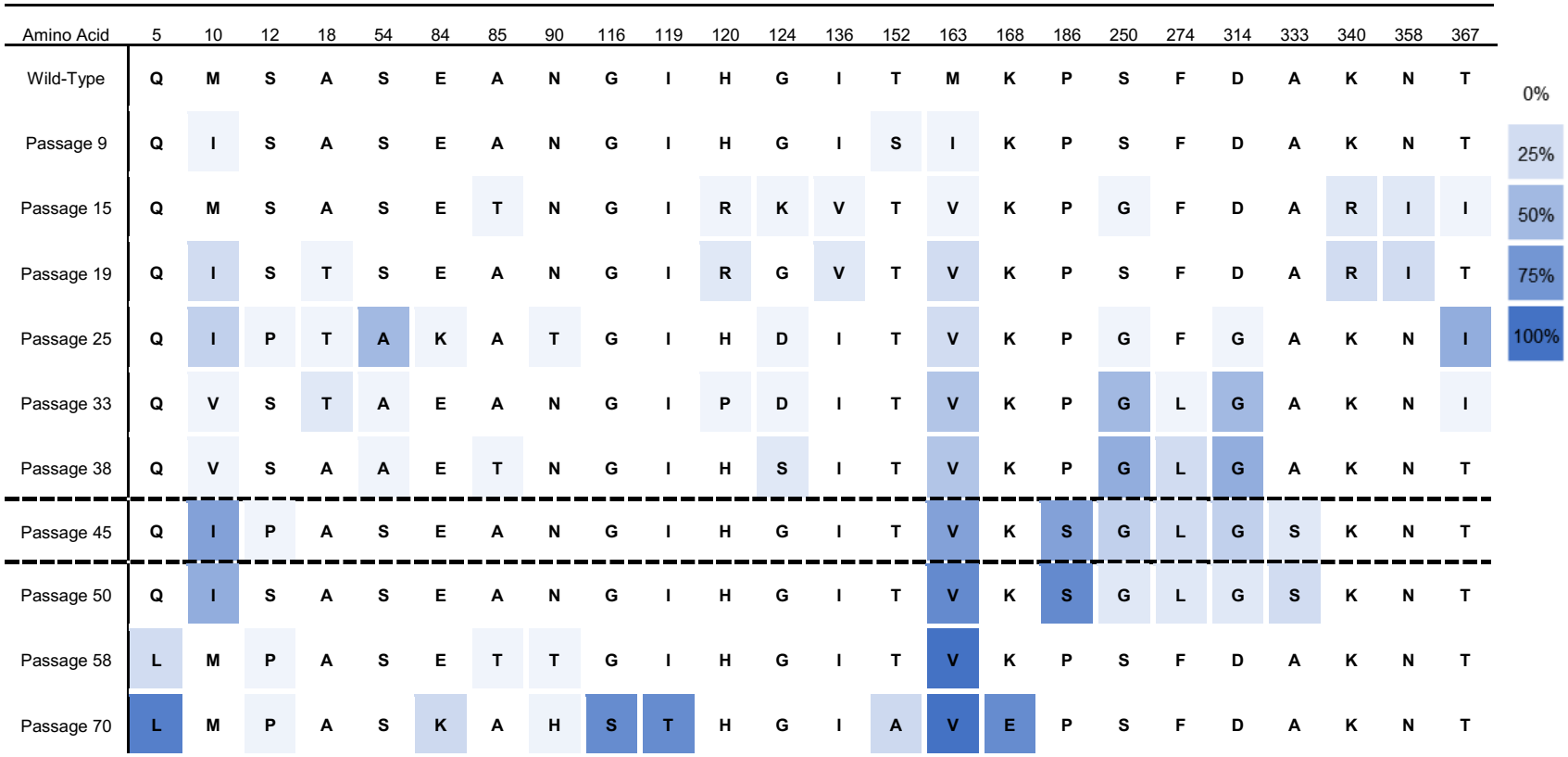


Figure 37: Sequencing Data for PANCE of *B. methanolicus mdh 1* and *mdh2*.
The degree of shading shows the percent convergence of the individuals sequenced for a given passage number for each mutation. Each passage shows sequencing data for 11 or 12 individuals. Dotted lines indicate points at which methanol concentrations were lowered for all following passages.

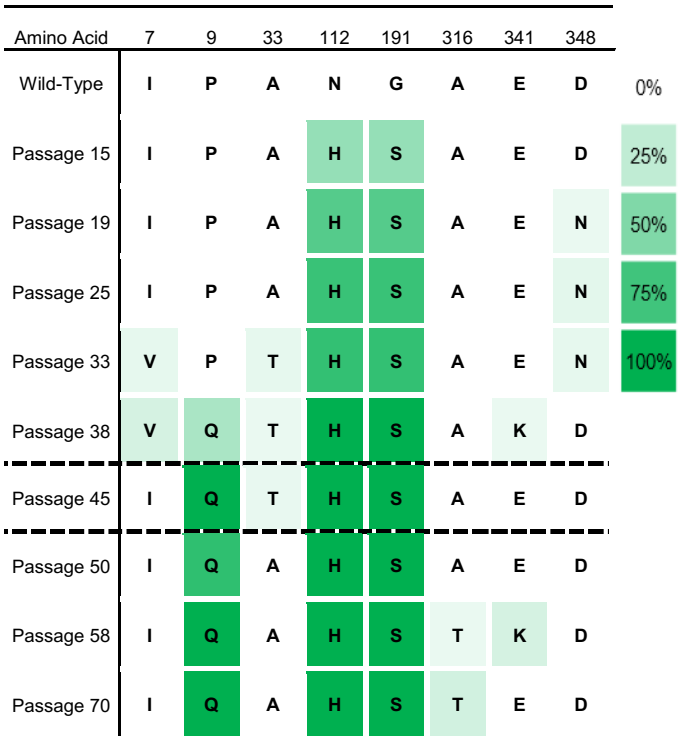


Figure 37 (continued): The degree of shading shows the percent convergence of the individuals sequenced for a given passage number for each mutation. Each passage shows sequencing data for 8-12 individuals. Dotted lines indicate points at which methanol concentrations were lowered for all following passages.

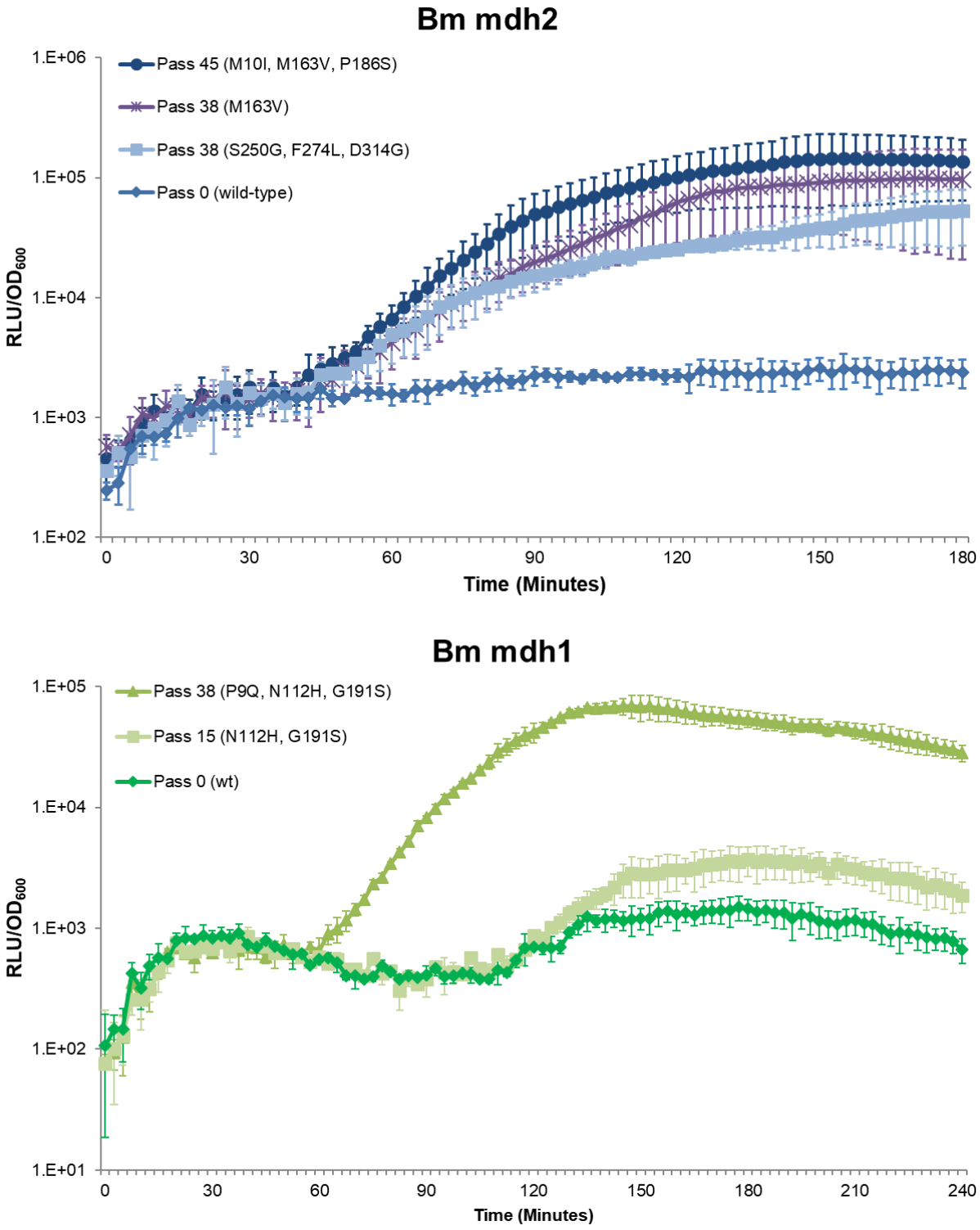


Figure 38: Individuals from PANCE intermediate populations show increased apparent activity.
Bm mdh2 was assayed using pTR47m4 with 40ng/mL ATc as previously reported[138]. *Bm mdh1* was assayed using the same AP used during PANCE (Figure 36A). Both plots show continuously monitored luciferase signal starting from induction of *mdh* expression from a P_{BAD} promoter with 10mM arabinose concurrent with addition of 500mM methanol at 0 minutes. RLU, Relative Luminescence Units. OD_{600}, Optical Density at 600nm.

Section 3.4: Plate-based non-continuous evolution of various methanol dehydrogenases

Bolstered by the clear improvements we observed in our two initial PANCE populations, we next wanted to see if we could take advantage of the iterative nature of this technique to address a key shortcoming of PACE: parallelization. Whereas a continuous flow system becomes prohibitive in parallelized setups without specialized fluidics platforms, PANCE can feasibly be performed in 96-well plates with little added labor compared to the culture-tubes used in our initial experiments. To demonstrate this, we evolved three replicate populations of the same phage starting materials with 6 different concentrations of methanol, totaling 18 separate evolving populations per Mdh target. For our targets, we chose to once again evolve both *mdh1* and *mdh2* from *B. methanolicus*[121], as well as the previously evolved *mdh2* CT4-1 from *C. necator*[100]. We used the same drift-selection cadence as our initial PANCE for this plate-based experiment, adjusting stringency by changing dilution rates between rounds (**Figure 38**). Given the difficulty of obtaining titers for 54 separate populations in parallel for each passage, phage population size was only monitored periodically or qualitatively assessed via PCR, which greatly minimized the degree of researcher intervention and resources required. As **Figure 39** shows, after 38 passages in these 96-well plates, phage populations were unsurprisingly largest for populations fed with higher concentrations of methanol. SPs encoding *mdh* genes with known higher starting activities also predicted greater titers at lower methanol concentrations, with *Cn mdh2* CT4-1 phage (SP30) persisting at relatively high titers in all populations down to 125mM methanol. *Bm mdh1* and *mdh2* showed less success, with titers dropping precipitously for SP13 (*Bm mdh2*) between 250mM and 125mM methanol conditions, and titers dropping between 500mM and 250mM for SP14 (*Bm mdh1*). One SP14 population washed out across all methanol doses entirely. Sequencing of each well shows that mutational convergence was surprisingly diverse between populations despite all conditions being held equal. As was seen with the single-culture PANCE experiment, individuals within each population frequently showed signs of separate converged solutions (**Figure 39**).

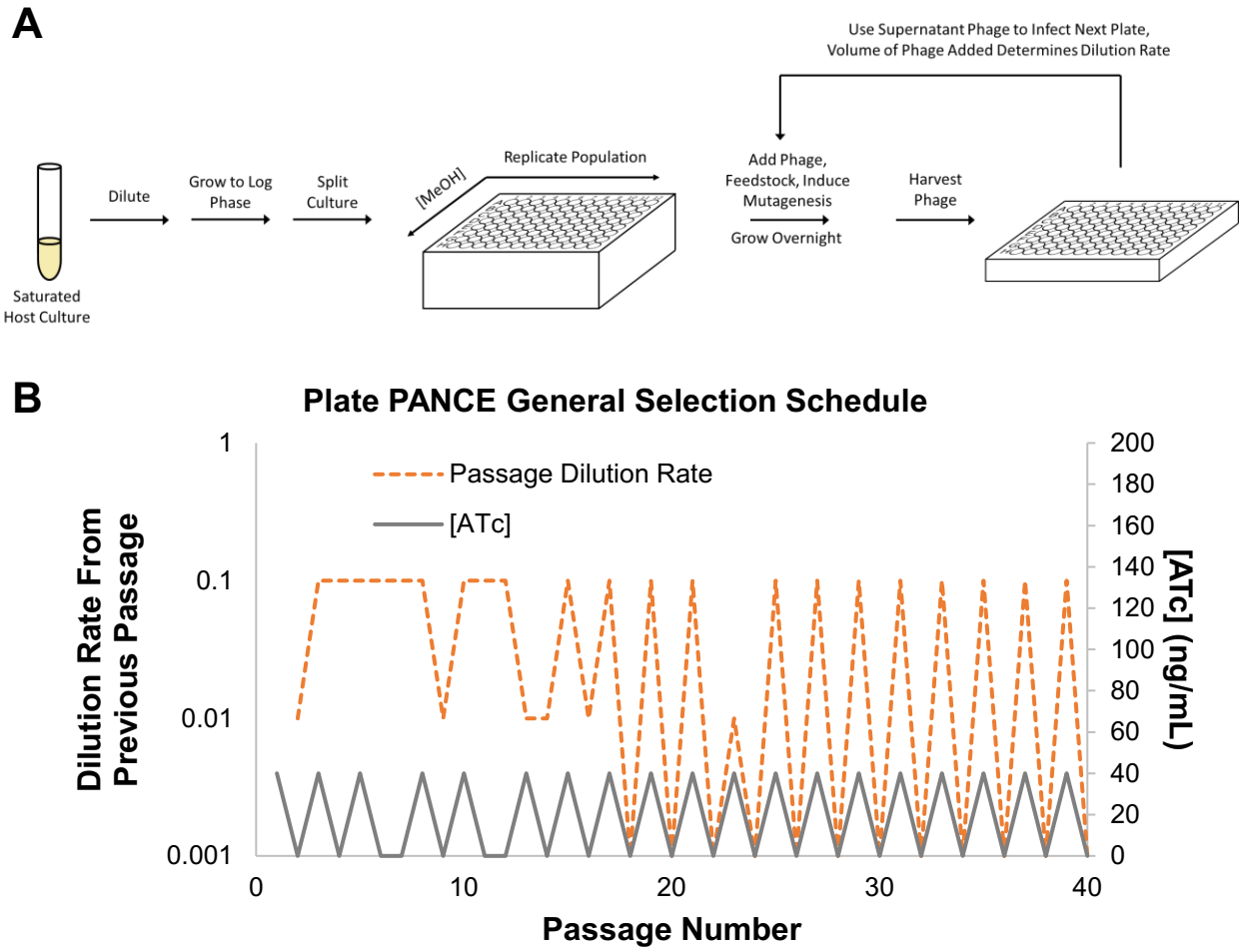


Figure 39: Plate-Based PANCE Evolution Allows for Many Populations to Evolve Simultaneously.
A. General protocol for plate-based PANCE. For Mdh evolutions, selection hosts were identical to the original, single-culture PANCE experiment. **B.** General selection schedule for plate-based PANCE of Mdh. ATc controls induction of the PSP-gene III circuit on MP variant DP6 as shown in Figure 35B

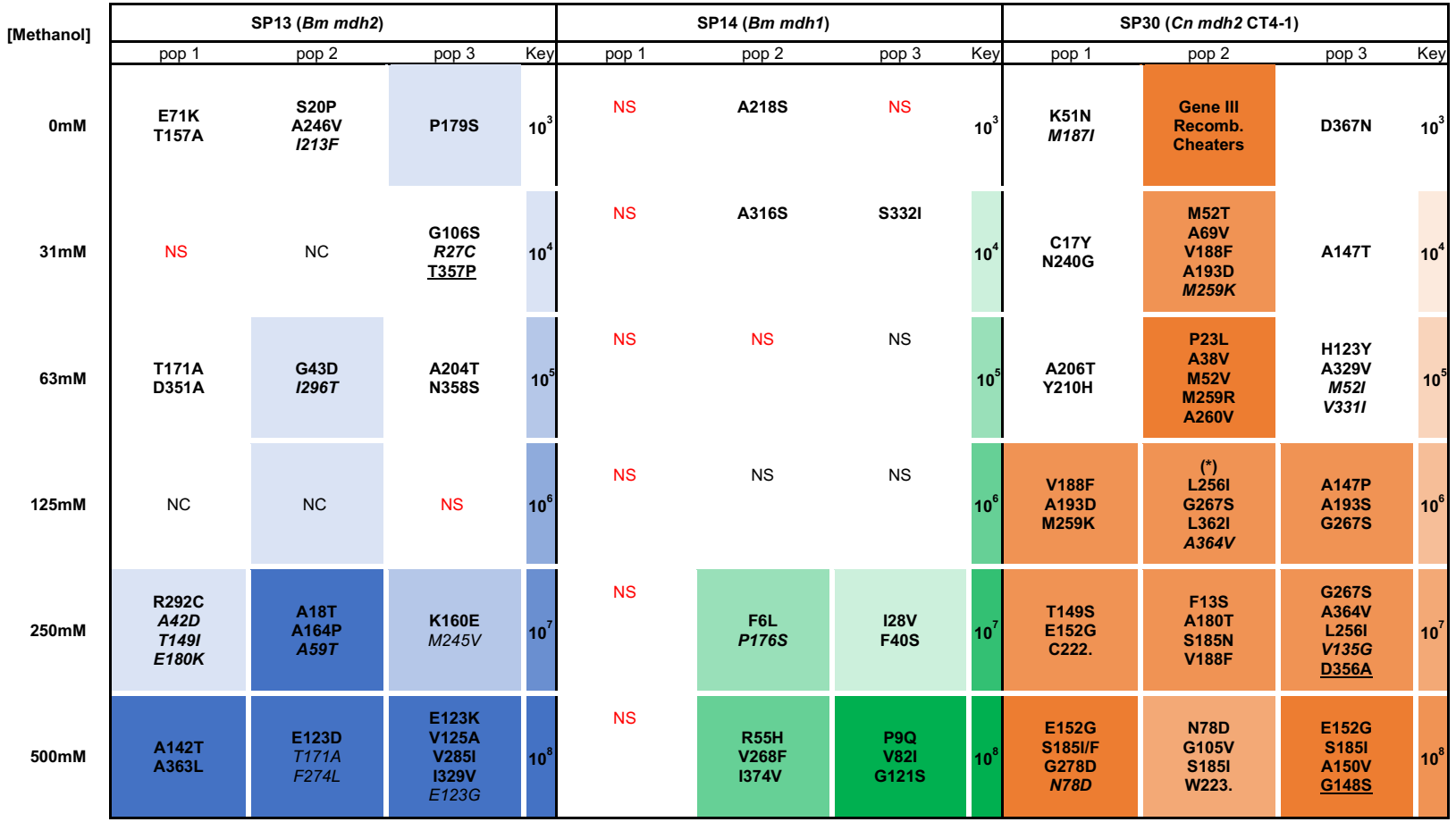


Figure 40: Sequencing and titer data for 96-well plate PANCE of methanol dehydrogenases.
Data are shown for passage 38 of the experiment. The "Key" column for each SP shows the phage titer at this passage for the given concentration of methanol in plaque-forming units per mL. Each condition was evolved in triplicate, notated by "pop 1," "pop 2" and "pop 3." Mutations shown for each population were coding mutations observed in at least 2 individuals sequenced out of up to 8 per population. Separate genotypes are notated with bold, italicization, and underline – mutually exclusive mutation sets will only have one or the other, whereas overlapping sets will have both. "NC" denotes a lack of coding mutation consensus between any 2 individuals sequenced. "NS" populations were not sequenced, and "NS" populations in red text did not form any visible plaques during titer measurement. "Gene III Recomb. Cheaters" indicates that the region of the phage typically carrying the gene of interest was sequence-confirmed to be a copy of gene III from the AP, allowing phage in this population to bypass the selection entirely. The (*) shows the only culture that evolved recombinant cheaters early in the experiment, which was re-infected the following passage using phage from the well immediately below it on the table. Despite this common ancestry, the two populations still show divergent genotypes.

While we were encouraged by the high titers of our SP30 populations, we noticed a significant problem in which *Cn mdh2* genes subcloned from variants in the populations would show strong apparent activity in our luciferase reporter assays even when expressed in the absence of methanol substrate. Additionally, our attempts to determine crude lysate activity in strains expressing these same variants failed to show any detectable NAD+ reduction at any methanol concentration, implying a complete loss of catalytic activity. This phenotype combined with the high frequency of premature stop codons fixing in these populations led us to suspect an unexpected "cheater" phenotype in these populations. Conversely, the *Bm mdh1* containing phage populations had the opposite problem of the *Cn mdh2* populations, in which almost all of them washed out, and those that survived showed similar genotypes to what we had previously seen in our single-culture evolution. Attempts to refine these populations in follow-up, standard PACE experiments (*i.e.* in a continuous flow system rather than iterative overnight passaging) resulted in a strong convergence of cheaters in SP30 populations and no significant changes in genotype to SP14 populations, suggesting both lineages had hit evolutionary dead-ends. Thus, we focused our analysis on the four *Bm mdh2* populations showing high (>10^6 pfu/mL) phage titers after 38 passages, all of which showed significantly different mutations than our initial single-culture PANCE experiment.

Among these new converged mutations in *Bm mdh2*, the A164P mutation was the most obvious expected mutation based on the evolution of *Cn mdh2 CT4-1* (though, interestingly, the *Cn mdh2* result was an A to V mutation, not an A to P) as well as our earlier evolution of M163V. Two other mutations also stood out: first, E123 was found to be mutated to one of three different amino acids in three distinct genotypes across two separate populations. Second, A363L, a mutation requiring two nucleotide changes in the codon to access, was strongly converged upon in yet another population. Characterization of sub-cloned mutant genes containing these mutations showed similar increases to apparent activity as M163V-containing variants. This demonstrated to us that a 96-well plate approach enables a wide array of evolution conditions to be tested in parallel

rather than iteratively using PACE. At 500mM methanol (the least stringent condition tested) we could repeat our single-culture PANCE results, while at a more stringent condition (250mM methanol) we saw increasing incidence of washout, which increased in severity as methanol concentrations continued to lower.

Section 3.5: Phage Assisted Continuous Evolution of PANCE-evolved methanol dehydrogenases

To determine which of our PANCE populations was the most fit and attempt to further evolve our target *mdh* genes, a subsequent PACE experiment was run using these *Bm mdh2* PANCE populations as input. Initially we separated the two PANCE experiments into two lagoons. The first tested the fitness of our 70-passage, single-culture experiment, while the second determined which of the plate-evolved populations was most fit. As shown in **Figure 41A**, both sets of PANCE-evolved populations were able to survive even the relatively stringent PACE conditions of 0.5% Methanol (v/v) (125mM) and 2 lagoon volumes per hour. Sequencing of individuals from the end of this experiment showed a clear convergence of derivatives from the initial, single-culture PANCE of *Bm mdh*2, but showed a less clear convergence for the pooled populations from the plate-based PANCE of the same gene. Instead, this second lagoon ended on a mixture of several, separate genotypes, each a likely derivative of a replicate population evolved during PANCE at 500mM or 250mM methanol (**Figure 41C**). To further hone in on the best possible *mdh2* variant from this set, a follow-up PACE was run combining these two PACE endpoints (**Figure 41B**). The lagoon converged almost immediately on derivatives from the single culture PANCE experiment. However, characterization of the consensus mutations suggests that these individuals are capable of de-repressing FrmR in the absence of a methanol feedstock (**Figure 42**).

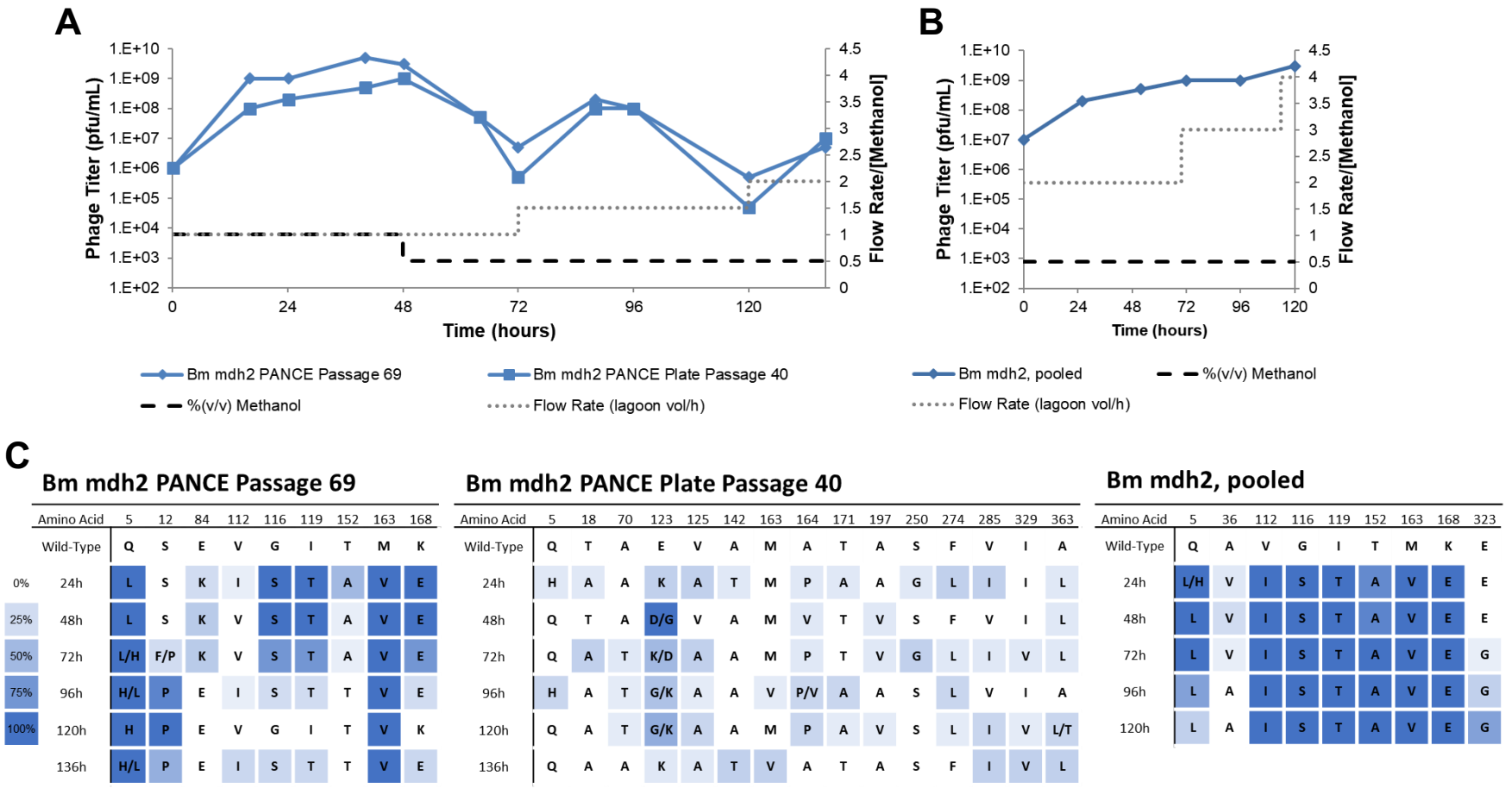


Figure 41: PACE of *Bm mdh2* PANCE populations.
A. Initial PACE, with separate lagoons seeded with the indicated passages from PANCE. For Plate Passage 40, equal volumes of all populations evolved above 0mM Methanol were pooled and used to infect the lagoon. Methanol concentrations are listed as % (v/v), with 2% being roughly equivalent to 500mM. pfu, plaque forming units. **B.** Second PACE experiment seeded by pooling the final timepoints for the lagoons in A. **C.** Sequencing data for the PACE experiments. The plate passage shows low convergence for many mutations, reflecting a diverse range of separate genotypes co-existing within a single lagoon rather than convergence on a single genotype.

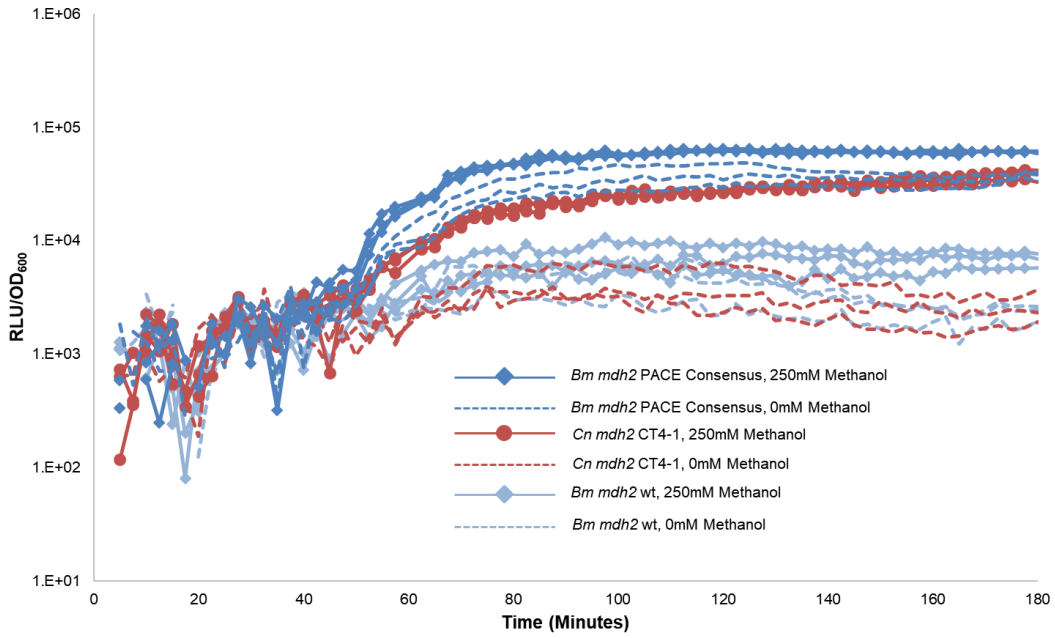


Figure 42: Converged PACE Genotype De-Represses FrmR in the Absence of Methanol.
RLU, relative luminescence units. OD_{600}, optical density measured at 600nm.

These results were almost identical to those we had seen for *Cn mdh2* during our plate-based PANCE, and the lagoon we ran in parallel using *Cn mdh2* populations as input quickly converged on a genotype highly similar to a known inactive variant, much like we seemed to be seeing for *Bm mdh2*.

Section 3.6: Development of a negative selection against inactive Mdh cheaters

With the evolution of these "cheaters" and no clear hypothesis as to their mechanism of cheating outside of a direct interaction with FrmR, we developed a negative selection host[8,11] to see if this lineage could be suppressed or reverted by punishing FrmR de-repression in the absence of methanol feeding (**Figure 43A**). Initial characterization of this negative selection host showed that it would propagate functional *mdh*-carrying SPs only in the absence of methanol, consistent with our desired phenotype. We also saw significant de-enrichment of one of our evolved populations of

SP13 encoding known cheaters when we passaged them even a single time in this strain both with and without methanol, suggesting this host as a potent means of preventing cheaters from sweeping a population during PANCE (**Figure 43B**). Because we validated this negative selection using a mixed population of cheaters, we assumed that the residual PCR product indicating survival of some members of this population resulted from either functional or completely inactive individuals that had been isolated along with the majority population of cheaters.

We implemented this new negative selection strategy by revisiting our initial SP13 PANCE. We started at the timepoint immediately after our first drop in methanol concentration, since we hypothesized that the more stringent, lower methanol concentrations from the initial evolution rewarded non-formaldehyde-producing variants more significantly and led to evolution of cheaters. We switched to a chloramphenicol antibiotic stock dissolved in DMSO instead of ethanol during positive selections and retained the use of the ethanol stock only during negative selections to help select against possible promiscuous enzyme activities biasing our selection. While it seemed unlikely that we would achieve the high levels of acetaldehyde needed to give a robust FrmR response (See: Chapter 2, **Figure 24**)**,** this would in theory help limit promiscuous oxidation from biasing our results during positive selection. Despite propagating this population for over 30 passages with a negative selection cycle inserted between drift and positive selection (**Figure 44A**), no new mutations arose when methanol concentrations were maintained at 1% (v/v) (~250mM) and cheaters eventually evolved after methanol concentrations were lowered to 0.25% (v/v) (~62.5mM). We tried to weed out these cheaters by switching to an aggressive negative selection schedule and evolving two populations side-by-side: one population being the cheater-enriched passage 102 from our initial failed negative selection, and one branching off from an early timepoint of that population that we had sequence-confirmed as not having any strong indication of cheater genotypes (**Figure 44B**). No new, enabling mutations arose to prominence in the population before cheater mutations appeared, possibly due to the lack of strong selection pressure

for methanol oxidation activity compared to the number of negative selection cycles. Interestingly, the negative selection created its own, divergent lineage of cheater mutations distinct from those we observed in descendants of our first experiment.

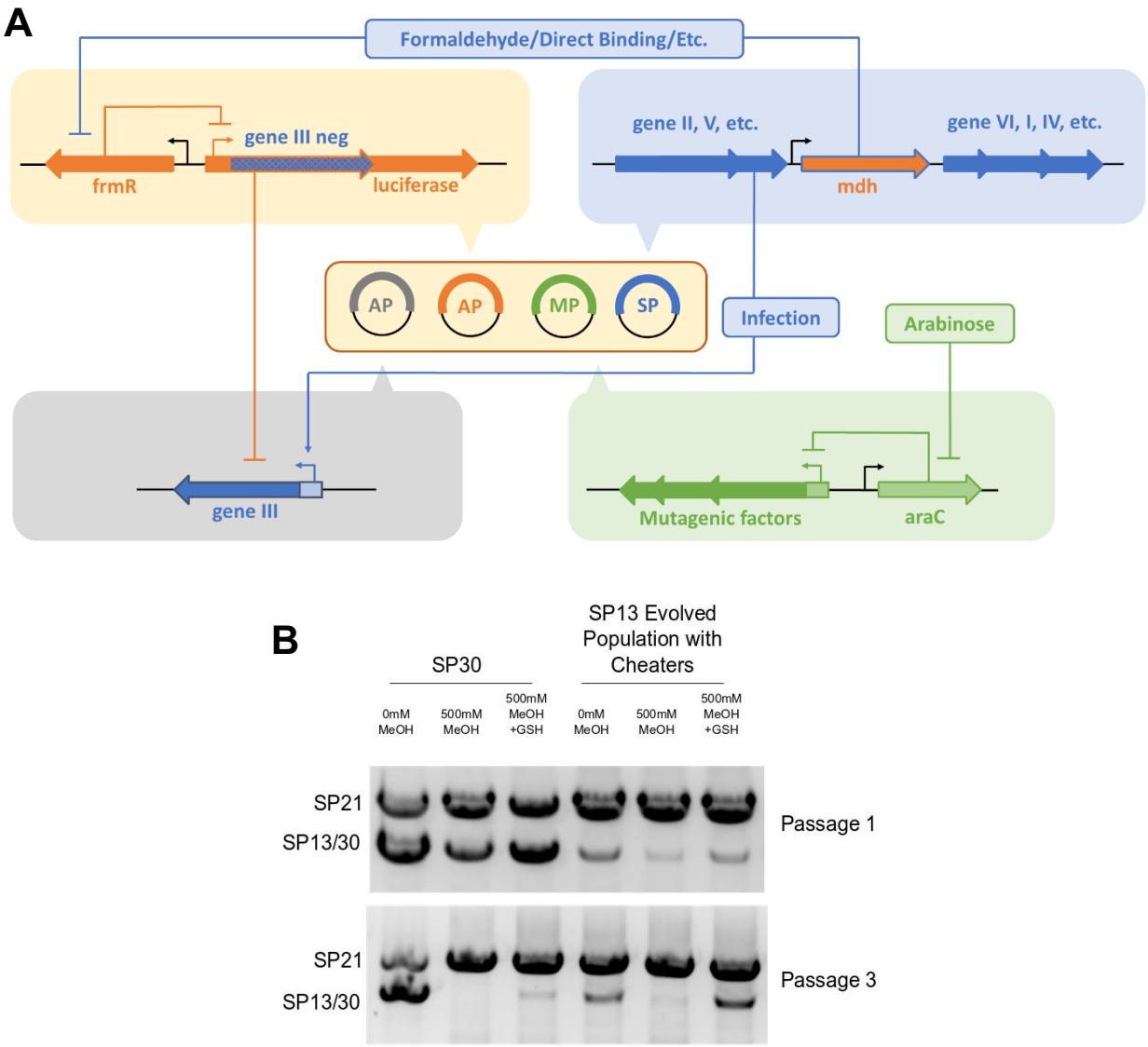


Figure 43: Design and Validation of a Negative Selection for Mdh Evolution.
A. Diagram of negative selection host cells. De-repression of FrmR triggers expression of gene III neg, which impedes the utility of gene III from a secondary AP. **B.** Application of negative selection host without mutagenesis to a population of equal titers of SP21 (non-Mdh containing) and SP30 or SP13 (Mdh-containing) phage under various conditions. MeOH, methanol, GSH, glutathione.

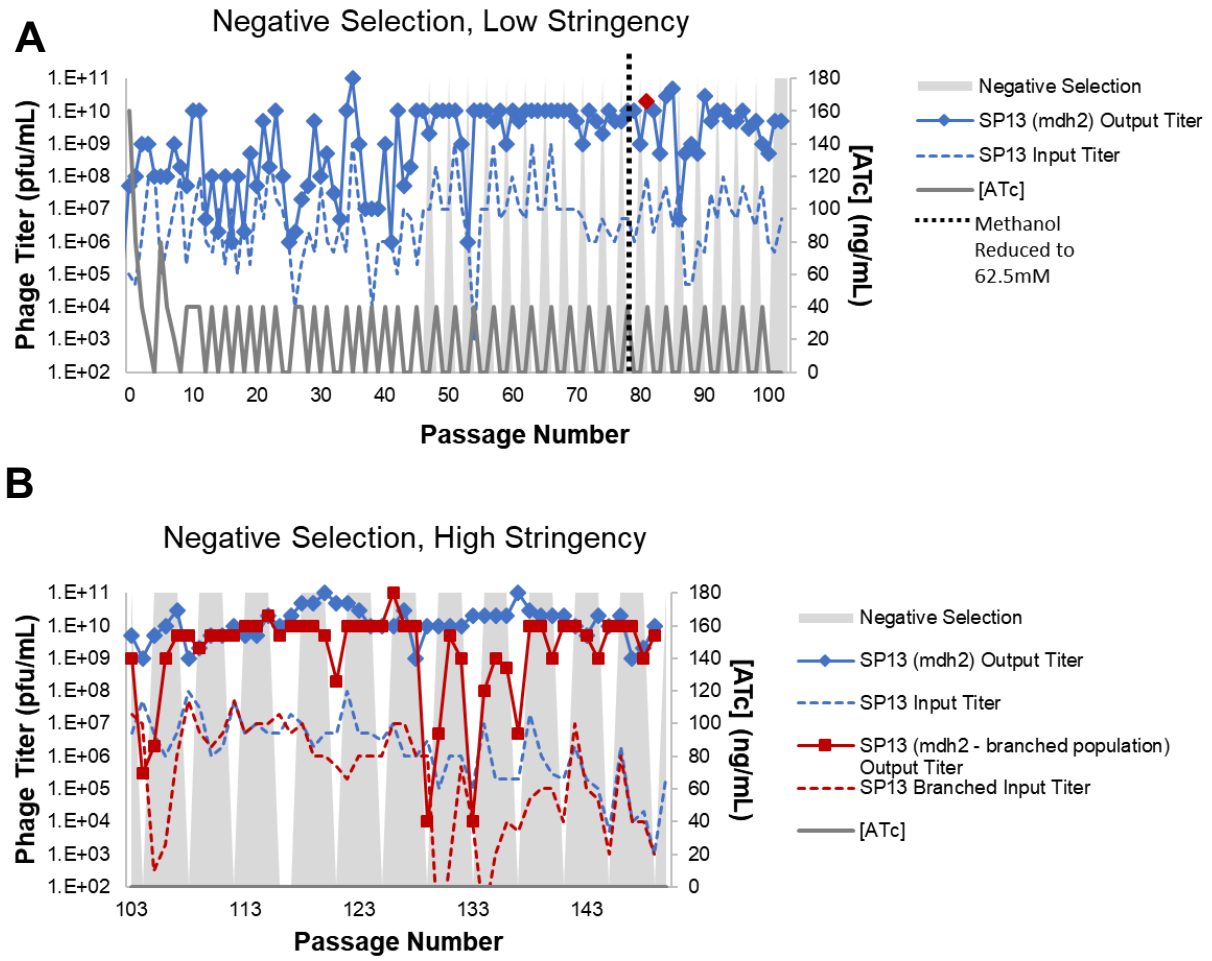


Figure 44: Negative Selection PANCE Fails to Eliminate Cheater Phenotypes Using Two Different Selection Schedules.
A. Initial attempt at using negative selection passages (grey) within the original PANCE drift-positive alternating selection schedule. Passages 1 through 45 are equivalent to those shown in Figure 35, with the negative selection starting at passage 46. The passage that branches into B is indicated as the red point. B. Continuation of negative selection using both the cheater-enriched population 103 and a branched population from the red point in A. In this selection schedule, drift was replaced with two extra negative selection passages between positive selection rounds. pfu, plaque forming units.

Section 3.7: Characterization of *B. methanolicus* Mdh2 variants showing direct interaction with FrmR

To test our hypothesis about direct interaction with the FrmR repressor protein being the mechanism of cheating, we cloned a variant of our typical AP (**Figure 35A**) that contained a mutant *frmR* gene with a C35S mutation. Mutations at this reactive cysteine are known to eliminate the FrmR protein's response to formaldehyde concentrations[108–110], but to our knowledge no one has reported on whether or not other mechanisms exist that de-repress FrmR from its DNA binding site. By inducing our "cheater" variants with or without methanol, we easily confirmed that de-repression is robust in every dominant cheater genotype pulled from our selections, regardless of methanol dosing. Fortunately, this cheating phenotype was not present in some of our earliest *Bm mdh2* variants, with our data clearly showing that the common ancestor of all eventual *Bm mdh2* chaters – Bm Mdh2 Q5L M163V E180 – exhibits no de-repression of FrmR C35S. As expected, induction from the wild-type *Bm mdh2* and *Cn mdh2* CT4-1 genes also failed to de-repress this mutant FrmR (**Figure 45**).

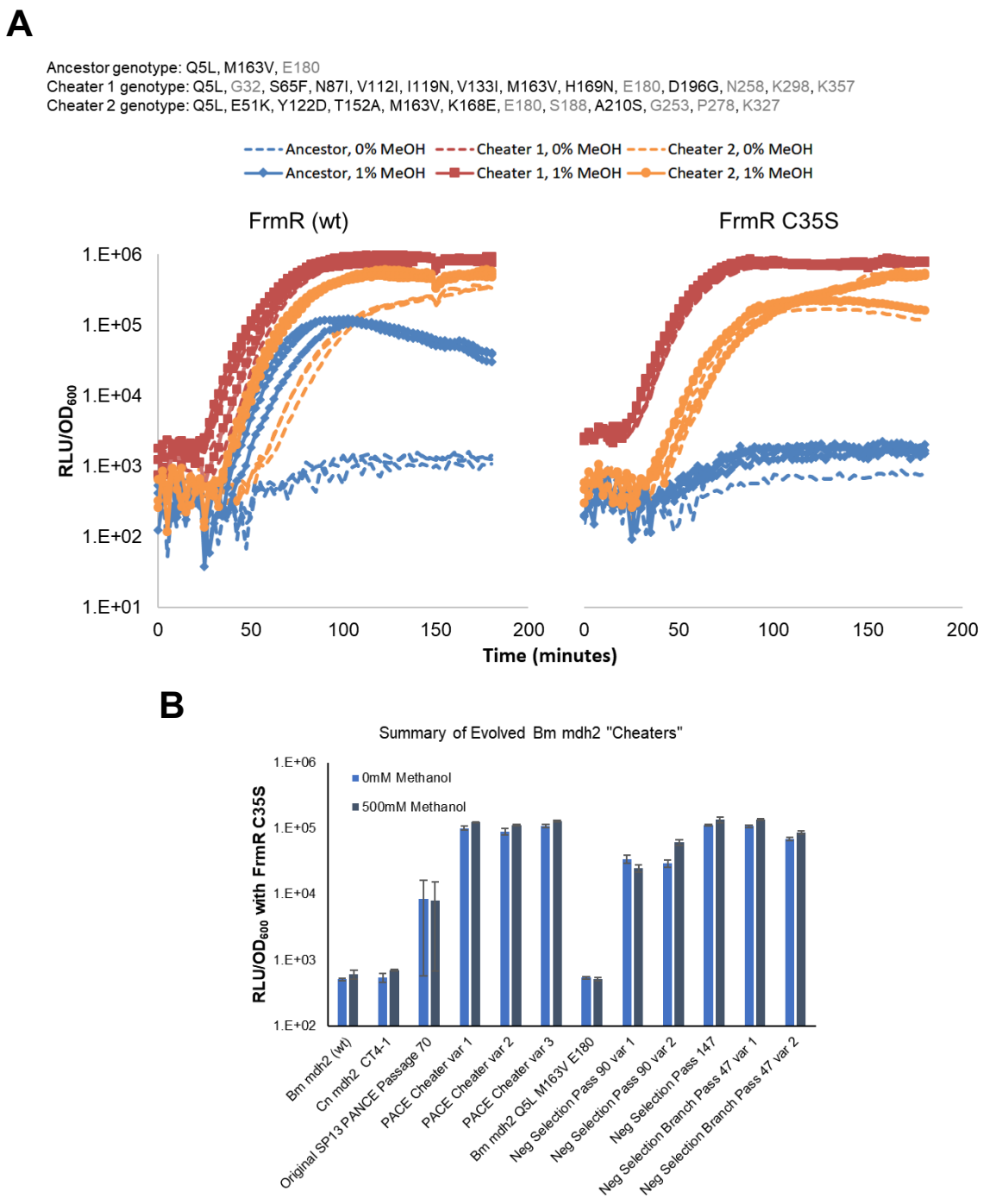


Figure 45: Cheater Characterization Implies a Direct Interaction between FrmR and Evolved Mdh Variants. **A.** Comparison of cheater interactions with both wild-type FrmR and FrmR C35S. The reporter construct used was pTR47m4, with 40ng/mL ATc inducing FrmR expression prior to induction of *mdh2* variants from a separate plasmid by addition of 10mM Arabinose at 0min. **B.** Reporter data using the PANCE AP with a FrmR C35S mutant repressor that does not react with formaldehyde. Bars show the average error bars the standard deviation of the maximum signal obtained from continuous monitoring of three biological replicates over four hours. RLU, Relative Luminescence Units; OD_{600}, Optical Density at 600nm.

Interestingly, both through PACE and this negative selection PANCE, cheaters with more dramatic FrmR interactions appeared, showing much stronger de-repression of a formaldehyde-insensitive C35S version of the repressor. Cell lysate NAD+ consumption assays failed to show any discernable activity for the first evolved "cheater" mutant from passage 70 of our first PANCE experiment, a variant which contains only 5 coding mutations, 3 of which (G116S, I119T, and K168E) seem responsible for the FrmR interaction given that a separate evolved variant containing the other 2 (Q5L, M163V) had already been confirmed not to have any detectable cheater phenotype (**Figure 45B**). It's possible that there is a combined fitness benefit to evolving a FrmR-binding phenotype while also losing a formaldehyde production phenotype, which otherwise would cause additional stress to host cells during the evolution. Although not comprehensive, we examined a few cheater mutation reversions in the branched lineage of the negative selection PANCE and confirmed that for some reversions, all methanol oxidation apparent activity had been completely lost, suggesting that intermediate genotypes prior to cheating are not a likely source of desired improvements. While the exact mechanism of cheating is unknown, our results strongly point toward a direct, protein-protein interaction between methanol dehydrogenases and the FrmR protein. Given that two separate methanol dehydrogenases from two separate organisms managed to evolve this strong phenotype within three to five amino acid changes, we suspect there may be some natural basis for this interaction. In some ways, this makes sense, given that a natural interaction of this nature would allow a cell to couple expression of a formaldehyde-producing enzyme with expression of formaldehyde detoxification factors. However, it is unclear what advantage this would have over the existing mechanism to sense the toxic formaldehyde molecule itself. This may, alternatively, point toward a more promiscuous tendency of FrmR or Mdh enzymes to bind to other proteins. Regardless of the mechanism, at this point it seemed clear that a FrmR-based selection for Mdh activity is fundamentally limited by the propensity of Mdh2 to directly induce expression from its promoter. Perhaps further optimization of the negative selection or

restarting a fresh lineage in PANCE with negative selection included from the very beginning would help circumvent the results we observed, but it seemed inevitable to us that evolving our *mdh*-encoding SPs at low methanol concentrations would drive them to find solutions that did not require a methanol oxidation phenotype.

Section 3.8: Evolved *B. methanolicus* Mdh2 variants show improved activity *in vitro* and *in vivo*

From sequencing data and a preliminary set of cell lysate analyses, it seemed to us that the most likely candidate mutations for improved methanol oxidation were E123G, M163V, A164P and A363L. It struck us as odd that despite many rounds of PANCE and a subsequent PACE experiment, none of these mutations recombined or evolved in sequence. PACE is known to permit recombination between phage genotypes[10], and yet pooling high titers of at least four separate, improved *Bm mdh2* genotypes did not result in a single, winning genotype, even after several days of stringent selective pressure. Using the Q5L and E180 mutations that persisted with the M163V genotype throughout our PACE and negative selection populations as a background strain, we manually cloned every possible single mutant and combination of these four mutations into a set of expression plasmids to see any of them exhibited higher activity than non-cheater variants we evolved in PANCE and PACE. We also cloned several other strains to test some of our assumptions about our combinatorial mutant genes. We already had one strain containing the M163V mutation without the Q5L or E180 mutations from our early PANCE experiments (See **Figure 38**, *Bm mdh2* pass 38), which we included in our analysis to determine if there is any effect conferred by these two mutations. We also tested two other A363 mutations (A363S, A363V) representing the single-nucleotide mutations required to reach the A363L double-mutation we observed in PANCE. Finally, we also tested our most representative *Bm mdh1* variant from our initial PANCE experiment to see if it would compare at all to our evolved *Bm mdh2* variants. We initially ran luciferase reporter

assays in both the Δ*frmA* and standard S1030 strains and compared them to cell lysate kinetics assays for each of these variants in S1030 cells (**Figure 46**). We omitted the reporter from the lysate strains to prevent FrmR acting as a formaldehyde sink and biasing observed rates of NAD+ reduction. From our data, it seems clear that A363L outperforms either single-nucleotide A363 mutation by a significant margin. Additionally, Q5L and E180 appear have a neutral-to-slightly-beneficial impact on activity in both crude lysate and luciferase assays both on their own and when combined with M163V. Outside of these controls, we were pleased to see that every single mutant we subcloned showed significantly improved activity over wild-type in both reporter and cell-lysate assays. While luciferase reporter assay data suggests that some of these single mutations in combination may have a slight improvement in activity (less than a 2-fold change in relative luminescence signal), crude lysate data suggests that, in general, the A164P and A363L mutations on their own provide the best improvements. For crude lysate activity, recombining these two best-performing mutations with each other or E123G/M163V either has no effect on activity compared to each individual mutation or significantly reduces activity, suggesting a negative epistatic interaction. However, reporter assays show these combinations as generally better than individual mutants, particularly for A363L. Since the reporter assay uses the same AP as the selections, it was possible that the discrepancy in reporter data and cell lysate data stems from the same "cheater" phenotype observed during evolution – this would likewise explain why some variants show little or no difference in reporter assay between the formaldehyde-sensitized knockout strain S1030 Δ*frmA* and the wild-type strain while others maintain at least a 2-fold higher signal for the Δ*frmA* background. However, no FrmR binding phenotype was observed when expressing these Mdh proteins without methanol, nor when expressing them against FrmR C35S at any tested concentration of methanol, with all normalized luminescence signals being below 1000 RLU/OD_{600} regardless of methanol dosing or induction of *mdh* expression. The development of cheaters during evolution and the discrepancies between *in vitro* and *in vivo* data for our variants highlight the

important differences in reporter-based assays done *in vivo* and more direct assays for either *in vitro* activity or substrate production in culture. While FrmR monitoring of formaldehyde is clearly sufficient for comparing larger differences in activity, our data show that it hits saturation for more active variants at high methanol concentrations and likely requires more fine-tuning to discern the subtler differences between these activities.

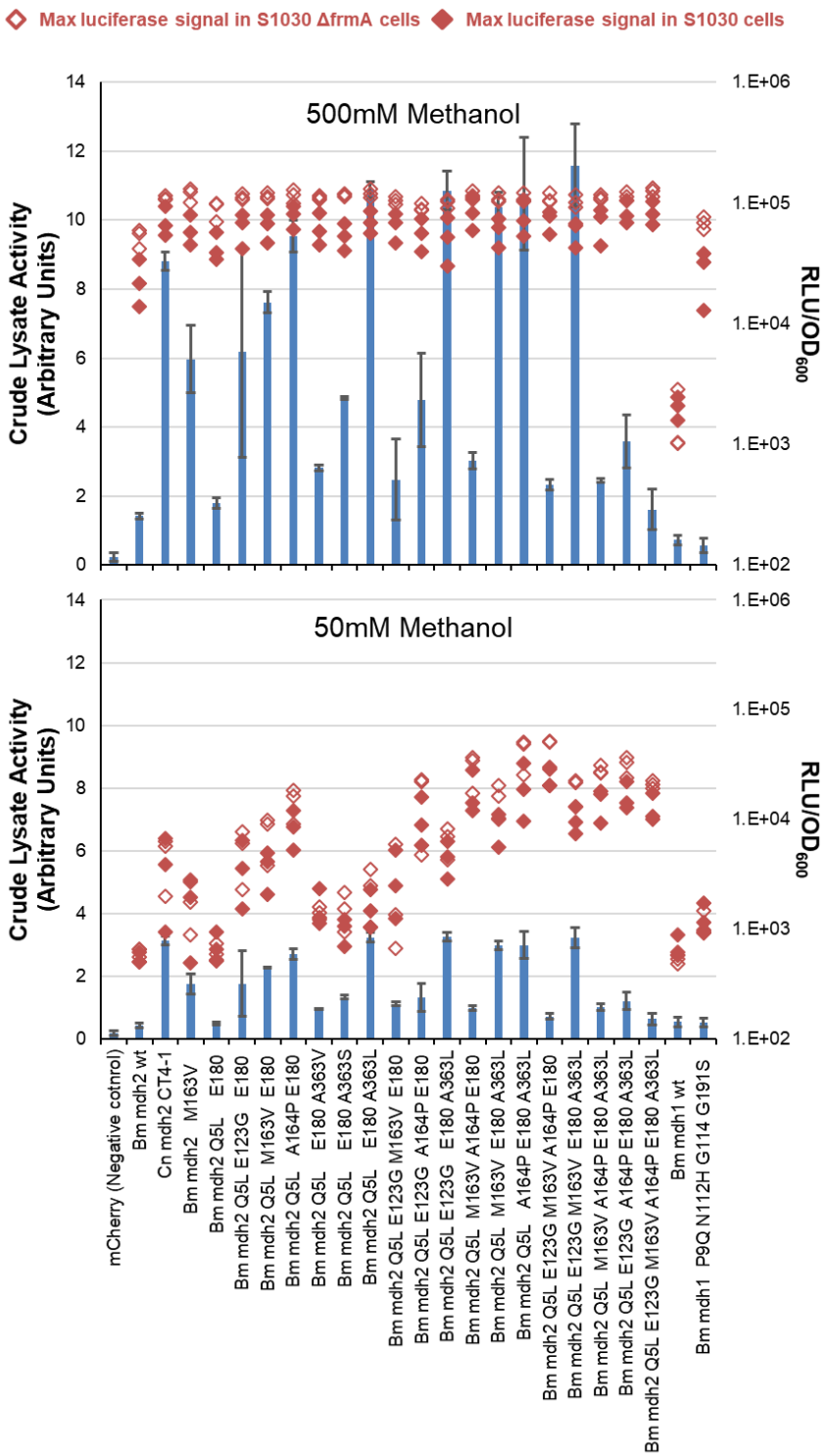


Figure 46: Cell Lysate and Luciferase Reporter Data for Combined Mutation Sets for Bm mdh2.
Diamonds show the maximum luminescence signal seen during continuous measurements taken for 4 hours after induction of Mdh and addition of methanol to log-phase cultures in DRM. Separate diamonds for each strain show biological replicates measured across separate experiments testing 0mM, 50mM or 500mM Methanol on the same plate. All reporter values at 0mM Methanol were below 800 RLU/OD_{600}. Crude Lysate Activity bars show the average for two replicates, with error bars showing the distance between each individual value. RLU, Relative Luminescence Units; OD_{600}, Optical Density at 600nm.

We next purified 6x His-tagged variants of each of the Q5L E123G/M163V/A164P/A363L variants for further *in vitro* testing to determine if their increased signal in cell-free lysate assays was reflected in changes to their kinetic parameters. We also purified the A164P A363L double-mutant to test the general trend of unaffected or lowered crude lysate activity for combined mutations. We recorded the initial velocities for these enzymes across nine to eleven methanol concentrations at a fixed NAD+ concentration of 1mM. We fit this data to the Michaelis-Menten equation to calculate and compare V_{max} and K_m values for each variant (**Figure 47**, **Table 3**). While the state-of-the-art Cn Mdh2 CT4-1 still showed a significantly lower K_m than any of our Bm Mdh2 enzymes (consistent with previous *in vitro* characterizations[100,121]), we observed a slightly higher V_{max} for Bm Mdh2 Q5L A363L and a general trend of improved V_{max} and K_m for all evolved mutants compared to wild-type Bm Mdh2, effectively corroborating the results of our selections prior to their evolution of cheaters.

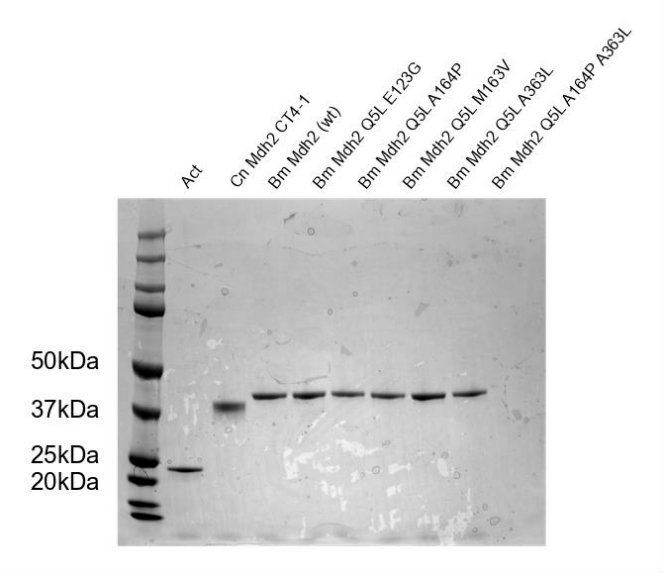


Figure 47: Purified Bm Mdh2 Variants Show Improved *in vitro* activity.
SDS-PAGE gel showing purified enzymes. Act is a Nudix Hydrolase known to improve the activity of many Mdh enzymes *in vitro*[119,120,139], but was not ultimately used in this study.

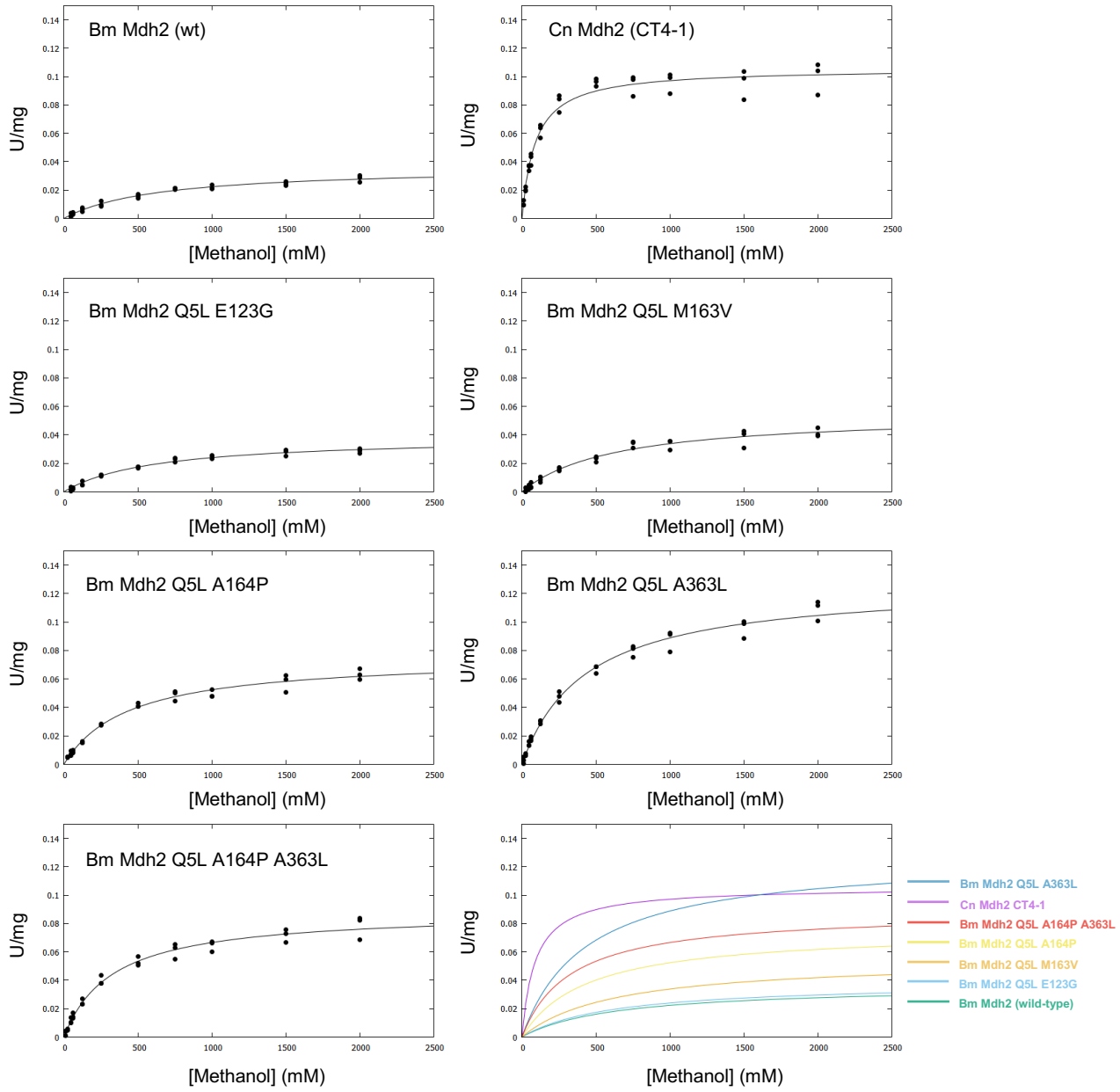


Figure 47 (Continued.): All tested mutant enzymes exhibit improved *in vitro* activity compared to wild-type Bm Mdh2. Graphs show initial velocities measured in triplicate at various methanol concentrations for each enzyme based on measuring NADH absorbance at 340nm. NAD+ was added to each reaction at 1mM. Lines represent the fit of a given dataset to the Michaelis-Menten equation. The bottom-right graph shows an overlay of all Michaelis-Menten curves for tested enzymes. U/mg, Units per milligram protein, with 1 Unit defined as 1 µmol NADH per minute.

Table 3: Kinetic Parameters for Evolved Bm Mdh2 Enzymes. Values are given based on the Michaelis-Menten fit for the datasets shown in Figure 46 plus-or-minus the asymptomatic standard error. mU/mg, milli-Units per milligram of enzyme, with 1 Unit defined as 1 μmol NADH per minute.

Enzyme	$V_{max\ MeOH}$ (mU/mg)	$K_{m\ MeOH}$ (mM)
Cn Mdh2 CT4-1	**105.8** ± 2.1	**88.8** ± 7.8
Bm Mdh2 (wt)	**36.5** ± 1.7	**636.1** ± 74.1
Bm Mdh2 Q5L E123G	**38.8** ± 1.6	**614.9** ± 65.9
Bm Mdh2 Q5L M163V	**55.0** ± 3.1	**627.3** ± 89.2
Bm Mdh2 Q5L A164P	**75.4** ± 2.3	**439.5** ± 38.6
Bm Mdh2 Q5L A363L	**127.2** ± 3.3	**431.8** ± 32.3
Bm Mdh2 Q5L A164P A363L	**88.5** ± 2.3	**328.5** ± 27.6

To further finalize these results, we wanted to compare the effects of these evolved mutations on the *in vivo* incorporation of methanol into central metabolites using the Hps-Phi-mediated ribulose phosphate pathway required for engineering synthetic methylotrophy[98]. Since PACE and PANCE are selections for *in vivo* production rather than strictly *in vitro* kinetic parameters, this experiment would help us best assess whether our evolved variants successfully outperformed our starting material and Cn Mdh2 CT4-1. Using C13-methanol feedstocks and a xylose-feeding strategy known to establish Mdh activity as the rate-determining step for methanol assimilation into central metabolism[140], we compared our three best variants to both the wild-type Bm Mdh2 and Cn Mdh2 CT4-1. By comparing the C13 incorporation across 4 metabolic intermediates from methanol assimilation, we found that all evolved variants we tested effectively double the amount of methanol assimilation in *E. coli* compared to both wild-type Bm Mdh2 and Cn Mdh2 CT4-1 (**Figure 48**). Interestingly, Bm Mdh2 seems to perform as well as Cn Mdh2 CT4-1

despite the large disparity in their *in vitro* activities, perhaps explaining why our evolved Bm Mdh2 derivatives—which themselves show higher *in vitro* activity than the wild-type enzyme, but not for Cn Mdh2 CT4-1—are significantly better *in vivo* than Cn Mdh2 CT4-1. Our FrmR reporter system consistently supports this trend when comparing evolved Bm Mdh2 to Cn Mdh2 CT4-1, but still corroborates the assumed inferiority of Bm Mdh2 wild-type to Cn Mdh2 CT4-1 that we didn't observe in our C13 incorporation data. The reason for the discrepancies between these different *in vivo* assays remains unclear.

Crucially, all our data show little-to-no benefit from combining these different mutations in a single enzyme. The reason these combinations did not occur during the selection may simply be that none of them represented a better solution than any one individual mutation. This is partially supported by the PACE sequencing data for the Plate-PANCE lagoon shown in **Figure 41C,** which suggests the "solution" to our selection was effectively any single mutation characterized, but none in combination. It thus remains unclear whether there exists a solution beyond what we have reported in this study that does not require significantly altering the sequence space of the enzyme, creating chimeric fusions with another protein, starting with a different set of alcohol dehydrogenases altogether—some of which have been previously shown to have significant methanol oxidation activity[98].

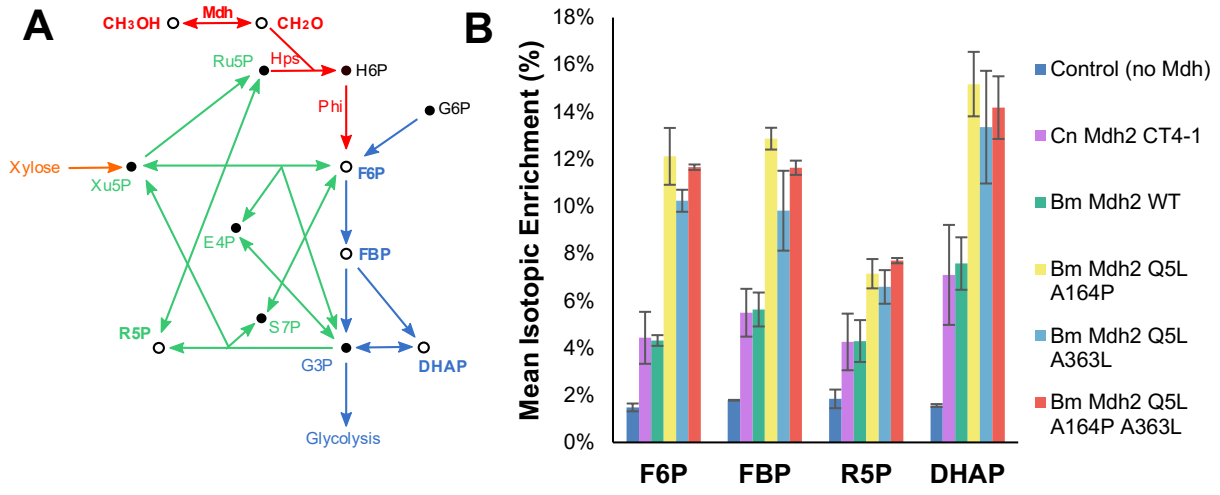


Figure 48: Incorporation of C13-labeled Methanol Feedstocks is Significantly Enhanced by Evolved Bm mdh2 Compared to the State of the Art.
A. Metabolic network map showing the path of C13-labeled methanol through hps-phi and subsequent assimilation into central metabolism. Red lines show assimilation pathway reactions, blue lines show glycolysis reactions, and green lines show pentose-phosphate pathway reactions. Empty circle nodes and corresponding bolded metabolites downstream of formaldehyde represent measured metabolites in labeling experiment analysis. **B.** C13 enrichment for assimilation pathways using different Mdh enzymes under equivalent Xylose-fed conditions. Bars show the average and errors bars the standard deviation across 3 biological replicates. F6P, Fructose 6-Phosphate; FBP, Fructose 1,6-Bisphosphate; R5P, Ribose 5-Phosphate; DHAP, Dihydroxyacetone Phosphate; H6P, D-arabino-3-Hexulo-6-Phosphate; G6P, Glucose 6-Phosphate; Ru5P, Ribulose 5-Phosphate; G3P, Glyceraldehyde 3-Phosphate; E4P, Erythrose 4-Phosphate; S7P, Sedoheptulose 7-Phosphate; Xu5P, Xylulose 5-Phosphate.

Section 3.9: Homology models suggest regions of interest for further Mdh engineering

To better hypothesize why we observed so many individual mutations in separate PANCE populations that all independently improve activity, we constructed a homology model of *B. methanolicus* Mdh2 with close alcohol dehydrogenase homologues using the Phyre 2 software developed by Kelley et al[141]. Bm Mdh2 has strong (>40%) sequence homology with many alcohol dehydrogenases, but few, if any, that are annotated as methanol dehydrogenases. We aligned our homology model to the *Zymomonas mobilis* ZM4 ADH2 alcohol dehydrogenase[142], which allowed us to approximate the location of the NAD+ molecule within the active site of our homology model. We then aligned this to a known, NAD-dependent alcohol dehydrogenase to predict the cofactor binding pocket of Bm Mdh2 and highlighted our mutated residues (**Figure 49**). We also highlighted a selection of residues found in cheaters to see if there was a consistent explanation for this FrmR-

binding phenotype. All our key mutated residues (E123, M163, A164 and A363) localize to the same general region proximal to the NAD cofactor binding pocket, close to where the expected redox-active region of the cofactor would be. Interestingly, E123, M163 and A164 all localize to one domain, while A363 localizes to a second domain on the other side of this binding pocket. Highlighting the four mutated residues we tested, it becomes clear that the residues map close to the redox-active end of the NAD molecule where we would expect the substrate methanol to bind, suggesting they play a role in association of one or more of the enzyme's substrates to this active site. However, they don't bind immediately next to residues predicted to bind to the zinc ion required for oxidation, which are predicted to sit on the opposite side of this NAD molecule based on the iron-coordinating site on the aligned ZM4 ADH2[142]. Residues associated with cheaters observed in the selection typically map to the outer face of the domain containing the E123, M163 and A164 mutations, which may help support why their luciferase reporter signals appeared to be much higher in combination with one another than we observed *in vitro* through cell-free lysate studies. Interestingly, an alignment of Cn Mdh2 CT4-1 to the same ZM4 ADH2 shows cheater residues evolving on the outside face of the opposite domain (**Figure 49B**), though similarly they appear to be primarily surface-exposed residues, corroborating our hypothesis that the method of cheating is a direct protein-protein interaction with FrmR. Highlighting the beneficial mutations for these enzymes, most notably the M163/A164 residues at Bm Mdh2 and the A169 residue for Cn Mdh2 cleanly shows the similarities in these improvements and further suggests this region of homologous enzymes as a key position in modulating methanol oxidation activities. Site-saturation mutagenesis of these residues and other residues near them may represent a more straight-forward path toward further improvements to Bm Mdh2, and the similarity of these residues to the improvements seen in Cn Mdh2 suggests that other methanol dehydrogenases could also benefit from targeted mutagenesis within this region.

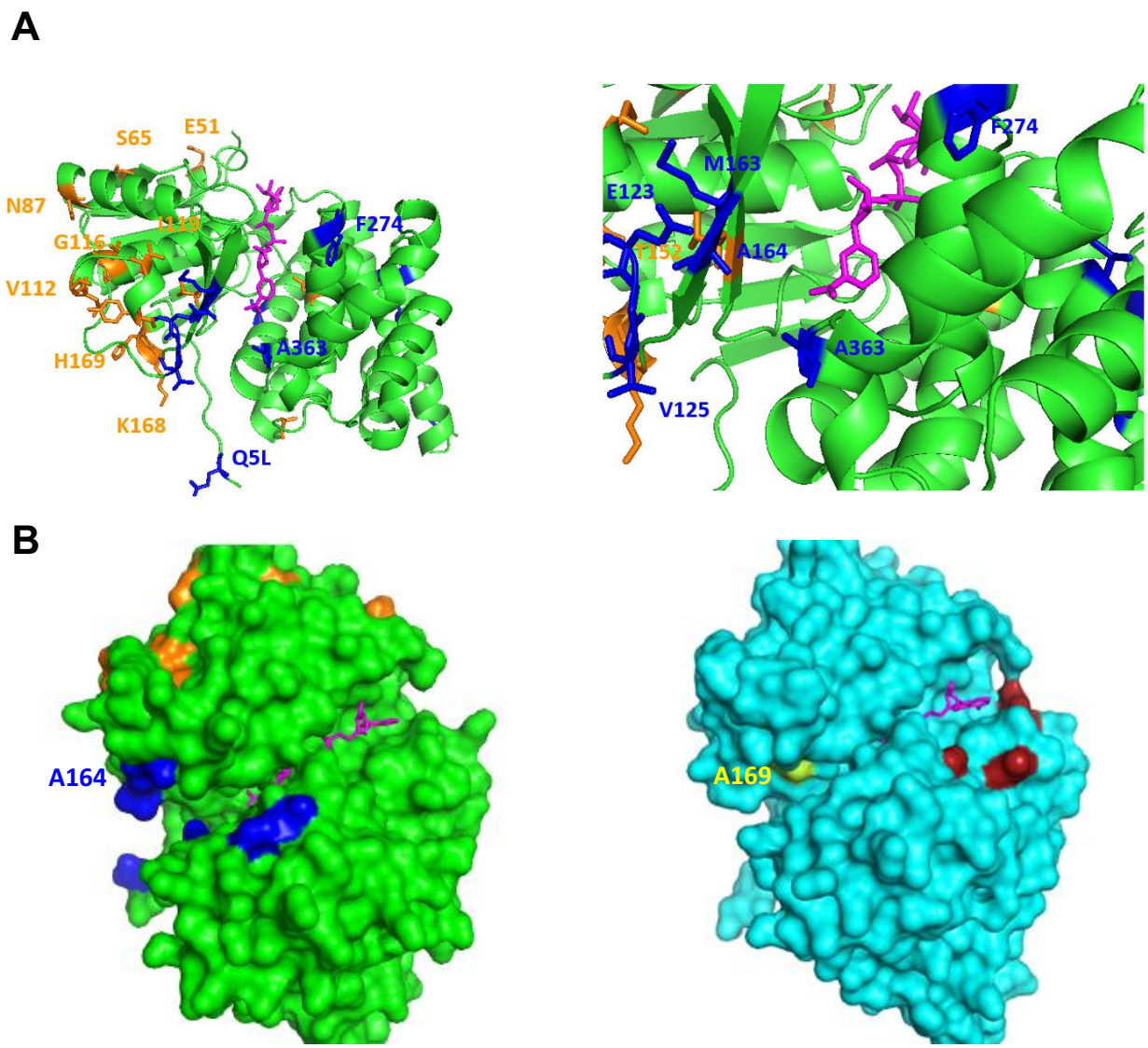


Figure 49: Representative Mutations to Bm mdh2 show possible mechanisms of improvement and cheating. **A.** Homology model of Bm Mdh2 aligned to ZM4 ADH2[142]. Residues highlighted in blue were observed in individuals from evolving populations that exhibited higher activities in FrmR-reporter assays and/or crude lysates and subsequent *in vitro* characterizations. Residues in orange represent a subset of mutations seen in cheater genotypes. NAD is shown in pink. **B.** Surface-display models of Bm Mdh2 (left) and Cn Mdh2 CT4-1 (right) showing beneficial residues in blue or yellow and cheater residues in orange or red.

Section 3.10: Discussion

Despite many unforeseen difficulties with cheater phenotypes, our results clearly indicate that we successfully evolved variants of *B. methanolicus* Mdh2 with significantly improved *in vitro* and *in vivo* activities. We also demonstrated that despite our assumption that Cn Mdh2 CT4-1 represented the state-of-the-art for Mdh enzymes used for synthetic methylotrophy based on its

superior *in vitro* kinetic parameters, the Bm Mdh2 enzyme shows roughly equivalent performance in terms of flux through methanol assimilation pathways. This suggests that *in vivo* activity might not always correlate with *in vitro* kinetics parameters. One possible explanation for this is the pH difference between the intracellular environment and the typical assay conditions used when comparing Mdh activities *in vitro* (pH 9.5). Further *in vitro* studies might consider testing the assumed assay conditions more thoroughly so as not to bias Mdh selection based on data not immediately relevant to the goals of synthetic methylotrophy *in vivo*.

To this end, it seems clear that had we not run into such a strong cheater phenotype after our initial evolutions, an *in vivo* selection for Mdh activity is poised to be the perfect solution to this engineering problem. In 5 separate populations evolved in PANCE, we observed clear, orthogonal mutations that each improve Mdh activity. Future work with our PANCE system should focus on improving the negative selection to better prevent these cheaters from sweeping the population. Given the clear limitations that the extraordinarily high K_m values for these dehydrogenases represent for improving their kinetic parameters, gaining the ability to run our selections at low concentrations of methanol without evolving cheaters would enable us to push these enzymes further and further toward improved activity. Selection techniques like FACS may provide a better solution to the cheater problem, making counter-selection against activity in the absence of substrate far simpler to gate out than in our phage-assisted methods. As we mentioned earlier, from our homology models it seems reasonable to assume that rational approaches might be an attractive alternative as well, since we now have a clearer picture of which residues are poised to impact activity the most for a given alcohol dehydrogenase with high sequence homology to Cn Mdh2 and Bm Mdh2. Whatever the approach may be, we've shown that these Mdh activities can be improved via *in vivo* phage-assisted selections, with evolved variants showing the desired phenotype of increased flux through assimilation pathways.

Section 3.11: Materials and methods

Reagents

Unless specified, all chemical reagents were purchased in the highest grade available from Sigma–Aldrich. For luciferase reporter assays, M9 salts and LB, 2xYT, Agar, and casamino acids were purchased from US Biological Life Sciences. Trace Elements (MD-TMS) and Vitamin Solution(MD-VS) were purchased from ATCC. Antibiotics and Arabinose were purchased from GoldBioTechnology Inc.

Strains and plasmids

Invitrogen Mach 1 T1R (Invitrogen) or NEB Turbo (NEB) chemically competent *E. coli* strains were used as cloning hosts. Luciferase reporter assays, phage-based assays, and all evolutions were carried out using E. coli S1030[8]. The original plasmid containing the *C. necator phaABCP* cassette from which the SP21 *phaC*-encoding SP was cloned (pMC001579) was gifted by Michelle Chang's Research group at University of California Berkeley. pETM6-mCherry was a gift from Mattheos Koffas (Addgene plasmid #66534).

Cloning

Plasmids and selection phage were constructed using USER cloning or KLD Enzyme Mix (NEB, Ipswich, MA). DNA fragments were generated by PCR using Pfu Turbo Cx Hotstart DNA Polymerase (Agilent), VeraSeq 2.0 High Fidelity DNA Polymerase (Enzymatics), or Phusion U Hot Start DNA Polymerase (Thermo Fisher Scientific). All amplicons were purified using kits from Qiagen and digested with DpnI during PCR fragment assembly for USER cloning. All restriction endonucleases and USER enzyme were purchased from NEB. Assembled vectors were transformed into chemically competent *E. coli* of various strains and verified by Sanger sequencing after

amplification from individual colonies using illustra TempliPhi DNA Amplification kits (GE Healthcare). Cell growth for cloning purposes was carried out using 2xYT media supplemented with appropriate antibiotics (Kanamycin, 50μgml^{-1}; Carbenicillin, 50μgml^{-1}; Spectinomycin, 100μgml^{-1}). For phage cloning and replication, phage were transformed directly into S1059, S1381 or an equivalent phage containing a PSP-gene III plasmid (e.g. pJC175e or similar)[8] and grown overnight. All transformed phage cultures were then plated on lawns of this same strain and individual plaques were picked into fresh media, grown for a minimum of 6 hours, and verified by Sanger sequencing after amplification using illustra TempliPhi DNA Amplification kits (GE Healthcare).

Luciferase-based reporter assays

For all reporter assays, the desired number of individual colonies from transformation plates were grown overnight at 37°C in Davis Rich Media (DRM). Cultures were diluted 100 to 1000-fold into fresh medium supplemented with anhydrotetracycline (ATc) if needed for FrmR expression from the pTR47m4 reporter plasmid and grown until early exponential phase (OD approximately 0.4). At this stage, arabinose was added to induce expression of Mdh from the P_{BAD} promoter, with full induction assumed at 1-10mM. For continuous monitoring of luminescence, 200μl of sample were placed into a black, Corning black clear bottom 96 well plate and analyzed for optical density at 600nm and luciferase activity at 37°C on an Infinite Pro M1000 plate reader (Tecan). For discrete timepoints, 150uL of culture was transferred to the same plates at given times post-induction and measured the same way. All liquid cultures and continuous cultures during plate-reading were grown with regular shaking or stirring.

Plaque Assays for Phage Titer Quantification

For plaque assays used to clone and titer phage, the desired host strain for plaquing was grown at 37°C to late log-phase (OD_{600} 0.6-1.2). 100-200μL of host culture was infected with 10μL

of a phage dilution series and diluted into 1mL of molten 0.75% (w/v) agar in 2xYT and immediately plated on 1.5% (w/v) agar in 2xYT media. Agar was cooled and set before plates were inverted and grown overnight at 37°C.

PACE Experiments

PACE experiments were performed as previously described[7–12,58]. All chemostat and lagoon systems were maintained using Masterflex Digital Pump systems (Cole-Parmer) at fixed RPM values manually calculated to provide a desired flow-rate for the tube diameter used during the experiment. TSS chemically competent *E. coli*[85] S1030 were transformed with desired AP and MP and plated on 2xYT agar containing 0.5-2% glucose (w/v). A single colony was grown to saturation overnight at 37°C in DRM containing appropriate antibiotics and diluted the next day 100- to 1000-fold into a chemostat at 37°C containing 50-100mL of Davis Rich Media supplemented with appropriate antibiotics for the AP/CP/MP used (Carbenicillin, 50μgml^{-1}; Chloramphenicol, 40μgml$^{-1;}$ Spectinomycin, 50μgml^{-1}). Once the chemostat reached an OD_{600} of ~0.8-1.2, dilution was started and adjusted in order to best maintain this OD_{600} range, which varied by host but typically fell in the range of 0.5-1.2 chemostat volumes per hour. Chemostat media was flowed into lagoons at a desired flow-rate. Lagoons were treated with 1M arabinose solution to 10mM final lagoon concentration pumped from a syringe pump (New Era Pump Systems) and 1.5mL per hour of a 5-20% Methanol and glutathione solution from a second syringe pump to achieve a desired final methanol concentration of 0.5-2% (v/v) methanol (125mM-500mM) and 5mM glutathione. Phage were injected into lagoons to start the evolution and collected from lagoon waste needles or waste lines at desired timepoints. Phage titers were determined by plaquing onto pJC175e-containing S1030 derivatives, typically S1381.

Sequencing data was collected by picking individual plaques into fresh media and growing overnight. Overnight cultures were spun down and the supernatant used as template material for

rolling circle amplification using illustra TempliPhi DNA Amplification kits (GE Healthcare). Sequences were determined by Sanger sequencing and results were aligned using SeqMan alignment software (DNAStar) and manually analyzed and recorded. All liquid cultures and were grown with continuous shaking or stirring.

PANCE Experiments

TSS chemically competent *E. coli*[85] S1030 were transformed with desired AP and MP and plated on 2xYT agar containing 0.5-2% glucose (w/v). A single colony was grown to saturation overnight at 37°C in DRM containing appropriate antibiotics and diluted the next day 100- to 1000-fold into fresh DRM. Cultures were grown to log-phase (OD 0.3-0.6), treated with 10mM arabinose to induce mutagenesis, the desired amount of anhydrotetracycline for a given passage (typically 0 or 40ng/mL), glutathione solution to a final concentration of 10mM, and the desired amount of methanol (0 to 500mM). Treated cultures were split into the desired number of either 2mL cultures in single culture tubes or 500uL cultures in a 96-well plate, and infected with selection phage. Infected cultures were grown overnight at 37°C and harvested the next day via centrifugation (max RCF for 2 min). Supernatant containing evolved phage was isolated with optional filtration through a 0.2μm Costar spin filter (Corning) and stored at 4°C. Isolated phage were then used to infect the next passage and the process repeated for however many passages were desired for the selection. Phage were diluted passage to passage a maximum of 10-fold and a minimum of 1000-fold. Phage titers were determined by plaquing onto pJC175e-containing S1030 derivative strain S1381.

Sequencing data was collected by picking individual plaques into fresh media and growing overnight. Overnight cultures were spun down and the supernatant used as template material for rolling circle amplification using illustra TempliPhi DNA Amplification kits (GE Healthcare). Sequences were determined by Sanger sequencing and results were aligned using SeqMan

alignment software (DNAStar) and manually analyzed and recorded. All liquid cultures and were grown with continuous shaking or stirring.

Crude Lysate Mdh Assays

TSS chemically competent *E. coli*[85] S1030 were transformed with desired pTR48 plasmid derivative encoding a given *mdh* gene and plated on 2xYT agar containing spectinomycin (100μgml^{-1}). Colonies were cultured overnight in DRM media containing spectinomycin (50μgml^{-1}), then diluted 100-fold into fresh media in the morning, and incubated with shaking (200 RPM) at 37°C until OD 0.5-0.6. Mdh expression was induced with 10mM arabinose, and growth continued for another two hours. Cells were harvested by centrifugation (20,000 g, room temperature, 5 minutes), washed once with PBS, and the pellets frozen at -20C. Lysis was achieved with B-PER complete (0.1 mL BPER per 1 mL culture, room termperature, 15 minutes), and the soluble fraction isolated by centrifugation (20,000g, 4°C, 20 minutes) and stored on ice until used. Methanol dehydrogenase activity was measured by following the methanol-dependent reduction of NAD+ at 340 nm in 250 μL final volume using a clear, flat-bottom 96 well plate and SpectraMax model M2e plate reader with SoftMax Pro 6.5 software. The assay consisted of 100 mM glycine-KOH (pH 9.5), 5 mM MgSO4 and 1 mM NAD, and was initiated by the addition of methanol to a final concentration of 50mM or 500mM. Product formation at 37°C was quantified in comparison to a NADH standard curve. Total protein concentration in crude lysate was determined by Pierce BCA Assay (ThermoFisher). Values are reported as arbitrary units, with one unit defined as 1 μmol NADH per minute normalized to total lysate protein concentration.

Mdh and Act Purification

pET vectors encoding 6x-His-Mdh genes were transformed into BL21* (de3) Chemically Competent *E. coli* (Invitrogen) and plated on LB Agar with 50μg/mL kanamycin at 37°C. A single

colony of each variant was inoculated into 2mL of LB media with 50µg/mL kanamycin and grown overnight at 37°C. Overnight cultures diluted 1000-fold into to 200mL of LB with 50µg/mL kanamycin, 20mM $MgSO_4$, and 100µM $ZnCl_2$ and grown to OD 0.6-1.2, induced with 0.1mM IPTG, and grown overnight at 22°C (220 RPM). Cultures were spun down at 10000g for 15 minutes at 4°C, decanted, and resuspended in 3mL B-PER reagent treated with EDTA-free protease inhibitor (Roche cOmplete™, Mini, EDTA-free) and left at room temperature for 30-60 minutes. Lysed cells were spun down for 20 minutes at 20000g at 4°C. Soluble lysate was decanted directly onto 2.5mL of Ni-NTA resin (5mL of a 50% solution washed with PBS, pH 7.4 with 10mM imidazole to remove all ethanol). Flow through was collected by gravity and/or light application of vacuum to speed up the initial flow. The column was then washed with 2-10mL of ice cold wash buffer (PBS pH 7.4 with 25mM imidazole), followed by elution with 2mL ice cold elution buffer (PBS pH 7.4 with 250mM imidazole). Elution fractions were checked via SDS-PAGE gel, pooled, and transferred to Amicon 3kDa CO Spin Columns and concentrated. Proteins were exchanged a total of 5 times into Tri-HCl (pH 7.5) to remove imidazole before overnight storage at 4°C and kinetics analysis the following day, with a final estimated imidazole concentration of no more than 1mM. Final protein concentrations were determined via Pierce BCA Protein Assay (ThermoFisher). 1ug of each enzyme was analyzed via SDS-PAGE using a Bolt 4-12% Bis-Tris Plus Gel (ThermoFisher) and visualized with Instant Blue stain (Expedeon) to verify the purity and size of each enzyme. Enzyme not used for kinetics was flash-frozen in liquid nitrogen and stored at -80°C.

Mdh Kinetics

NAD+ was added to reaction buffer (100 mM glycine-NaOH (pH 9.5), 50mM $MgSO_4$), and was incubated for 5 minutes with about 10ug of Mdh enzyme at 37°C. The methanol dehydrogenase reaction was initiated by the addition of methanol to a final concentration of 0mM, 10mM, 25mM, 50mM, 62.5mM, 125mM, 250mM, 500mM, 750mM, 1000mM, 1500mM, or 2000mM at a final

volume of 100uL in a black, Corning black clear bottom 96 well plate. Product formation at 37°C was quantified in comparison to an NADH standard curve prepared in reaction buffer. Total protein concentrations were re-confirmed by Pierce BCA Protein Assay (ThermoFisher). We note here that 50mM $MgSO_4$ is ten times more concentrated than typical assay conditions, but this did not noticeably alter our observed reaction rates. Reactions were all blanked to a well containing reaction buffer with 1mM NAD+. Initial velocities were determined from the slope of a plot of the calculated concentration of NADH based on absorbance at 340nm. Steady-state kinetic parameters were calculated by fitting these velocities to the Michaelis-Menten equation using gnuplot software.

13C Labeling and Analysis

Freshly transformed colonies were precultured overnight in Medium A (LB + 5g/L xylose, 20mM $MgSO_4$, 100μM $ZnCl_2$, 50 ug/mL kanamycin), then diluted 100-fold into fresh Medium A and incubated with shaking (200 RPM) at 37C until OD 0.5-0.6. Pathway enzyme expression was induced with 100μM IPTG, and growth continued for another two hours. The culture was then centrifuged (10 min at 3500g), washed once with an equal volume of M9 medium, and resuspended in an equal volume of M9 containing 5 g/L D-xylose, and 250mM 13C methanol (Cambridge Isotope Labs, 99%). Cells were incubated with shaking for a further hour before intracellular metabolites were extracted as described previously[143]. Briefly, 1 mL of liquid culture was filtered through a 0.45μm nylon filter, washed with 10mL room-temperature water, and then the filter was transferred into a 50mL falcon tube containing 5mL of extraction solution (40:40:20 acetonitrile:methanol:water) at -20°C. After 30 minutes, the filter was removed, the samples were centrifuged, and the supernatants dried overnight under air. The next morning, dried metabolites were resuspended in 150μL water, centrifuged at 13000 RPM for 40 minutes, and injected into an LC-MS/MS system as previously described[143].

Conclusions

While many aspects of process chemistry will need to be addressed to fully implement microbial fermentation toward issues such as methane assimilation (for which capture of methane and conversion to methanol remain significant barriers) or plastics production (for which isolation of intracellular polymer granules is typically inefficient), better yields from microbial systems will never hurt the odds of success. As a result, metabolic engineering will likely continue to grow as an industry over the next few decades as DNA synthesis and microbial engineering technologies become more robust and cost-effective. However, there is never a guarantee that all enzymatic activities present in nature will meet the demands of industry, both in terms of their catalytic rate and chemical specificity. Screens and selections for engineering these enzymes will remain invaluable for the foreseeable future and help us continue to elucidate how to better rationally design these catalysts for a variety of applications.

Industry has many cutting-edge options for selections, but smaller laboratories are not always equipped to use the state-of-the-art techniques. Phage Assisted Non-Continuous Evolution (PANCE) as described in this work is hardly the most robust or comprehensive enzymatic selection technique, but it represents a highly-accessible, in-cell option for any activities that can be connected to genetic regulation. The technical requirements for PANCE are essentially the bare-minimum for any microbiology lab, and the elimination of library construction from the selection pipeline removes the need for specialized kits, enzymes, primers, or competent host cells. For the initial stages of metabolic engineering research, particularly for younger labs with smaller resource pools, having a widely accessible, low-cost method for evolving pathway components may provide a new route toward small-scale pathway optimization compared to the standard synthetic biology toolkit currently available. Even for larger labs, such cheap and simple techniques can offer a useful parallel to standard procedures, helping diversify the range of selection conditions sampled. Most

of my plate-based PANCE selections were run without routine monitoring of population size, and while this led to some number of populations washing out or evolving cheaters, the populations that survived the selection still yielded valuable results; thus, rather than serving as a replacement for existing techniques, I see PANCE as a complement or starting point for more expensive and involved approaches such as FACS or plate-based screens coupled to direct chemical analysis. A researcher can maintain dozens of PANCE populations without devoting more than an hour or two each day to the process, and as the technique gets refined for future selections, I anticipate that innovators in the field will find improved methods for assessing phage populations quickly and efficiently compared to the phage titer assays and PCR techniques I report in this work.

Compared to PACE, PANCE is a much easier platform to parallelize, as I showed in both Chapters 1 and 3. Increasing the number of populations able to be evolved in parallel not only enables evolution of many target selection phage with replicate populations—which I showed was a valuable approach given the variety of divergent mutations achieved in the Mdh evolution—but also allows for the simultaneous evolution of identical selection phage in a variety of host conditions, as I did when evolving PhaC stereochemical changes. For metabolic phenotypes, PANCE also reduces the relatively high burden of synthetic biology work required to tune gene III expression to fit the kinetics of the starting genes of interest. For PHB pathways, for example, full pathway expression from phage often failed to show robust propagation until 6 to 8 hours after infection, resulting in washout in continuous flow, but significant population expansion for each PANCE passage. Furthermore, the lowered stringency enables the use of less rich growth media, helping keep selection conditions closer to tractable growth conditions and removing noise from background metabolite production. This parallelization hedges bets against the most common modes of selection failure, including improper tuning of selection conditions (fed substrate concentrations, AP stringency, etc.) and gene III recombinant cheaters.

There are still many limitations to PANCE, not the least of which is the need to build a genetic circuit in response to a metabolic phenotype. For things like feedstock assimilation, simpler selections based on growth rate might remain a more straightforward path, whereas for specific chemical catalysts used in drug synthesis pipelines or similar, it is highly unlikely that a sensor exists that would allow PANCE to be used in place of direct screens for desired chemical outputs. Even if a selection circuit exists for a desired pathway product, my Mdh evolution results show the wide-reaching potential for cheater phenotypes, a caveat that requires much closer examination for any future selections using this technique. PANCE is further limited in the size of the target evolutionary transcript based on the permissible size of the selection phage genome, but any selection seeking to sample a large library of mutations will eventually run into a similar issue and I doubt this problem will ever find a solution outside of improved methods in rational protein engineering. Despite these limitations, metabolic engineering strategies that can connect an activity of interest to a well-characterized output molecule with a known, robust biosensor could help circumvent many of these issues. As an example, using PHA synthesis pathways that accept long-chain 3-hydroxyacyl-CoA substrates in a host strain would allow for polymerization and subsequent detection of fatty acid products. Compared to direct detection of fatty acids in the media, this method would sequester outputs of interest to an intracellular inclusion, helping ensure more accurate resolution of individual variants in a population. In this manner, my hope is that researchers in metabolic engineering will consider not just products with known biosensors, but also the capacity for biosensors of downstream byproducts as proxies for direct detection. This would greatly expand the range of accessible targets for PANCE and open the door for more generalizable selection platforms based on this technique while retaining its cheap and accessible properties. The growing field of biosensor design will also play a crucial role in improving options for this and other selections, allowing us to sense an increasing number of molecules *in vivo*.

One of the crucial components missing from this thesis is a comparison of PACE and PANCE to alternative techniques based on biosensor readouts in evolving populations. In the future, I hope that one or more labs will consider comparing PANCE to well-established techniques such as FACS or plate-based screens to better speak to the unique advantages of each. However, even if established techniques prove to be more effective than PANCE, the niche for low-cost, low-expertise techniques in biology will likely always exist, whether for pedagogical purposes or simply to offset the inefficient distribution of research funds. In my opinion, it is paramount for researchers in this field to consider not just how to maximize their approach in terms of output per unit time, but also in terms of financial investment and technological accessibility. Even compared to PACE and established plate-based screens, PANCE is one of the cheapest and most accessible selection techniques I have encountered, and I hope that the increasing spread and influence of synthetic biology techniques will only improve its utility to all researchers in the field.

References

1. Wang, J. *et al.* Revealing a 5,000-y-old beer recipe in China. *PNAS* **113,** 6444–6448 (2016).

2. Koh, A. J., Yasur-Landau, A. & Cline, E. H. Characterizing a Middle Bronze Palatial Wine Cellar from Tel Kabri, Israel. *PLOS ONE* **9,** e106406 (2014).

3. Clark, J. H., Luque, R. & Matharu, A. S. Green Chemistry, Biofuels, and Biorefinery. *Annual Review of Chemical and Biomolecular Engineering* **3,** 183–207 (2012).

4. Woolston, B. M., Edgar, S. & Stephanopoulos, G. Metabolic Engineering: Past and Future. *Annual Review of Chemical and Biomolecular Engineering* **4,** 259–288 (2013).

5. Nielsen, J. & Keasling, J. D. Engineering Cellular Metabolism. *Cell* **164,** 1185–1197 (2016).

6. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16,** 379–394 (2015).

7. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472,** 499–503 (2011).

8. Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A. & Liu, D. R. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature Chemical Biology* **10,** 216–222 (2014).

9. Dickinson, B. C., Leconte, A. M., Allen, B., Esvelt, K. M. & Liu, D. R. Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *PNAS* **110,** 9007–9012 (2013).

10. Badran, A. H. *et al.* Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance. *Nature* **533,** 58–63 (2016).

11. Bryson, D. I. *et al.* Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **13,** 1253–1260 (2017).

12. Packer, M. S., Rees, H. A. & Liu, D. R. Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nature Communications* **8,** 956 (2017).

13. Lemoigne, Maurice. Produit de déshydratation et de polymérisation de l'acide b-oxybutyrique. *Bulletin De La Societe De Chimie Biologique* **8,** 770–782 (1926).

14. Rehm, B. H. A., Mitsky, T. A. & Steinbüchel, A. Role of Fatty Acid De Novo Biosynthesis in Polyhydroxyalkanoic Acid (PHA) and Rhamnolipid Synthesis by Pseudomonads: Establishment of the Transacylase (PhaG)-Mediated Pathway for PHA Biosynthesis inEscherichia coli. *Appl. Environ. Microbiol.* **67,** 3102–3109 (2001).

15. Peoples, O. P. & Sinskey, A. J. Poly-beta-hydroxybutyrate biosynthesis in Alcaligenes eutrophus H16. Characterization of the genes encoding beta-ketothiolase and acetoacetyl-CoA reductase. *J. Biol. Chem.* **264,** 15293–15297 (1989).

16. Peoples, O. P. & Sinskey, A. J. Poly-beta-hydroxybutyrate (PHB) biosynthesis in Alcaligenes eutrophus H16. Identification and characterization of the PHB polymerase gene (phbC). *J. Biol. Chem.* **264,** 15298–15303 (1989).

17. Qi, Q., Rehm, B. H. & Steinbüchel, A. Synthesis of poly(3-hydroxyalkanoates) in Escherichia coli expressing the PHA synthase gene phaC2 from Pseudomonas aeruginosa: comparison of PhaC1 and PhaC2. *FEMS Microbiol. Lett.* **157,** 155–162 (1997).

18. Gross, R. A., DeMello, C., Lenz, R. W., Brandl, H. & Fuller, R. C. The biosynthesis and characterization of poly(β-hydroxyalkanoates) produced by Pseudomonas oleovorans. *Macromolecules* **22,** 1106–1115 (1989).

19. Preusting, H., Nijenhuis, A. & Witholt, B. Physical characteristics of poly(3-hydroxyalkanoates) and poly(3-hydroxyalkenoates) produced by Pseudomonas oleovorans grown on aliphatic hydrocarbons. *Macromolecules* **23,** 4220–4224 (1990).

20. Matsusaki, H. *et al.* Cloning and Molecular Analysis of the Poly(3-hydroxybutyrate) and Poly(3-hydroxybutyrate-co-3-hydroxyalkanoate) Biosynthesis Genes in Pseudomonas sp. Strain 61-3. *J. Bacteriol.* **180,** 6459–6467 (1998).

21. Taguchi, S. *et al.* A microbial factory for lactate-based polyesters using a lactate-polymerizing enzyme. *PNAS* **105,** 17323–17327 (2008).

22. Hori, C., Oishi, K., Matsumoto, K., Taguchi, S. & Ooi, T. Site-directed saturation mutagenesis of polyhydroxylalkanoate synthase for efficient microbial production of poly[(R)-2-hydroxybutyrate]. *Journal of Bioscience and Bioengineering* (2018).

23. Li, Z.-J., Qiao, K., Che, X.-M. & Stephanopoulos, G. Metabolic engineering of Escherichia coli for the synthesis of the quadripolymer poly(glycolate-co-lactate-co-3-hydroxybutyrate-co-4-hydroxybutyrate) from glucose. *Metabolic Engineering* **44,** 38–44 (2017).

24. Choi, S. Y. *et al.* One-step fermentative production of poly(lactate-*co*-glycolate) from carbohydrates in *Escherichia coli*. *Nature Biotechnology* **34,** nbt.3485 (2016).

25. Matsumoto, K., Shiba, T., Hiraide, Y. & Taguchi, S. Incorporation of Glycolate Units Promotes Hydrolytic Degradation in Flexible Poly(glycolate-co-3-hydroxybutyrate) Synthesized by Engineered Escherichia coli. *ACS Biomater. Sci. Eng.* (2016).

26. Song, S., Ma, H., Gao, Z., Jia, Z. & Zhang, X. [Construction of recombinant Escherichia coli strains producing poly (4-hydroxybutyric acid) homopolyester from glucose]. *Wei Sheng Wu Xue Bao* **45,** 382–386 (2005).

27. Zhou, X.-Y. *et al.* Hyperproduction of poly(4-hydroxybutyrate) from glucose by recombinant Escherichia coli. *Microbial Cell Factories* **11,** 54 (2012).

28. Valentin, H. E. & Dennis, D. Production of poly(3-hydroxybutyrate-co-4-hydroxybutyrate) in recombinant Escherichia coli grown on glucose. *Journal of Biotechnology* **58,** 33–38 (1997).

29. Jendrossek, D. Polyhydroxyalkanoate Granules Are Complex Subcellular Organelles (Carbonosomes). *J Bacteriol* **191,** 3195–3202 (2009).

30. Pötter, M., Madkour, M. H., Mayer, F. & Steinbüchel, A. Regulation of phasin expression and polyhydroxyalkanoate (PHA) granule formation in Ralstonia eutropha H16. *Microbiology (Reading, Engl.)* **148,** 2413–2426 (2002).

31. York, G. M., Stubbe, J. & Sinskey, A. J. The Ralstonia eutropha PhaR Protein Couples Synthesis of the PhaP Phasin to the Presence of Polyhydroxybutyrate in Cells and Promotes Polyhydroxybutyrate Production. *J. Bacteriol.* **184,** 59–66 (2002).

32. Eggers, J. & Steinbuchel, A. Impact of Ralstonia eutropha's Poly(3-Hydroxybutyrate) (PHB) Depolymerases and Phasins on PHB Storage in Recombinant Escherichia coli. *Applied and Environmental Microbiology* **80,** 7702–7709 (2014).

33. Sudesh, K., Maehara, A., Gan, Z., Iwata, T. & Doi, Y. Direct observation of polyhydroxyalkanoate granule-associated-proteins on native granules and on poly(3-hydroxybutyrate) single crystals by atomic force microscopy. *Polymer Degradation and Stability* **83,** 281–287 (2004).

34. Yamada, M., Wakuda, A. & Taguchi, S. Morphological Change in Cellular Granule Formation of Poly[(*R*)-3-hydroxybutyrate] Caused by DNA-Binding-Related Mutations of an Autoregulated Repressor PhaR. *Bioscience, Biotechnology, and Biochemistry* **71,** 1572–1576 (2007).

35. Mezzina, M. P. & Pettinari, M. J. Phasins, Multifaceted Polyhydroxyalkanoate Granule-Associated Proteins. *Appl. Environ. Microbiol.* **82,** 5060–5067 (2016).

36. Maehara, A., Taguchi, S., Nishiyama, T., Yamane, T. & Doi, Y. A Repressor Protein, PhaR, Regulates Polyhydroxyalkanoate (PHA) Synthesis via Its Direct Interaction with PHA. *Journal of Bacteriology* **184,** 3992–4002 (2002).

37. Chou, M.-E. & Yang, M.-K. Analyses of binding sequences of the PhaR protein of Rhodobacter sphaeroides FJ1: PhaR-binding sequence of Rhodobacter. *FEMS Microbiology Letters* **302,** 138–143 (2010).

38. Eugenio, L. I. de *et al.* Biochemical Evidence That phaZ Gene Encodes a Specific Intracellular Medium Chain Length Polyhydroxyalkanoate Depolymerase in Pseudomonas putida KT2442. *J. Biol. Chem.* **282,** 4951–4962 (2007).

39. Sznajder, A. & Jendrossek, D. To Be or Not To Be a Poly(3-Hydroxybutyrate) (PHB) Depolymerase: PhaZd1 (PhaZ6) and PhaZd2 (PhaZ7) of Ralstonia eutropha, Highly Active PHB Depolymerases with No Detectable Role in Mobilization of Accumulated PHB. *Appl. Environ. Microbiol.* **80,** 4936–4946 (2014).

40. Knoll, M., Hamm, T. M., Wagner, F., Martinez, V. & Pleiss, J. The PHA Depolymerase Engineering Database: A systematic analysis tool for the diverse family of polyhydroxyalkanoate (PHA) depolymerases. *BMC Bioinformatics* **10,** 89 (2009).

41. Martínez, V. *et al.* Identification and Biochemical Evidence of a Medium-Chain-Length Polyhydroxyalkanoate Depolymerase in the Bdellovibrio bacteriovorus Predatory Hydrolytic Arsenal. *Appl. Environ. Microbiol.* **78,** 6017–6026 (2012).

42. Amara, A., Steinbuchel, A. & Rehm, B. In vivo evolution of the Aeromonas punctata polyhydroxyalkanoate (PHA) synthase: isolation and characterization of modified PHA synthases with enhanced activity. *Applied Microbiology and Biotechnology* **59,** 477–482 (2002).

43. Kichise, T., Taguchi, S. & Doi, Y. Enhanced Accumulation and Changed Monomer Composition in Polyhydroxyalkanoate (PHA) Copolyester by In Vitro Evolution of Aeromonas caviae PHA Synthase. *Applied and Environmental Microbiology* **68,** 2411–2419 (2002).

44. Sheu, D.S., Chen, W. M., Lai, Y. W. & Chang, R.-C. Mutations Derived from the Thermophilic Polyhydroxyalkanoate Synthase PhaC Enhance the Thermostability and Activity of PhaC from Cupriavidus necator H16. *Journal of Bacteriology* **194,** 2620–2629 (2012).

45. Watanabe, Y. *et al.* Development and validation of an HPLC-based screening method to acquire polyhydroxyalkanoate synthase mutants with altered substrate specificity. *Journal of Bioscience and Bioengineering* **113,** 286–292 (2012).

46. Law, J. H. & Slepecky, R. A. Assay of Poly-B-Hydroxybutyric Acid. **82,** 33–36 (1961).

47. Braunegg, G., Sonnleitner, B. & Lafferty, R. M. A rapid gas chromatographic method for the determination of poly-β-hydroxybutyric acid in microbial biomass. *European J. Appl. Microbiol. Biotechnol.* **6,** 29–37 (1978).

48. Li, T. *et al.* Semirational Approach for Ultrahigh Poly(3-hydroxybutyrate) Accumulation in *Escherichia coli* by Combining One-Step Library Construction and High-Throughput Screening. *ACS Synthetic Biology* **5,** 1308–1317 (2016).

49. Taguchi, S. *et al.* Analysis of mutational effects of a polyhydroxybutyrate (PHB) polymerase on bacterial PHB accumulation using an in vivo assay system. *FEMS Microbiology Letters* **198,** 65–71 (2001).

50. Spiekermann, P., Rehm, B. H. A., Kalscheuer, R., Baumeister, D. & Steinbüchel, A. A sensitive, viable-colony staining method using Nile red for direct screening of bacteria that accumulate polyhydroxyalkanoic acids and other lipid storage compounds. *Archives of Microbiology* **171,** 73–80 (1999).

51. Zuriani, R., Vigneswari, S., Azizan, M. N. M., Majid, M. I. A. & Amirul, A. A. A high throughput Nile red fluorescence method for rapid quantification of intracellular bacterial polyhydroxyalkanoates. *Biotechnology and Bioprocess Engineering* **18,** 472–478 (2013).

52. Gorenflo, V., Steinbüchel, A., Marose, S., Rieseberg, M. & Scheper, T. Quantification of bacterial polyhydroxyalkanoic acids by Nile red staining. *Appl Microbiol Biotechnol* **51,** 765–772 (1999).

53. Opgenorth, P. H., Korman, T. P. & Bowie, J. U. A synthetic biochemistry molecular purge valve module that maintains redox balance. *Nature Communications* **5,** 4113 (2014).

54. Grage, K. *et al.* Bacterial Polyhydroxyalkanoate Granules: Biogenesis, Structure, and Potential Use as Nano-/Micro-Beads in Biotechnological and Biomedical Applications. *Biomacromolecules* **10,** 660–669 (2009).

55. Ma, H.-K. *et al.* Application of polyhydroxyalkanoate (PHA) synthesis regulatory protein PhaR as a bio-surfactant and bactericidal agent. *Journal of Biotechnology* **166,** 34–41 (2013).

56. Zhang, F., Carothers, J. M. & Keasling, J. D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology* **30,** 354–359 (2012).

57. Yamashita, K., Yamada, M., Numata, K. & Taguchi, S. Nonspecific Hydrophobic Interactions of a Repressor Protein, PhaR, with Poly[(*R*)-3-hydroxybutyrate] Film Studied with a Quartz Crystal Microbalance. *Biomacromolecules* **7,** 2449–2454 (2006).

58. Hubbard, B. P. *et al.* Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nature Methods* **12,** 939–942 (2015).

59. Davis, J. H., Rubin, A. J. & Sauer, R. T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Research* **39,** 1131–1141 (2011).

60. Yamada, M. *et al.* Autoregulator Protein PhaR for Biosynthesis of Polyhydroxybutyrate [P(3HB)] Possibly Has Two Separate Domains That Bind to the Target DNA and P(3HB): Functional Mapping of Amino Acid Residues Responsible for DNA Binding. *Journal of Bacteriology* **189,** 1118–1127 (2007).

61. Brissette, J. L., Weiner, L., Ripmaster, T. L. & Model, P. Characterization and sequence of the Escherichia coli stress-induced psp operon. *Journal of Molecular Biology* **220,** 35–48 (1991).

62. Okamura, E., Tomita, T., Sawa, R., Nishiyama, M. & Kuzuyama, T. Unprecedented acetoacetyl-coenzyme A synthesizing enzyme of the thiolase superfamily involved in the mevalonate pathway. *Proceedings of the National Academy of Sciences* (2010).

63. Bond-Watts, B. B., Bellerose, R. J. & Chang, M. C. Y. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nature Chemical Biology* **7,** nchembio.537 (2011).

64. Shen, C. R. *et al.* Driving forces enable high-titer anaerobic 1-butanol synthesis in Escherichia coli. *Appl. Environ. Microbiol.* **77,** 2905–2915 (2011).

65. Lan, E. I. & Liao, J. C. ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. *Proc Natl Acad Sci U S A* **109,** 6018–6023 (2012).

66. Karr, D. B., Waters, J. K. & Emerich, D. W. Analysis of poly-β-hydroxybutyrate in Rhizobium japonicum bacteroids by ion-exclusion high-pressure liquid chromatography and UV detection. *Applied and environmental microbiology* **46,** 1339–1344 (1983).

67. Badran, A. H. & Liu, D. R. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat Commun* **6,** 8425 (2015).

68. Walker, M. C. *et al.* Expanding the Fluorine Chemistry of Living Systems Using Engineered Polyketide Synthase Pathways. *Science* **341,** 1089–1094 (2013).

69. Normi, Y. M. *et al.* Characterization and Properties of G4X Mutants of Ralstonia eutropha PHA Synthase for Poly(3-hydroxybutyrate) Biosynthesis in Escherichia coli. *Macromolecular Bioscience* **5,** 197–206 (2005).

70. Normi, Y. M. *et al.* Site-directed saturation mutagenesis at residue F420 and recombination with another beneficial mutation of Ralstonia eutropha polyhydroxyalkanoate synthase. *Biotechnology Letters* **27,** 705–712 (2005).

71. Stubbe, J. & Tian, J. Polyhydroxyalkanoate (PHA) homeostasis: the role of the PHA synthase. *Natural Product Reports* **20,** 445 (2003).

72. Rehm, B. H. A., Antonio, R. V., Spiekermann, P., Amara, A. A. & Steinbu, A. Molecular characterization of the poly(3-hydroxybutyrate) (PHB) synthase from Ralstonia eutropha: in vitro evolution, site-specific mutagenesis and development of a PHB synthase protein model. *Biochimica et Biophysica Acta* 13

73. Wittenborn, E. C., Jost, M., Wei, Y., Stubbe, J. & Drennan, C. L. Structure of the Catalytic Domain of the Class I Polyhydroxybutyrate Synthase from Cupriavidus necator. *J. Biol. Chem.* jbc.M116.756833 (2016). doi:10.1074/jbc.M116.756833

74. Kim Yeo-Jin *et al.* Structure and function of the N-terminal domain of Ralstonia eutropha polyhydroxyalkanoate synthase, and the proposed structure and mechanisms of the whole enzyme. *Biotechnology Journal* **12,** 1600649 (2016).

75. Kim, J., Kim, Y.-J., Choi, S. Y., Lee, S. Y. & Kim, K.-J. Crystal structure of Ralstonia eutropha polyhydroxyalkanoate synthase C-terminal domain and reaction mechanisms. *Biotechnol. J.* **12,** 1600648 (2017).

76. Kourtz, L. *et al.* A novel thiolase-reductase gene fusion promotes the production of polyhydroxybutyrate in Arabidopsis. *Plant Biotechnology Journal* **3,** 435–447 (2005).

77. Zhang, F., Rodriguez, S. & Keasling, J. D. Metabolic engineering of microbial pathways for advanced biofuels production. *Current Opinion in Biotechnology* **22,** 775–783 (2011).

78. Dueber, J. E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nature Biotechnology* **27,** 753–759 (2009).

79. Hiroe, A., Tsuge, K., Nomura, C. T., Itaya, M. & Tsuge, T. Rearrangement of Gene Order in the phaCAB Operon Leads to Effective Production of Ultrahigh-Molecular-Weight Poly[(R)-3-Hydroxybutyrate] in Genetically Engineered Escherichia coli. *Applied and Environmental Microbiology* **78,** 3177–3184 (2012).

80. Steinbüchel, A. & Valentin, H. E. Diversity of bacterial polyhydroxyalkanoic acids. *FEMS Microbiology Letters* **128,** 219–228 (1995).

81. Reichardt, R. & Rieger, B. Poly(3-Hydroxybutyrate) from Carbon Monoxide. in *Synthetic Biodegradable Polymers* (eds. Rieger, B. et al.) **245,** 49–90 (Springer Berlin Heidelberg, 2011).

82. Blaisse, M. R., Dong, H., Fu, B. & Chang, M. C. Y. Discovery and Engineering of Pathways for Production of α-Branched Organic Acids. *J. Am. Chem. Soc.* **139,** 14526–14532 (2017).

83. Zheng, Z. *et al.* Mutation on N-terminus of polyhydroxybutyrate synthase of Ralstonia eutropha enhanced PHB accumulation. *Appl Microbiol Biotechnol* **72,** 896–905 (2006).

84. Wang, Q., Xia, Y., Chen, Q. & Qi, Q. Incremental truncation of PHA synthases results in altered product specificity. *Enzyme and Microbial Technology* **50,** 293–297 (2012).

85. Chung, C. T. & Miller, R. H. [43] Preparation and storage of competent Escherichia coli cells. in *Methods in Enzymology* (ed. Wu, R.) **218,** 621–627 (Academic Press, 1993).

86. Hiroe, A., Hyakutake, M., Thomson, N. M., Sivaniah, E. & Tsuge, T. Endogenous Ethanol Affects Biopolyester Molecular Weight in Recombinant *Escherichia coli*. *ACS Chemical Biology* **8,** 2568–2576 (2013).

87. Monographs on the evaluation of carcinogenic risk to humans: Formaldehyde, 2-butoxyethanol and 1-tert-butoxypropan-2-ol. *International Agency for Research on Cancer* **88,** 1–390 (2006).

88. Salthammer, T. Formaldehyde in the ambient atmosphere: from an indoor pollutant to an outdoor pollutant? *Angew. Chem. Int. Ed. Engl.* **52,** 3320–3327 (2013).

89. Chistoserdova, L., Kalyuzhnaya, M. G. & Lidstrom, M. E. The expanding world of methylotrophic metabolism. *Annu. Rev. Microbiol.* **63,** 477–499 (2009).

90. Cregg, J. M., Madden, K. R., Barringer, K. J., Thill, G. P. & Stillman, C. A. Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. *Mol Cell Biol* **9,** 1316–1323 (1989).

91. Müller, J. E. N., Heggeset, T. M. B., Wendisch, V. F., Vorholt, J. A. & Brautaset, T. Methylotrophy in the thermophilic Bacillus methanolicus, basic insights and application for commodity production from methanol. *Appl. Microbiol. Biotechnol.* **99,** 535–551 (2015).

92. Yurimoto, H., Oku, M. & Sakai, Y. Yeast Methylotrophy: Metabolism, Gene Regulation and Peroxisome Homeostasis. *International Journal of Microbiology* (2011). doi:10.1155/2011/101298

93. Kalyuzhnaya, M. G., Puri, A. W. & Lidstrom, M. E. Metabolic engineering in methanotrophic bacteria. *Metab. Eng.* **29,** 142–152 (2015).

94. Whitaker, W. B., Sandoval, N. R., Bennett, R. K., Fast, A. G. & Papoutsakis, E. T. Synthetic methylotrophy: engineering the production of biofuels and chemicals based on the biology of aerobic methanol utilization. *Current Opinion in Biotechnology* **33,** 165–175 (2015).

95. Kalász, H. Biological role of formaldehyde, and cycles related to methylation, demethylation, and formaldehyde production. *Mini Rev Med Chem* **3,** 175–192 (2003).

96. Yurimoto, H., Kato, N. & Sakai, Y. Assimilation, dissimilation, and detoxification of formaldehyde, a central metabolic intermediate of methylotrophic metabolism. *Chem Rec* **5,** 367–375 (2005).

97. Nash, T. The colorimetric estimation of formaldehyde by means of the Hantzsch reaction. *Biochem. J.* **55,** 416–421 (1953).

98. Müller, J. E. N. *et al.* Engineering Escherichia coli for methanol conversion. *Metabolic Engineering* **28,** 190–201 (2015).

99. Whitaker, W. B. *et al.* Engineering the biological conversion of methanol to specialty chemicals in Escherichia coli. *Metab. Eng.* **39,** 49–59 (2017).

100. Wu, T.-Y. *et al.* Characterization and evolution of an activator-independent methanol dehydrogenase from Cupriavidus necator N-1. *Appl Microbiol Biotechnol* **100,** 4969–4983 (2016).

101. US EPA, O. EPA Method 8315A (SW-846): Determination of Carbonyl Compounds by High Performance Liquid Chromatography (HPLC). *US EPA* (2015). Available at: https://www.epa.gov/homeland-security-research/epa-method-8315a-sw-846-determination-carbonyl-compounds-high-performance. (Accessed: 28th March 2018)

102. Zhang, J., Jensen, M. K. & Keasling, J. D. Development of biosensors and their application in metabolic engineering. *Current Opinion in Chemical Biology* **28,** 1–8 (2015).

103. Zhang, J. *et al.* Engineering an NADPH/NADP+ Redox Biosensor in Yeast. *ACS Synth. Biol.* **5,** 1546–1556 (2016).

104. Zhang, J. *et al.* Development of a Transcription Factor-Based Lactam Biosensor. *ACS Synth. Biol.* **6,** 439–445 (2017).

105. Raman, S., Rogers, J. K., Taylor, N. D. & Church, G. M. Evolution-guided optimization of biosynthetic pathways. *PNAS* **111,** 17803–17808 (2014).

106. Brockman, I. M. & Prather, K. L. J. Dynamic metabolic engineering: New strategies for developing responsive cell factories. *Biotechnol J* **10,** 1360–1369 (2015).

107. Gonzalez, C. F. *et al.* Molecular Basis of Formaldehyde Detoxification: Characterization of two S-Formylglutathione Hydrolases from Escherichia coli, FrmB and YeiG. *Journal of Biological Chemistry* **281,** 14514–14522 (2006).

108. Denby, K. J. *et al.* The mechanism of a formaldehyde-sensing transcriptional regulator. *Scientific Reports* **6,** 38879 (2016).

109. Law, J. *Molecular Basis of Bacterial Formaldehyde Sensing*. (University of Manchester, 2012).

110. Osman, D. *et al.* The Effectors and Sensory Sites of Formaldehyde-responsive Regulator FrmR and Metal-sensing Variant. *J. Biol. Chem.* **291,** 19502–19516 (2016).

111. Herring, C. D. & Blattner, F. R. Global Transcriptional Effects of a Suppressor tRNA and the Inactivation of the Regulator frmR. *J. Bacteriol.* **186,** 6714–6720 (2004).

112. Tralau, T. *et al.* An internal reaction chamber in dimethylglycine oxidase provides efficient protection from exposure to toxic formaldehyde. *J. Biol. Chem.* **284,** 17826–17834 (2009).

113. Rohlhill, J., Sandoval, N. R. & Papoutsakis, E. T. Sort-Seq Approach to Engineering a Formaldehyde-Inducible Promoter for Dynamically Regulated Escherichia coli Growth on Methanol. *ACS Synth Biol* **6,** 1584–1595 (2017).

114. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405,** 590–593 (2000).

115. Iwig, J. S. & Chivers, P. T. DNA Recognition and Wrapping by Escherichia coli RcnR. *Journal of Molecular Biology* **393,** 514–526 (2009).

116. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79–88 (2006).

117. Irla, M. *et al.* Transcriptome analysis of thermophilic methylotrophic Bacillus methanolicus MGA3 using RNA-sequencing provides detailed insights into its previously uncharted transcriptional landscape. *BMC Genomics* **16,** 73 (2015).

118. Müller Jonas E. N. *et al.* Proteomic analysis of the thermophilic methylotroph Bacillus methanolicus MGA3. *PROTEOMICS* **14,** 725–737 (2014).

119. Arfman, N., Van Beeumen, J., De Vries, G. E., Harder, W. & Dijkhuizen, L. Purification and characterization of an activator protein for methanol dehydrogenase from thermotolerant Bacillus spp. *J. Biol. Chem.* **266,** 3955–3960 (1991).

120. Kloosterman, H., Vrijbloed, J. W. & Dijkhuizen, L. Molecular, Biochemical, and Functional Characterization of a Nudix Hydrolase Protein That Stimulates the Activity of a Nicotinoprotein Alcohol Dehydrogenase. *J. Biol. Chem.* **277,** 34785–34792 (2002).

121. Krog, A. *et al.* Methylotrophic Bacillus methanolicus encodes two chromosomal and one plasmid born NAD+ dependent methanol dehydrogenase paralogs with different catalytic and biochemical properties. *PloS one* **8,** e59188 (2013).

122. Kotrbova-Kozak, A., Kotrba, P., Inui, M., Sajdok, J. & Yukawa, H. Transcriptionally regulated adhA gene encodes alcohol dehydrogenase required for ethanol and n-propanol utilization in Corynebacterium glutamicum R. *Appl. Microbiol. Biotechnol.* **76,** 1347–1356 (2007).

123. World Health Organization. Formaldehyde in Drinking-water. (2005).

124. Tseng, H.-C., Martin, C. H., Nielsen, D. R. & Prather, K. L. J. Metabolic engineering of Escherichia coli for enhanced production of (R)- and (S)-3-hydroxybutyrate. *Appl. Environ. Microbiol.* **75,** 3137–3145 (2009).

125. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences* **97,** 6640–6645 (2000).

126. US EPA, O. Global Anthropogenic Non-CO2 Greenhouse Gas Emissions: 1990-2030. *US EPA* (2016). Available at: https://www.epa.gov/global-mitigation-non-co2-greenhouse-gases/global-anthropogenic-non-co2-greenhouse-gas-emissions. (Accessed: 30th March 2018)

127. National Academy of Sciences, National Academy of Engineering, and National Research Council. America's Energy Future: Technology and Transformation. *The National Academies Press.* (2009).

128. Elvidge, C. D. *et al.* A Fifteen Year Record of Global Natural Gas Flaring Derived from Satellite Data. *Energies* **2,** 595–622 (2009).

129. Fei, Q. *et al.* Bioconversion of natural gas to liquid fuel: Opportunities and challenges. *Biotechnology Advances* **32,** 596–614 (2014).

130. Conrado, R. J. & Gonzalez, R. Envisioning the Bioconversion of Methane to Liquid Fuels. *Science* **343,** 621–623 (2014).

131. Haynes, C. A. & Gonzalez, R. Rethinking biological activation of methane and conversion to liquid fuels. *Nat. Chem. Biol.* **10,** 331–339 (2014).

132. Schrader, J. *et al.* Methanol-based industrial biotechnology: current status and future perspectives of methylotrophic bacteria. *Trends in Biotechnology* **27,** 107–115 (2009).

133. Anthony, C. & Zatman, L. J. The microbial oxidation of methanol. The alcohol dehydrogenase of Pseudomonas sp. M27. *Biochem. J.* **96,** 808–812 (1965).

134. Quayle, J. R. & Ferenci, T. Evolutionary aspects of autotrophy. *Microbiol Rev* **42,** 251–273 (1978).

135. Peyraud, R. *et al.* Demonstration of the ethylmalonyl-CoA pathway by using 13C metabolomics. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 4846–4851 (2009).

136. Hu, B. & Lidstrom, M. E. Metabolic engineering of Methylobacterium extorquens AM1 for 1-butanol production. *Biotechnology for Biofuels* **7,** (2014).

137. Price, J. V., Chen, L., Whitaker, W. B., Papoutsakis, E. & Chen, W. Scaffoldless engineered enzyme assembly for enhanced methanol utilization. *Proceedings of the National Academy of Sciences* **113,** 12691–12696 (2016).

138. Woolston, B. M., Roth, T.B., Kohale, I., Liu, D. R. & Stephanopoulos, G. Development of a formaldehyde biosensor with application to synthetic methylotrophy. *Biotechnol. Bioeng.* **115**, 206-215 (2018).

139. Ochsner, A. M., Müller, J. E. N., Mora, C. A. & Vorholt, J. A. In vitro activation of NAD-dependent alcohol dehydrogenases by Nudix hydrolases is more widespread than assumed. *FEBS Letters* **588,** 2993–2999 (2014).

140. Woolston, B., King, J., Reiter, M., Van Hove, B. & Stephanopoulos, G. Improving formaldehyde consumption drives methanol assimilation in engineered E. coli. In Review.

141. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10,** 845 (2015).

142. Moon, J.-H. *et al.* Structures of Iron-Dependent Alcohol Dehydrogenase 2 from Zymomonas mobilis ZM4 with and without NAD+ Cofactor. *Journal of Molecular Biology* **407,** 413–424 (2011).

143. King, J. R., Woolston, B. M. & Stephanopoulos, G. Designing a New Entry Point into Isoprenoid Metabolism by Exploiting Fructose-6-Phosphate Aldolase Side Reactivity of Escherichia coli. *ACS Synth. Biol.* **6,** 1416–1426 (2017).