Title
Exploring in vivo biochemistry with C4 fuel and commodity chemical pathways

Author
Davis, Matthew Aaron

Publication Date
2015

Peer reviewed|Thesis/dissertation

Exploring in vivo biochemistry with C4 fuel and commodity chemical pathways

by

Matthew Aaron Davis

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy in
Molecular and Cell Biology
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:
Professor Michelle C. Y. Chang, Chair
Professor Jamie H. D. Cate
Professor Ming C. Hammond
Professor John E. Dueber

Summer 2015

Exploring in vivo biochemistry with C4 fuel and commodity chemical pathways

by Matthew Aaron Davis

Abstract

Exploring in vivo biochemistry with C4 fuel and commodity chemical pathways

by

Matthew Aaron Davis

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Michelle C. Y. Chang, Chair

The biological diversity found throughout the world contains equally wondrous chemical diversity that can operate with the precision, efficiency, and scale that humanity has yet to attain. This capacity is an untapped resource that must be understood and harnessed to address pressing global needs for food, energy, medicine, and materials. Wielding this power will require a deeper understanding of how a given biological process occurs in the context of a cell. Metabolic pathways are an ideal model system to study biochemical processes *in vivo* as they are integral to the cell's survival, they are regulated on multiple interlocking levels, and they have a broad dynamic range with many measurable inputs and outputs.

We have studied a synthetic metabolic pathway in *E. coli* as a means of gaining insight into biological regulatory networks, but also with the goal of optimizing production of the second-generation biofuel *n*-butanol. Our previous pathway suffered from poor substrate specificity in the final enzyme, leading to off-target products and decreased yield. This enzyme, AdhE2, is a bifunctional aldehyde alcohol dehydrogenase that catalyzes sequential reductions of acyl-CoAs to alcohols through aldehyde intermediates. The enzyme was biochemically characterized to determine its substrate specificity, coordination between active sites, and oligomerization behavior. The enzyme was found to be undesirable for butanol production and new classes of enzymes were explored.

To replace AdhE2 we employed bioinformatic methods to identify a family of monofunctional aldehyde dehydrogenases. This family was screened and a highly specific enzyme was identified. The improved butanol production pathway was then a suitable tool for exploring regulatory mechanisms controlling metabolism by employing whole genome mutagenesis and selection. A butanol production strain was engineered such that its growth under anaerobic conditions was directly linked to butanol production. This strain's genome was mutagenized and subjected to anaerobic growth selection to enrich for mutants producing elevated levels of butanol. We then sequenced the genomes of these strains to identify regulatory mechanisms impacting butanol production.

Finally, we expanded upon our butanol production pathway by leveraging the previously identified aldehyde dehydrogenase family for the production of the commodity chemicals 1,3-butanediol and 4-hydroxy-2-butanone. Aldehyde and alcohol dehydrogenases were identified by a variety of methods and screened for production. We developed several strategies to afford control over the ratio of products produced including pathway design and expression level tuning. Directed evolution methods including DNA shuffling and saturation mutagenesis were also used to further tailor aldehyde dehydrogenases for the desired products.

In sum we have extensively characterized a number of aldehyde and alcohol dehydrogenases from multiple families. Optimized pathways for production of *n*-butanol, 1,3-butanediol, and 4-hydroxy-2-butanone were developed. A genetic selection for metabolite production was developed and validated, and evolved strains were characterized to identify important regulatory mechanisms.

Table of Contents

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Appendices

List of Abbreviations

ADAM	array-based discovery of adaptive mutations
ADH	alcohol dehydrogenase
ALDH	aldehyde dehydrogenase
ATP	adenosine-5´-triphosphate
BDO	1,3-butanediol
CoA	coenzyme A
Cb	carbenicillin
Cm	chloramphenicol
CPM	7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methylcoumarin
dNTP	deoxynucleotide triphosphate
DMSO	dimethyl sulfoxide
DNase	deoxyribonuclease
DTT	diothiothreitol
dTTP	deoxythymidine triphosphate
EDTA	ethylenediaminetetraacetic acid
EMS	ethyl methanesulfonate
ePCR	error-prone polymerase chain reaction
ESI-MS	electron spray ionization mass spectrometry
FPLC	fast protein liquid chromatography
GC-FID	gas chromatography-flame ionization detection
GC-MS	gas chromatography-mass spectrometry
HB	4-hydroxy-2-butanone
IPTG	isopropyl β-D-1-thiogalactopyranoside
KCM	KCl, CaCl2, MgCl2
Km	kanamycin
LB	Luria broth with Miller's modification
LC-MS	liquid chromatography- mass spectrometry
MAGE	multiplex automated genome engineering
MR	molecular replacement
MOPS	3-(N-morpholino)propanesulfonic acid
OD_{600}	optical density at 600 nm
PCR	polymerase chain reaction
PDHc	pyruvate dehydrogenase complex
PMSF	phenylmethanesulfonyl fluoride
REGRES	recursive genomewide recombination and sequencing
RP-HPLC	reversed phase-high performance liquid chromatography
RPM	revolutions per minute
SADH	secondary alcohol dehydrogenase
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SSN	sequence similarity network
Sp	spectinomycin
TEMED	N,N,N′,N′-tetramethylethane-1,2-diamine
TB	terrific broth

Tet	tetracycline
TEV	tobacco etch virus
TCEP	tris(2-carboxyethyl)phosphine
Tris	trisaminomethane
WT	wild type

Acknowledgements

First and foremost, I would like to thank my advisor, Professor Michelle Chang. Her guidance and support over my graduate career has been essential to everything I have achieved. I will always admire and strive to live up to her keen attention to detail in all things; and I will equally admire her acknowledgement that some details are "just whatevers". I am grateful for her scientific mentorship, both in an experimental sense as well as in related skills including presentation and organization. Finally and most importantly, Michelle has been instrumental in providing the opportunity to enjoy a wide range of truly delicious food.

I would next like to thank the members of my thesis committee: Professor Jamie Cate, Professor Ming Hammond, and Professor John Dueber. They have all provided valuable insight scientifically, as well as kind words of encouragement. I am glad to have rotated with and taught for them respectively.

It is unlikely that I would have signed up for six extra years of education had it not been for the many teachers I've learned from along the way. I have not surprisingly had a long affinity for science teachers as I learned as much as I could about the world around me. Laura Schmied and Tom Stevenson gave me a start in grade school and junior high, and Rich Paul, Ann Sulivan, and Jeff Weyers continued my exploration in high school. Outside of science I was also fortunate to be taught by Gail Hopkins, Josh Piper, Cindy Wierzba, Tom Wierzba, Liz Driscoll, Fred Herink, and many others. As an undergraduate student I was very privileged to work with Professor Vernita Gordon and Professor Gerard Wong. They gave me a lab notebook, a set of keys, and then turned me loose on a very expensive microscope. I didn't break anything, and I am very grateful for that first research opportunity.

Some are cursed with terrible colleagues; many get by with average colleagues; a few are graced with excellent colleagues. I am glad to have been in the latter. When I joined the lab an excellent culture was already well established by Brooks Bond-Watts, Maggie Brown, Mark Walker, Amy Weeks, and Laura Fredriksen. All of them were extremely helpful both scientifically and personally throughout my career. I owe a special debt to Brooks for teaching me nearly everything I know, keeping the lab a well-oiled machine, and serving as a perfect officemate. I also owe a special debt to Maggie who has always served as a role model by displaying great intensity tempered with great empathy. I was very glad to have Ben Thuronyi and Stephanie Jones as companions joining the lab at the same time as me; they always provided the right advice at the right time. I will never be as efficient as Ben, and I will never be as feisty as Stephanie, but they are goals to work towards. I was fortunate to work with a number of talented postdocs over the years; Dr. Miao Wen, Dr. Jeff Hanson, Dr. Michiei Sho, Dr. Zhen Wang, Dr. Ningkun Wang, and Dr. Quanjiang Ji always provided seasoned advice when needed. As new students joined the lab I was happy to work with Mike Blaise who has been the source of many amusing conversations, Joe Gallagher who can be counted on to have knowledge of all available free food sources on campus, and Jon McMurry who, somehow, always has a new (insert arcane area of knowledge here) fact of the day. Rounding out the crew is Omer Ad who provides numerous new recipe ideas, Jase Gehring who can always be counted on for a

laugh, Vivian Yu who is relentlessly, soul-crushingly, optimistic and eager to help a friend in need, and Jorge Marchand who has an impressive supply of shaker-based puns. I am also glad to know our newest students, Monica Neugebauer and Sasilada Sirirungraung, who are already off to a good start in the lab. Finally I am glad to have mentored Thomas Le as an undergraduate and rotation students Andrew Dippel and Julia Lazzari-Dean. Mentoring students can be a humbling yet rewarding process, and all three were excellent to work with.

Science marches on at a relentless pace. Sometimes your project is scooped out from under you, and sometimes Professor Daniel Gibson invents Gibson Assembly. The clouds part, angels sing, and your cloning is magical. Then you get scooped next week.

My friends have never plated my transformations for me (a true sign of friendship), but they have always provided entertainment, distractions, and memories. Whether new grad school friends that are reassuringly awkward and nerdy, or old friends that have stuck around for so many years, they have all improved my life on a daily basis.

I would not be where I am today without the support of my family, and particularly my parents. As long as I can remember they have done everything possible to provide my brothers and I with a happy, healthy, and nurturing upbringing. I will probably never grasp the amount of time and energy they devoted to allowing me to participate in numerous enriching activities growing up. I am thankful they have shaped me into the person I am.

Finally I want to thank my wonderful girlfriend Charlene. She has supported me tirelessly through some of the best times and through some of the worst times. She is the reason I tackle each new day to the best of my ability, and she helps me to become a better person every day. She makes me laugh, she makes me happy, and I would be lost without her.

Chapter 1: *Introduction*

1.1 Introduction

As our human mark on our planet grows more indelible, our need to leverage the resources at our disposal grows more pressing. One resource that is so far underutilized is the chemical and biological diversity permeating every square inch of the biosphere[1-3]. This diversity routinely employs chemistry that we cannot harness in a flask at scales that we struggle to comprehend. We have begun to harness the power of biology to address our needs for food, fuel, medicine, and chemicals, but to fully leverage this resource a greater understanding of the biological processes will be required[4,5]. Understanding a biological process in a tube is incredibly informative to be sure, but understanding that process in the context of thousands of overlapping reactions confined in 1 μm^3 will help us more skillfully build biology.

1.2 Synthetic pathways report on biochemical networks *in vivo*

Living systems are capable of complex yet precise chemistry at unparalleled efficiency and scale. As human demands push natural resources to their limits the need to tap into this chemical resource continues to grow. Engineered biology will increasingly supply our food, chemicals, and materials, but greater understanding of how living systems achieve their chemical goals is needed to re-engineer these systems more efficiently. Synthetic phenotypes can become a readout for all cellular decision-making, and provide a platform to understand biology at scale so that we may harness biology at scale (*Figure 1.1*).

Development of microbially produced fuels and chemicals will require a deeper understanding of biological decision-making. Because these pathways must perform at high yield and flux, they will draw from and interact deeply with the native central metabolism of the cell. Achieving such tight integration with the native machinery gives us excellent insight into the regulatory networks that govern such integral cellular mechanisms. Synthetic metabolic pathways have a very large dynamic range that can be used to peer deeper into in vivo biochemistry than is otherwise possible, thus making them attractive model systems (*Figure 1.2*).

As an example, research on the pyruvate dehydrogenase complex (PDHc) stretches back to the 1960s if not earlier[6], and despite its key role in metabolism, new regulatory mechanisms controlling the PDHc continue to be discovered[7]. This 60-subunit masterpiece occupies a central position within metabolism as it is one of the gatekeepers controlling the fate of pyruvate[8]. Pyruvate can enter multiple different fermentation pathways, or it can enter several anapleurotic pathways, or it can become acetyl-CoA to be flung to the far corners of the cell. Currently, the PDHc is known to be regulated at the transcriptional, post-translational, and allosteric levels, with multiple opposing inputs at every level[9-13].

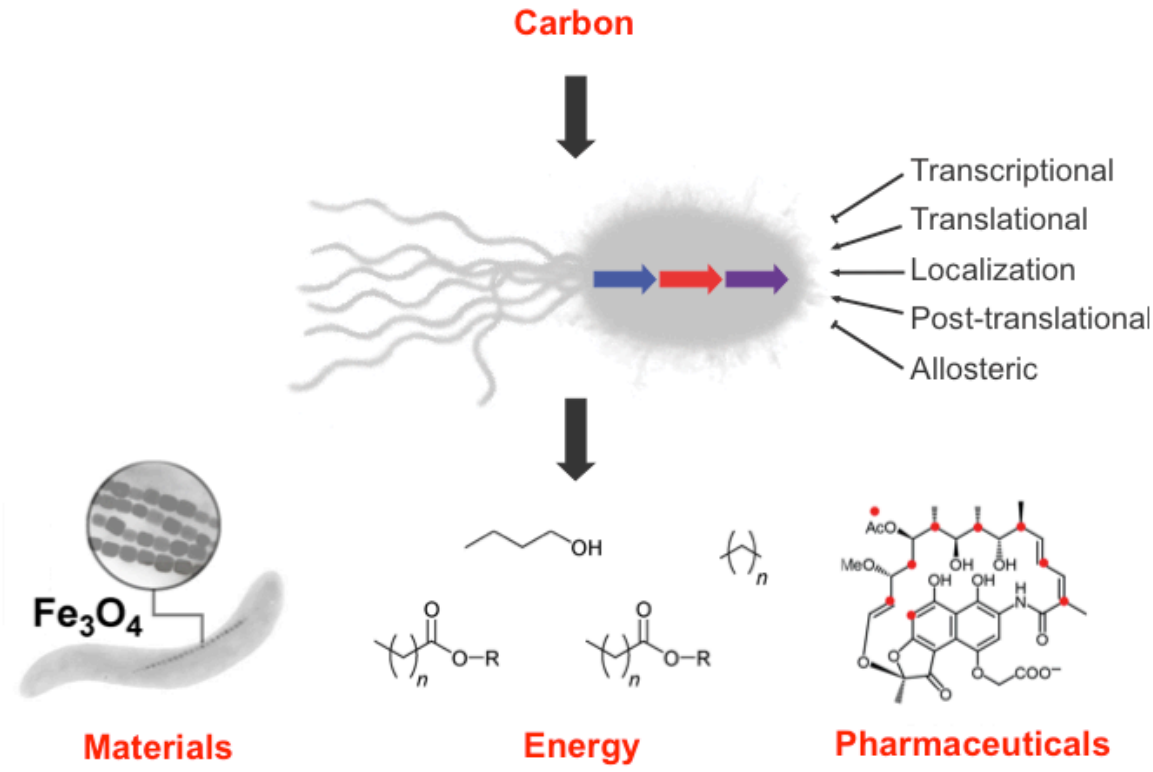


Figure 1.1 Synthetic phenotypes can report on in vivo biochemistry
Living systems are capable of complex yet precise chemistry at unparalleled efficiency and scale. As human demands push natural resources to their limits the need to tap into this chemical resource continues to grow. Engineered biology will increasingly supply our food, chemicals, and materials, but greater understanding of how living systems achieve their chemical goals is needed to re-engineer these systems more efficiently. Synthetic phenotypes become a readout for all cellular decision-making, and provide a platform to understand biology at scale so that we may harness biology at scale.

With so many inputs modulating PDHc activity, the complex is positioned to affect, and be affected by, a majority of cellular processes. If we then connect a reporter pathway such as our acetyl-CoA derived butanol production pathway to PDHc output we have an easy way to measure PDHc activity under a variety of growth conditions. This effect can be amplified when the host metabolism is made increasingly dependent on the reporter pathway. This can be done by growing the cells in minimal media, growing under anaerobic conditions, or by removing native fermentation pathways that the cell would typically use to balance redox requirements. The growth of this fermentation knockout strain is now directly tied to butanol production through the redox requirement needed to balance glycolytic flux and ATP production.

To further extract information about the cellular decision making process we can employ mutagenesis followed by growth selection to enrich for cells carrying mutations that positively impacted PDHc activity and thus butanol production[14]. In essence we can generate millions or billions of hypothesis about the importance of a given gene to cellular carbon fate decisions, and let the cells inform us through improved growth whether or not that gene is important[15,16]. This mutational investigation can be widened to

cover the entire genome[17], or narrowed to the binding interface of a single transcription factor[18]; all mutations that impact cellular metabolism will have a growth cost or benefit and will disappear or become enriched in the selection.

That is not to say that cells must simply be mutated and screened to determine the most important mechanisms governing their decision-making processes. Many mutagenesis methods will generate multiple mutations per cell[19], but only a small fraction of these are likely to have a large contribution to the phenotype in question. If the goal is simply to improve production of a metabolite there is little need for mechanistic understanding[20]. Mutagenesis and selection are simply applied in an iterative fashion until the production goal is met. However if the goal is to understand the means by which cells reroute their metabolism robust tools are necessary to dissect complex genotypes and map them to complex phenotypes.

As a first pass for reconstruction of mutagenized strains, robust methods have been developed for generating arbitrary mutations in the genome of *E. coli*[21]. However, depending on mutagenesis method, any given cell may contain tens or even hundreds of mutations[22]. To meet this challenge of sorting through hitchhiker mutations in search of causative alleles a number of strategies could be employed. Combinatorial reconstruction of mutant genotypes can be achieved by multiplex automated genome engineering (MAGE)[23]. MAGE allows for a pool of mutations to be constructed iteratively and at random, and intermediate strains may be screened throughout construction for comparison to the originally mutagenized strain. MAGE has the limitation that for a large pool of mutations the complete combinatorial space cannot be explored[24], and that new genotypes are built additively from the wild type genotype instead of subtractively from the mutagenized genotype, which could be inefficient if many mutations contribute to the phenotype.

Another approach that could be used is array-based discovery of adaptive mutations (ADAM)[25]. ADAM is a means of discovering and quantifying the contribution of mutations across the entire genome simultaneously. In comparison to the additive process of MAGE, ADAM is a subtractive process; a library of strains is created from the mutagenized strain such that each library member has been reverted to the wild type sequence at a single location. This library of single site revertants is then subjected to the same selection that initially generated the mutagenized strain. During this selection any revertant that has lost a causative allele will no longer have a growth advantage relative to the rest of the library, and that strain will be depleted from the population. Conversely a revertant that has lost a hitchhiker mutation will retain the evolved phenotype and persist in the population. A microarray or high-throughput sequencing is then used to measure the abundance of revertants with and without selection and thus quantify the contribution of every mutation genome wide. ADAM has been used in a variety of scenarios to effectively deconvolute complex genotypes giving rise to a phenotype[15,26,27].

A final method useful in dissecting complex genotype:phenotype linkages is recursive genomewide recombination and sequencing (REGRES)[28]. Similar to ADAM, REGRES is a subtractive method that systematically reverts a mutagenized strain to wild type while screening intermediate genotypes in comparison to the evolved phenotype. Here chimeric genomes are produced by conjugating the mutagenized strain with a library of wild type strains. Transconjugants retaining causative alleles will be retained in the library while

transconjugants that were reverted to wild type at these alleles will be depleted. The procedure is then iterated to arrive at the minimal genotype responsible for the phenotype in question, and the resulting strain is sequenced . This process was used to confirm the importance of two mutations out a background of over 70 mutations that arose throughout the long term evolution experiment in *E. coli*[29,30].

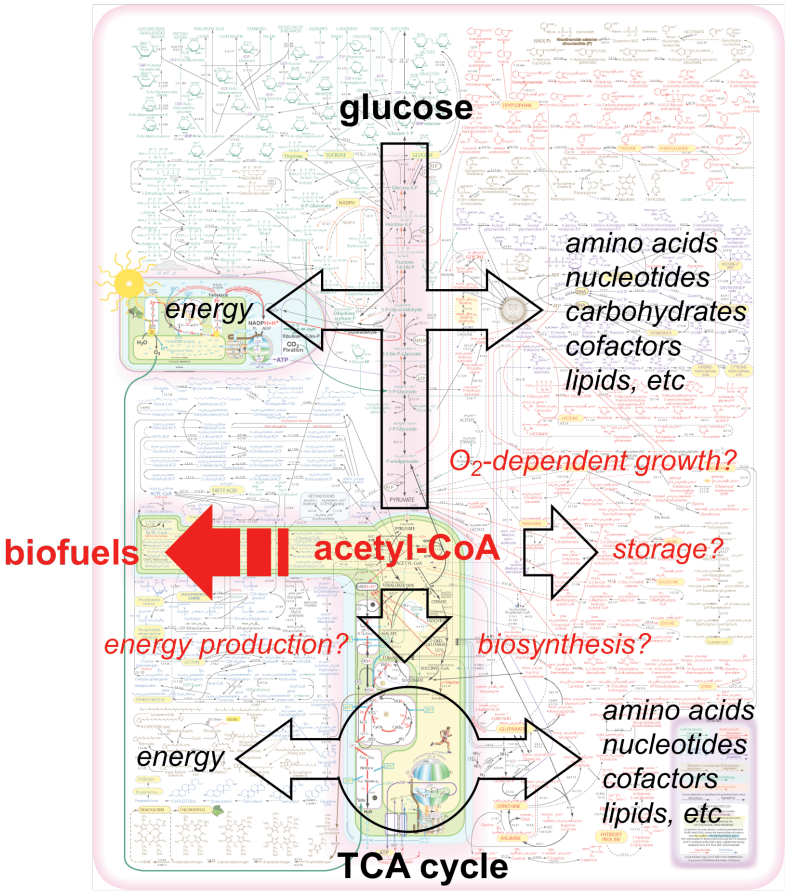


Figure 1.2 Metabolism as a model system to study in vivo biochemistry
Development of microbiallly produced fuels and chemicals will require a deeper understanding of biological decision making. Because these pathways must perform at high yield and flux, they will draw from and interact deeply with the native central metabolism of the cell. Achieving such tight integration with the native machinery gives us excellent insight into the regulatory networks that govern such integral cellular mechanisms. Synthetic metabolic pathways have a very large dynamic range that can be used to peer deeper into in vivo biochemistry than is otherwise possible.

1.3 The challenge of high-throughput screening for metabolites

If we are to aim to understand living systems more deeply in the context of their dense interlocking regulatory networks, we need tools that can probe these networks deeply and rapidly. High-throughput screens and selections are well suited to this task. Although the methodologies falling under this umbrella term are extremely diverse, they all provide a means to explore diverse phenotypes very quickly, usually in an unbiased way (when well implemented), and with a mind to iteration. Genetic selections have a rich and storied history as a means of discovering new biological processes and elucidating complex mechanisms[31,32]. Likewise high-throughput screens for directed evolution have delivered immense improvements to industrial biotechnology[17]. In the context of applying these methods to metabolic systems with the goal of uncovering core regulatory principles the power of iteration and dynamic range inherent to metabolism stand out as an especially good pairing. Iteration is almost always required to meet a directed evolution goal, and in the realm of metabolite production there is almost always higher ground to be sought out (budget providing of course). The three core challenges in applying high-throughput techniques to metabolite production are diversity generation, transformation efficiency, and detection of the desired metabolite.

Diversity generation is straightforward in that there are many ways to make many mutations[33]. Diversity generation is less straightforward in that many mutations do nothing, or worse, are harmful. The method of diversity generation should be tailored to the directed evolution goal. This will often dictate where mutations are desired, what kind of mutations are desired, and how many mutations are desired. For protein engineering, mutations that alter substrate specificity are more commonly found close to the active site, while mutations conferring greater thermal tolerance are likely to be distributed through the tertiary structure[34]. With some methods such as saturation mutagenesis, both the location and type of mutation can be controlled in advance; careful selection of the degenerate codons used can limit mutations likely to be harmful, such as early stop codons, while biasing toward a subset of desired codons[35]. The number of mutations desired can often be controlled, but in many scenarios external factors such as the transformation efficiency or screening capacity may artificially limit this.

Transformation efficiency poses the next hurdle, and again is often defined by the overall goal or other factors like the screening capacity. Organisms with well-developed transformation protocols are unlikely to limit pursuit of the goal, but if the screen must be carried out in an organism with limited transformation efficiency there may be very little recourse. Generally speaking *E. coli* libraries of 1×10^9 are common (although larger are feasible[36]), and libraries of 1×10^8 are available in *S. cerevisiae*. However transformation efficiency can be circumvented if the desired diversity can be achieved *in vivo*, for example through mutator strains[37,38]. Similar in concept to these mutator strains, new techniques such as phage-assisted continuous evolution[39] and compartmentalized partnered replication [40] combine most or all of the diversity generation and transformation steps into a single *in vivo* process, thus negating the loss of diversity that usually occurs when transferring among *in vitro* and *in vivo* processes.

Finally, the most problematic challenge of metabolite detection must be dealt with. The simple truth is that many metabolites of interest are largely invisible to high

throughput methods (*Figure 1.3*). Some metabolites are chromophores or fluorophores, some metabolites are natively essential for cell survival, but everything else will require clever screen or selection design. When faced with this hurdle surrogate metabolites that are more readily detectable may be considered[41,42], but the adage "you get what you screen for" does not exist apropos of nothing. A number of solutions to this problem do exist but may have some drawbacks for any given scenario. Fluorescent dyes that bind a molecule of interest may exist[43], a reporter strain carrying a marker and sensitive to the metabolite could be developed[44], or a transcription factor responsive to the compound may be used to drive expression of a reporter[45]. Work is also progressing on computational protein design methods that could allow a custom transcription factor to be developed[46]. Should none of these options prove suitable new mass-spectrometry instrumentation is becoming available that may deliver enough screening capacity to interrogate a well-designed library. Instruments capable of analyzing greater than 8,000 samples per day are available; running continuously for two weeks a moderately sized library of 1 x 10^5 could be attempted[47].

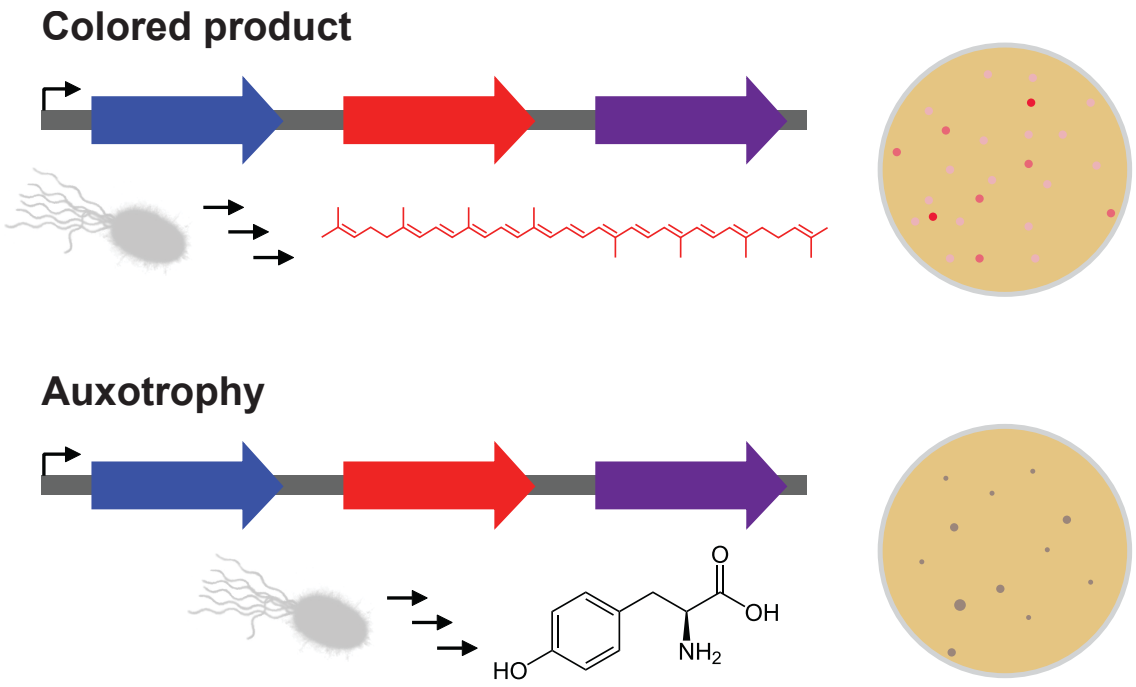


Figure 1.3 Conventional screens for conspicuous metabolites
High throughput screening methods for metabolite production are severely limited to the most conspicuous of metabolites. Pathways for lycopene production can be screened by searching for dark red cells, and pathways for tyrosine production can be screened by growth of an auxotrophic strain sensitive to tyrosine levels. Vastly more metabolites do not have easily identifiable screens or selections to enable high throughput methodology.

1.4 References

1. J. D. Keasling, Synthetic biology for synthetic chemistry, *ACS Chem. Biol.* **2008**, *3*, 64–76.
2. V. G. Yadav and G. Stephanopoulos, Reevaluating synthesis by biology, *Curr. Opin. Microbiol.* **2010**, *13*, 371–376.
3. A. M. Weeks and M. C. Y. Chang, Constructing de novo biosynthetic pathways for chemical synthesis inside living cells, *Biochemistry* **2011**, *50*, 5404–5418.
4. A. Glazer and H. Nikaido, Microbial biotechnology: Fundamentals of applied microbiology, **1995**
5. Z. An, *Handbook of Industrial Mycology*, *Mycology Series* (**2005**). *32*,
6. U. Henning, G. Dennert, R. Hertel and W. S. Shipp, Translation of the structural genes of the E. coli pyruvate dehydrogenase complex, *Cold Spring Harbor Symposia on Quantitative Biology* **1966**, *31*, 227–234.
7. Q. Wang, Y. Zhang, C. Yang, H. Xiong, Y. Lin, J. Yao, H. Li, L. Xie, W. Zhao, Y. Yao, *et al.*, Acetylation of Metabolic Enzymes Coordinates Carbon Source Utilization and Metabolic Flux, *Science* **2010**, *327*, 1004–1007.
8. D. P. Clark, The fermentation pathways of Escherichia coli, *FEMS Microbiol. Rev.* **1989**, *5*, 223–234.
9. J. L. Snoep, M. R. de Graef, A. H. Westphal, A. de Kok, M. J. Teixeira de Mattos and O. M. Neijssel, Differences in sensitivity to NADH of purified pyruvate dehydrogenase complexes of Enterococcus faecalis, Lactococcus lactis, Azotobacter vinelandii and Escherichia coli: implications for their activity in vivo, *FEMS Microbiol. Lett.* **1993**, *114*, 279–283.
10. J. Guest, S. Angier and G. Russell, Structure, expression, and protein engineering of the pyruvate dehydrogenase complex of Escherichia coli, *Ann. N. Y. Acad. Sci.* **1989**, *573*, 76–99.
11. M. de Graef, S. Alexeeva, J. Snoep and M. Teixeira de Mattos, The steady-state internal redox state (NADH/NAD) reflects the external redox state and is correlated with catabolic adaptation in Escherichia coli, *J. Bacteriol.* **1999**, *181*, 2351–2357.
12. M. A. Quail, D. J. Haydon and J. R. Guest, The pdhR-aceEF-lpd operon of Escherichia coli expresses the pyruvate dehydrogenase complex, *Mol. Microbiol.* **1994**, *12*, 95–104.
13. M. E. Spencer and J. R. Guest, Transcription analysis of the sucAB, aceEF and lpd genes of Escherichia coli, *Mol. Gen. Genet.* **1985**, *200*, 145–154.
14. Y. Kim, L. O. Ingram and K. T. Shanmugam, Dihydrolipoamide Dehydrogenase Mutation Alters the NADH Sensitivity of Pyruvate Dehydrogenase Complex of Escherichia coli K-12, *J. Bacteriol.* **2008**, *190*, 3851–3858.
15. H. S. Girgis, Y. Liu, W. S. Ryu and S. Tavazoie, A Comprehensive Genetic Characterization of Bacterial Motility, *PLoS Genet.* **2007**, *3*, e154.

16. H. S. Girgis, K. Harris and S. Tavazoie, Large mutational target size for rapid emergence of bacterial persistence, *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 12740–12745.

17. R. Rowlands, Industrial strain improvement: mutagenesis and random screening procedures, *Enzyme Microb. Technol.* **1984**,

18. H. Alper and G. Stephanopoulos, Global transcription machinery engineering: A new approach for improving cellular phenotype, *Metab. Eng.* **2007**, *9*, 258–267.

19. S. Parekh, V. A. Vinci and R. J. Strobel, Improvement of microbial strains and fermentation processes, *Appl. Microbiol. Biotechnol.* **2000**, *54*, 287–301.

20. I. Normansell, A. Hénaut and A. Danchin, Strain improvement in antibiotic-producing microorganisms, *J. Chem. Tech. Biotechnol.* **1982**,

21. K. A. Datsenko and B. L. Wanner, One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6640–6645.

22. D. Parkhomchuk, V. Amstislavskiy, A. Soldatov and V. Ogryzko, Use of high throughput sequencing to observe genome dynamics at a single cell level, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 20830–20835.

23. H. Wang, F. Isaacs, P. Carr, Z. Sun, G. Xu, C. Forest and G. Church, Programming cells by multiplex genome engineering and accelerated evolution, *Nature* **2009**, *460*, 894–898.

24. N. R. Sandoval, J. Y. H. Kim, T. Y. Glebes, P. J. Reeder, H. R. Aucoin, J. R. Warner and R. T. Gill, Strategy for directing combinatorial genome engineering in Escherichia coli, *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 10540–10545.

25. H. Goodarzi, A. K. Hottes and S. Tavazoie, Global discovery of adaptive mutations, *Nat. Methods* **2009**, *6*, 581–583.

26. S. Amini, A. K. Hottes, L. E. Smith and S. Tavazoie, Fitness landscape of antibiotic tolerance in Pseudomonas aeruginosa biofilms, *PLoS Pathog.* **2011**, *7*, e1002298.

27. H. Goodarzi, B. D. Bennett, S. Amini, M. L. Reaves, A. K. Hottes, J. D. Rabinowitz and S. Tavazoie, Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in E. coli, *Mol. Syst. Biol.* **2010**, *6*, 378.

28. E. M. Quandt, D. E. Deatherage, A. D. Ellington, G. Georgiou and J. E. Barrick, Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in Escherichia coli, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 2217–2222.

29. J. E. Barrick, D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski and J. F. Kim, Genome evolution and adaptation in a long-term experiment with Escherichia coli, *Nature* **2009**, *461*, 1243–1247.

30. Z. D. Blount, J. E. Barrick, C. J. Davidson and R. E. Lenski, Genomic analysis of a key innovation in an experimental Escherichia coli population, *Nature* **2012**, *489*, 513–518.

31. J. A. Barnett, A history of research on yeasts 10: foundations of yeast genetics, *Yeast* **2007**, *24*, 799–845.

32. T. D. Brock, The emergence of bacterial genetics, **1990**

33. J. A. Dietrich, A. E. McKee and J. D. Keasling, High-Throughput Metabolic Engineering: Advances in Small-Molecule Screening and Selection, *Annu. Rev. Biochem.* **2010**, *79*, 563–590.

34. K. L. Morley and R. J. Kazlauskas, Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* **2005**, *23*, 231–237.

35. M. T. Reetz and J. D. Carballeira, Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes, *Nat. Protoc.* **2007**, *2*, 891–903.

36. S. S. Sidhu, H. B. Lowman, B. C. Cunningham and J. A. Wells, Phage display for selection of novel binding peptides, *Methods Enzymol.* **2000**, *328*, 333–363.

37. A. Greener, M. Callahan and B. Jerpseth, An efficient random mutagenesis technique using an E. coli mutator strain, *Mol. Biotechnol.* **1997**, *7*, 189–195.

38. M. Camps, J. Naukkarinen, B. P. Johnson and L. A. Loeb, Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9727–9732.

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

40. J. W. Ellefson, A. J. Meyer, R. A. Hughes, J. R. Cannon, J. S. Brodbelt and A. D. Ellington, Directed evolution of genetic parts and circuits by compartmentalized partnered replication, *Nat. Biotechnol.* **2014**, *32*, 97–101.

41. C. L. Hendricks, J. R. Ross, E. Pichersky, J. P. Noel and Z. S. Zhou, An enzyme-coupled colorimetric assay for S-adenosylmethionine-dependent methyltransferases, *Anal. Biochem.* **2004**, *326*, 100–105.

42. K. Wagschal, D. Franqui-Espiet, C. C. Lee, G. H. Robertson and D. W. S. Wong, Enzyme-coupled assay for beta-xylosidase hydrolysis of natural substrates, *Appl. Env. Microbiol.* **2005**, *71*, 5318–5323.

43. I. R. Sitepu, L. Ignatia, A. K. Franz, D. M. Wong, S. A. Faulina, M. Tsui, A. Kanti and K. Boundy-Mills, An improved high-throughput Nile red fluorescence assay for estimating intracellular lipids in a variety of yeast species, *J. Microbiol. Methods* **2012**, *91*, 321–328.

44. B. F. Pfleger, D. J. Pitera, J. D. Newman, V. J. J. Martin and J. D. Keasling, Microbial sensors for small molecules: Development of a mevalonate biosensor, *Metab. Eng.* **2007**, *9*, 30–38.

45. J. A. Dietrich, D. L. Shis, A. Alikhani and J. D. Keasling, Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis, *ACS Synth. Biol.* **2013**, *2*, 47–58.

46. S. Raman, J. K. Rogers, N. D. Taylor and G. M. Church, Evolution-guided optimization of biosynthetic pathways, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 17803–17808.

47. P. T. Rye and W. A. LaMarr, Measurement of glycolysis reactants by high-throughput solid phase extraction with tandem mass spectrometry: Characterization of pyrophosphate-dependent phosphofructokinase as a case study, *Anal. Biochem.* **2015**, *482*, 40–47.

Chapter 2: *Biochemical characterization of the bifunctional aldehyde-alcohol dehydrogenase from Clostridium acetobutylicum*

2.1 Introduction

AdhE2 is a 94 kDa bifunctional enzyme composed of two distinct aldehyde and alcohol dehydrogenase domains linked as a fusion protein[1] (*Figure 2.1*). The first domain catalyzes reduction of short-chain acyl-CoAs to aldehydes with NAD(P)H as a cofactor. The second domain catalyzes a subsequent reduction of aldehydes to alcohols, again with NAD(P)H as a cofactor. This unique domain organization may serve to coordinate the activities of the two domains and potentially enable substrate channeling of volatile and reactive aldehydes.

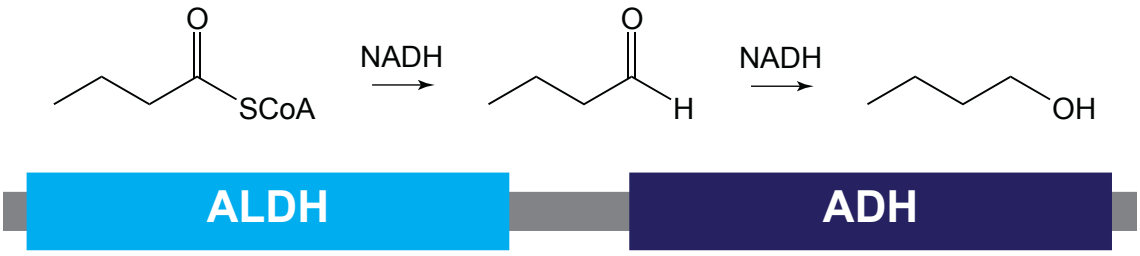


Figure 2.1 Domain architecture of AdhE2
AdhE2 is composed of two distinct aldehyde and alcohol dehydrogenase domains linked as a fusion protein. The first domain catalyzes reduction of short-chain acyl-CoAs to aldehydes with NAD(P)H as a cofactor. The second domain catalyzes a subsequent reduction of aldehydes to alcohols, again with NAD(P)H as a cofactor. This unique domain organization may serve to coordinate the activities of the two domains and potentially enable substrate channeling of volatile and reactive aldehydes.

AdhE2 was identified in *Clostridium acetobutylicum*, the most widely studied member of a family of gram-positive spore-forming strict anaerobes that carry out a unique biphasic fermentation that results in the production of acetone, butanol, and ethanol[2,3]. This two stage fermentation initially secretes butyrate and acetate, perhaps as a means of rapidly consuming available carbon sources and using them to drop the pH of the surrounding environment, thus inhibiting growth of potential competitors and sequestering carbon for future use. These organic acids are then taken back up by the cell and reduced to butanol and ethanol, along with acetone production. AdhE2 homologs are known in a number of other species including some eukaryotic parasites[4,5], but none are known to produce butanol outside of *Clostridia.*

AdhE2 is thought to be a central enzyme in this fermentation pathway as its bifunctional nature means it can complete two subsequent reductions of an acyl-CoA to the final alcohol through an aldehyde intermediate. The bifunctional architecture of this enzyme is somewhat unique: of the 210,000 aldehyde dehydrogenase (ALDH) domains in the Pfam[6] database 77% are monofunctional ALDHs and only 8% are found in a bifunctional protein coupled to an alcohol dehydrogenase (ADH) domain. However the unique possibilities of a bifunctional enzyme may be especially well suited to a critical fermentation enzyme like AdhE2.

As fermentation pathways are high flux by necessity and critical to maintaining redox balance to extract maximum energy from limited resources, it is essential that they are streamlined and have minimal off-pathway loss. This may be especially important where one of the pathway intermediates is a volatile aldehyde that could be lost from the cell. Furthermore, these aldehyde intermediates are reactive and could prove toxic to the cell if they leak from a high flux pathway at an appreciable rate[7,8]. Scenarios such as this are benefited greatly by direct or shielded transfer of the product of one active site to the active site of the subsequent enzyme. Direct transfer is best exemplified by the exquisite pyruvate dehydrogenase complex[9], and shielded transfer is well understood in tryptophan synthase[10].

As a potential second layer of substrate channeling, homologs of AdhE2 are know to assemble into helical ultrastructures called "spirosomes"[5,11,12] (*Figure 2.2*, reproduced from Reference [13]). These ultrastructures contain between 20-60 monomers and are easily visualized by electron microscopy, but little else is understood about their function. At the time of their discovery it was not well recognized that many metabolic enzymes are found in various ultrastructures[14-16], but this may be yet another example of leveraging local concentration effects.

We have successfully employed AdhE2 as part of a synthetic butanol production pathway in *E. coli*, but limitations in its substrate specificity, ease of expression, and overall activity warranted further study. AdhE2 is an intriguing enzyme in its own right, but greater understanding could also aid us in improving its applicability to the construction of high flux fuel and commodity chemical pathways.

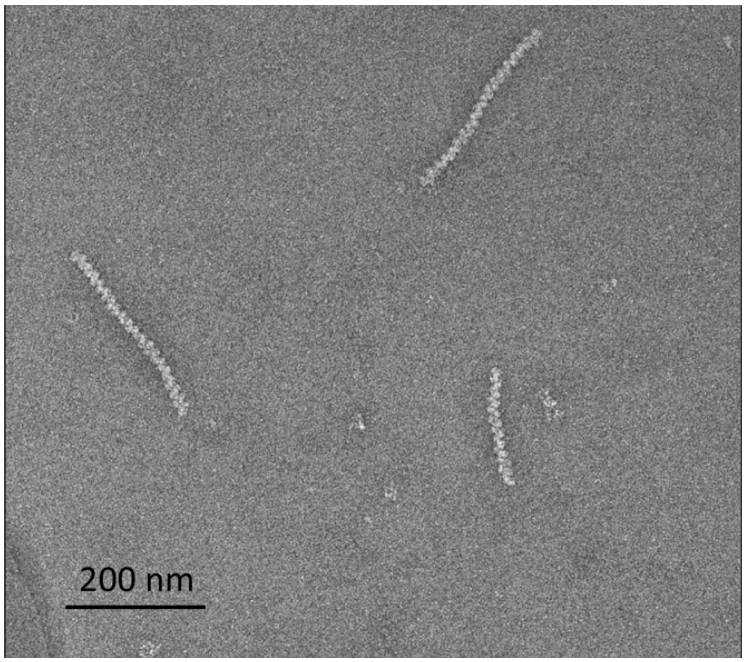


Figure 2.2 Spirosome ultrastructure of the E. coli AdhE2 homolog
Reproduced from Reference [13]

The Streptococcus pneumoniae homolog of AdhE2 has been shown to form large helical "spirosome" ultrastructures by electron microscopy. Monomers are arranged in a head-to-tail fashion of between 20-60 subunits per helical structure resulting in a complex of 5 mDa or larger. This oligomeric assembly could serve to further enhance substrate channeling of volatile and reactive aldehyde substrates, and may provide a mechanism for inter-subunit domain coordination.

2.2 Materials and methods

Commercial materials. Luria-Bertani (LB) Broth Miller, LB Agar Miller, and Terrific Broth (TB) were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), isopropyl-β-D-thiogalactopyranoside (IPTG), phenylmethanesulfonyl fluoride (PMSF), tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), sodium chloride, dithiothreitol (DTT), kanamycin (Km), ethyl acetate and ethylene diamine tetraacetic acid disodium dihydrate (EDTA), were purchased from Fisher Scientific (Pittsburgh, PA). Coenzyme A trilithium salt (CoA), acetyl-CoA, nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), β-mercaptoethanol, sodium phosphate dibasic hepthydrate, and N,N,N',N'-tetramethyl-ethane-1,2-diamine (TEMED) were purchased from Sigma-Aldrich (St. Louis, MO). Acrylamide/Bis-acrylamide (30%, 37.5:1), electrophoresis grade sodium dodecyl sulfate (SDS), Bio-Rad protein assay dye reagent concentrate and ammonium persulfate were purchased from Bio-Rad Laboratories (Hercules, CA). Restriction enzymes, T4 DNA ligase, Phusion DNA polymerase, T5 exonuclease, and Taq DNA ligase were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs) and Platinum Taq High-Fidelity polymerase (Pt Taq HF) were purchased from Invitrogen (Carlsbad, CA). PageRuler™

Plus prestained protein ladder was purchased from Fermentas (Glen Burnie, Maryland). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a stock concentration of 100 μM in 10 mM Tris-HCl, pH 8.5, and stored at either 4°C for immediate use or -20°C for longer term use. DNA purification kits and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). Amicon Ultra 10,000 centrifugal concentrators were purchased from EMD Millipore (Billerica, MA).

Bacterial strains. *E. coli* DH10B-T1^R and BL21(de3)T1^R were used for DNA construction and heterologous protein production, respectively. *E. coli* DH1 and DH1 ΔackA-pta ΔadhE ΔldhA ΔpoxB ΔfrdBC (MC1.24) were used for metabolite production.

Gene and plasmid construction. Restriction enzyme cloning, Gibson assembly, and Golden Gate assembly were used to carry out plasmid construction. All PCR amplifications were carried out with Phusion or Platinum Taq High Fidelity DNA polymerases. All constructs were verified by sequencing (Quintara Biosciences; Berkeley, CA).

Expression of Strep-tagged proteins. TB (1 L) containing carbenicillin (50 μg/mL) in a 2.8 L Fernbach baffled shake flask was inoculated to OD_{600} = 0.05 with an overnight TB culture of freshly transformed *E. coli* containing the appropriate overexpression plasmid. The cultures were grown at 37°C at 200 rpm to OD_{600} = 0.6 to 0.8 at which point cultures were cooled on ice for 20 min, followed by induction of protein expression with 1 mM IPTG and overnight growth at 16°C. Cell pellets were harvested by centrifugation at 9,800 × *g* for 7 min and resuspended at 20 mL/L of culture with Buffer W (100 mM Tris-HCl, 150 mM sodium chloride, 1 mM EDTA, pH 8.0) supplemented with 2 mg/mL lysozyme and 2 uL/50 mL final volume Benzonase and frozen at -80°C.

Purification of Strep-tagged proteins. Frozen cell suspensions were thawed and frozen twice before finally thawing and adding 0.5 mM PMSF as a 50 mM stock solution in ethanol dropwise. The cell suspension was lysed at with a Misonix 3000 probe sonicator at full power with a 15 second on, 60 second off cycle for a total sonication time of 2.5 minutes. The lysate was centrifuged at 15,300 × *g* for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction by addition of 0.5% polyethylenimine as a 15% v/v stock solution added dropwise. The precipitated DNA was removed by centrifugation at 15,300 × *g* for 20 min at 4°C. The lysate was loaded onto a Strep-tactin Superflow High Capacity column (IBA, 1 mL resin/L expression culture) by gravity flow. The column was washed with 20 column volumes Buffer W. The protein was then eluted with 2.5 mM desthiobiotin in Buffer W. Fractions containing ALDH protein by A_{280} were pooled and concentrated in an Amicon Ultra 10,000 MWCO concentrator. Concentrated protein was supplemented with glycerol to 10% v/v and stored at -80°C.

Size exclusion chromatography. Purified protein was loaded on a Superose 6 or Superdex 200 size exclusion column connected to an Akta Purifier FPLC (GE Healthcare). Separation was carried out at 0.5 mL/minute for analytical Superose 6 runs or 1 mL/min for preparatory Superdex 200 runs in 100 mM Tris-HCl, 150 mM sodium chloride, 1 mM EDTA, pH 8.0 and fractions were collected for preparatory purifications.

Fraction activity was assayed immediately after separation for oligomer activity determination.

Enzyme assays. Activity of ALDH proteins was measured by monitoring the oxidation of NADH at 340 nm at 25 °C. The assay mixture (400 μL) contained 100 μM NADH in 100 mM Tris 1 mM DTT pH 7.5. The reaction was initiated by the addition of substrate. Kinetic parameters (k_{cat}, K_M) were determined by fitting the data using Microcal Origin to the equation: $v_o = v_{max}[S] / (K_M + [S])$, where v is the initial rate and [S] is the substrate concentration. Data are reported as mean ± s.e. (n = 3) unless otherwise noted with standard error derived from the nonlinear curve fitting. Error bars on graphs represent mean ± s.d. (n = 3). Error in k_{cat}/K_M is calculated by propagation of error from the individual kinetic parameters.

Cell culture. *E. coli* strains were transformed by electroporation using the appropriate plasmids. A single colony from a fresh transformation was then used to seed an overnight culture grown in Terrific Broth (TB) (EMD Biosciences) supplemented with 1.5% (w/v) glucose and appropriate antibiotics at 37 °C in a rotary shaker (200 rpm). Antibiotics were used at a concentration of 50 μg ml^{-1} for strains with a single resistance marker. For strains with multiple resistance markers, kanamycin and chloramphenicol were used at 25 μg ml^{-1} and carbenicillin was used at 50 μg ml^{-1}.

***In vivo* production of alcohols.** Overnight cultures of freshly transformed *E. coli* strains were grown for 12–16 h in TB at 37 °C and used to inoculate TB (50 ml) with glucose replacing the standard glycerol supplement (1.5% (w/v) glucose for aerobic cultures and 2.5% (w/v) glucose for anaerobic cultures) and appropriate antibiotics to an optical density at 600 nm (OD_{600}) of 0.05 in a 250 mL-baffled flask or a 250 mL-baffled anaerobic flask. The cultures were grown at 37 °C in a rotary shaker (200 rpm) and induced with IPTG (1.0 mM) at OD_{600} = 0.35–0.45. At this time, the growth temperature was reduced to 30 °C, and the culture flasks were sealed with Parafilm M (Pechiney Plastic Packaging) to prevent product evaporation for aerobic cultures. Anaerobic cultures were sealed and the headspace was sparged with argon for 3 minutes immediately follow induction. Aerobic cultures were unsealed for 10 to 30 min every 24 h then resealed with Parafilm M, and additional glucose (1% (w/v)) was added 1 day post-induction. Samples were quantified after 3 d of cell culture.

Quantification of alcohols. Samples (2 ml) were removed from cell culture and cleared of biomass by centrifugation at 20,817g for 2 min using an Eppendorf 5417R centrifuge. The supernatant or cleared medium sample was then mixed in a 9:1 ratio with an aqueous solution containing the isobutanol internal standard (10,000 mg l^{-1}). These samples were then analyzed on a Trace GC Ultra (Thermo Scientific) using an HP-5MS column (0.25 mm × 30 m, 0.25 μM film thickness, J & W Scientific). The oven program was as follows: 75 °C for 3 min, ramp to 300 °C at 45 °C min^{-1}, 300 °C for 1 min. Alcohols were quantified by flame ionization detection (FID) (flow: 350 ml min^{-1} air, 35 ml min^{-1} H2 and 30 ml min^{-1} helium). Samples containing n-butanol levels below 500 mg l^{-1} were requantified after extraction of the cleared medium sample or standard (500 μl) with toluene (500 μl) containing the isobutanol internal standard (100 mg l^{-1}) using a Digital Vortex Mixer (Fisher) for 5 min set at 2,000. The organic layer was then quantified using the same GC parameters with a DSQII single-quadrupole mass

spectrometer (Thermo Scientific) using single-ion monitoring (m/z 41 and 56) concurrent with full scan mode (m/z 35–80). Samples were quantified relative to a standard curve of 2, 4, 8, 16, 31, 63, 125, 250, 500 mg l^{-1} n-butanol for MS detection or 125, 250, 500, 1,000, 2,000, 4,000, 8,000 mg l^{-1} n-butanol/ethanol for FID detection. Standard curves were prepared freshly during each run and normalized for injection volume using the internal isobutanol standard (100 or 1,000 mg l^{-1} for MS and FID, respectively).

2.3 Results and discussion

Biochemical dissection with mutant and truncated enzymes

Biochemical study of the properties of AdhE2 began by teasing out the activity of its two fused aldehyde and alcohol dehydrogenase domains in various contexts. In addition to the full-length enzyme, various truncations and active site mutations were also explored. The boundaries of the aldehyde and alcohol dehydrogenase domains were easily distinguished by sequence analysis using the Pfam database. Active site residues were also inferred by homology or by comparison to literature determinations[1]. In total, five constructs of AdhE2 were cloned (*Figure 2.3*): full length, full length with a mutated ALDH domain (C244A), full length with a mutated ADH domain (H721A, H735A), truncated ALDH domain, and truncated ADH domain.

Each AdhE2 construct was tagged with an N-terminal Strep-tag for heterologous expression in *E. coli* and affinity purification. A Strep-tag was chosen due to their high-purity, one step purifications with very mild purification conditions. The expression vector contained two tandem Tac promoters to drive expression of each gene. More common T7 vectors had been attempted previously, but this led to a very high fraction of the protein residing in the insoluble fraction after cell lysis. It is possible that the intrinsic oligomerization nature of AdhE2 predisposes it towards low solubility, even more so when highly expressed. Lower expression levels from the double Tac promoters resulted in lower total production, but a sufficient quantity remained in the soluble fraction throughout purification.

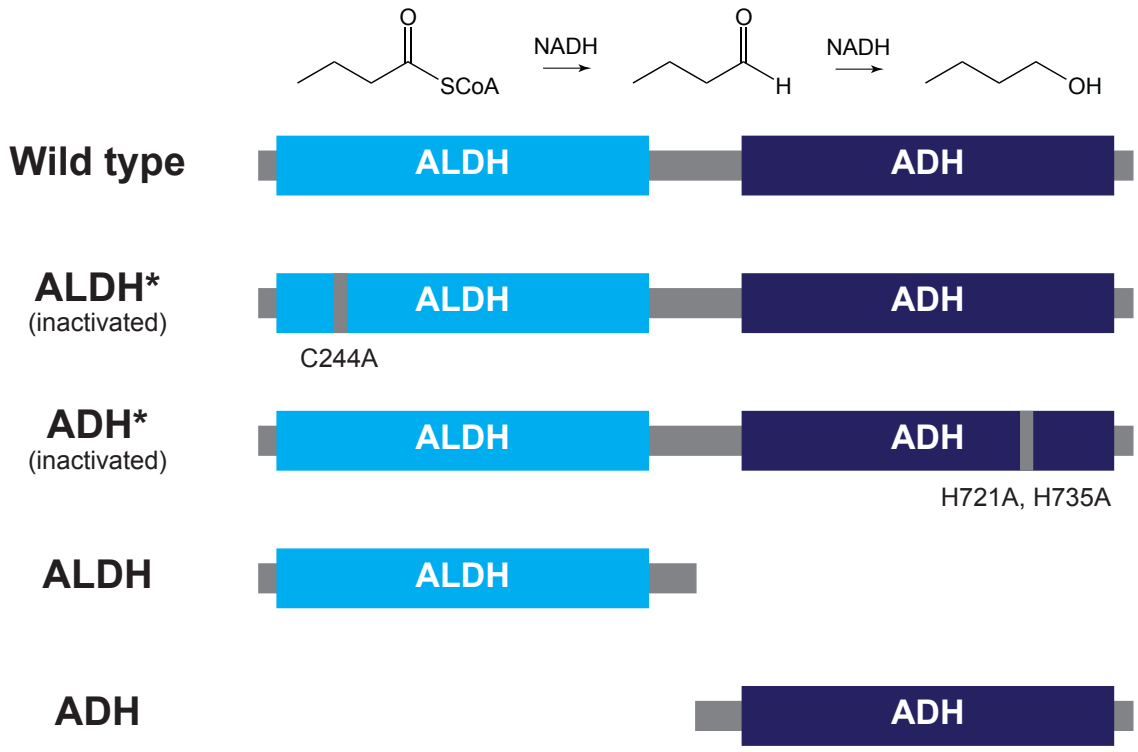


Figure 2.3 AdhE2 constructs for biochemical characterization
A collection of Adhe2 truncations and mutants were constructed for biochemical study. To examine the effect of a full-length enzyme with one domain inactivated, essential active site residues were mutated in the ALDH and ADH domains. Truncated enzymes were also prepared to study the activity of the domain in isolation.

In addition to lowering expression levels by switching from T7 to double Tac promoters, other expression condition variables were explored to maximize the amount of soluble protein that could be recovered. Expression in LB was compared to that of TB, but this had the effect of lowering total protein production without increasing the amount in the soluble fraction. The amount of IPTG used to induce the culture was varied, but this appeared to have minimal effect. Expression length and temperature were also evaluated with mixed results. Typical expressions were conducted at 30 °C for 4-5 hours. This was compared to overnight expressions at temperatures from 16-30 °C; overnight expression at low temperature was seen to mildly increase protein yield and soluble fraction, but standard expressions were generally carried out at 30 °C for 4-5 hours for time considerations. A typical purification is show in *Figure 2.4*. As seen in post-induction (lane 2) total protein production is relatively low. Most purifications of the full-length enzyme yielded approximately 5 mg of protein per liter of culture grown; this necessitated growing 6-12 liters of culture for typical expressions. As seen in lane 6, purifications were generally high-purity and free of significant contaminants. Following elution protein was either frozen at -80 °C with 10% glycerol for future assays, or dialyzed into SEC buffer overnight for oligomerization state studies.

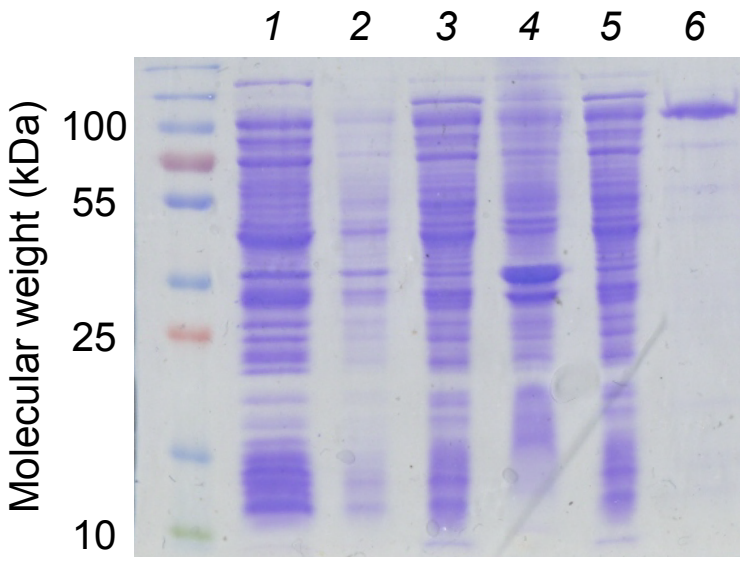


Figure 2.4 Purification of AdhE2
1. Pre-induction 2. Post-induction 3. Soluble fraction 4. Insoluble fraction 5. Flow-through 6. Elution of purified AdhE2. All AdhE2 constructs were purified to homogeneity using Strep-tags. AdhE2 is expressed poorly with a significant insoluble fraction, leading to typical yields of < 5 mg/L. Eluted protein was largely free from contaminants and was used without cleavage of the strep tag.

Given the dual active site architecture of AdhE2, *in vitro* kinetics afforded numerous areas for investigation[17-20], and we chose to focus on activity in the forward direction as the most relevant physiologically and to our butanol production pathway. Within the full-length wild type enzyme, three different activity measurements were possible for reactions in the forward direction, though not simultaneously. First, the activity of both domains active simultaneously was determined by measuring a decrease in absorbance at 340 nm, corresponding to oxidation of NADH within both active sites as acyl-CoA substrates were reduced to aldehydes by the ALDH domain and aldehyde substrates were reduced to alcohols by the ADH domain. This measurement is perhaps the most reflective of how the enzyme performs *in vivo*, but monitoring activity of both domains without the ability to distinguish between them is likely to obscure phenomena such as substrate channeling or domain coordination.

Second, the activity produced solely by the ALDH domain could be monitored by measuring release of free CoA after the acyl group is transferred to the active site cysteine. We initially tried to measure the release of free CoA with DTNB (Ellmans's reagent)[21]. DTNB reacts with free thiols and the reaction can be monitored by measuring an increase in absorbance at 412 nm. Although DTNB is commonly used for the quantification of thiols, this led to severely diminished activity of the enzyme. Some enzymes are not inhibited by DTNB and may be measured in a continuous fashion, but it is common for DTNB to result in loss of activity[22,23]. In this case a discontinuous assay could be performed by initiating the reaction, quenching at various time points, and then adding DTNB to react with the free CoA released during the reaction. This process has

the potential to be inaccurate and time-consuming, so other thiol reactive assays were investigated.

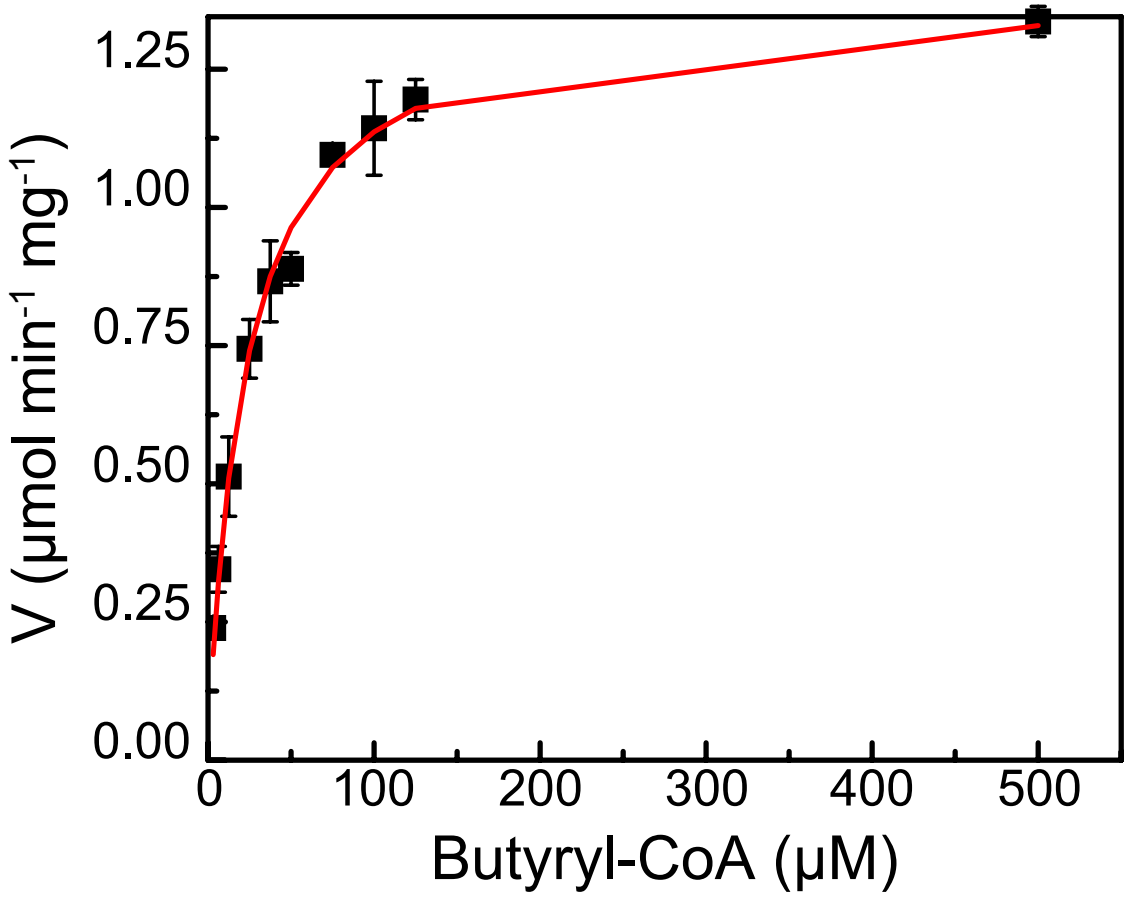


Figure 2.5 In vitro kinetics of AdhE2 with butyryl-CoA
AdhE2 displayed typical Michaelis–Menten kinetics with all substrates tested. The overall rate was relatively slow, highlighting the role of AdhE2 as a bottleneck in butanol production. Data are mean ± s.e. (n = 3).

A number of fluorogenic and chromogenic thiol-reactive dyes are known, and 7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methylcoumarin (CPM) was identified as being successfully used for *in vitro* free CoA releases assays[24]. Importantly, CPM did not appear to have any effect on the activity of AdhE2 and this allowed for a much faster continuous assay that could be carried out in a fluorescence plate reader.

The third activity measurement possible for the full-length wild type enzyme is measurement of NADH oxidation by the ADH domain when it is directly supplied with an aldehyde substrate. Measuring the activity solely produced by the ADH domain was thus relatively straightforward compared that required for measuring the ALDH domain independently.

A distinction should be made that in the case of supplying aldehyde substrates directly to the ADH domain it is the only activity occurring and the other domain is presumably inert. This is very different from the case of measuring ALDH domain activity through a free CoA release assay, as this confines the measurement to the ALDH domain despite the fact that both domains are still actively processing substrates. This distinction may be important if there is any coordination between the two domains: the free CoA release assay of the ALDH domain will capture any stimulatory (or inhibitory) effect the ADH domain exerts as it processes substrates, but the aldehyde substrate ADH assay will not include any stimulatory (or inhibitory) effect created by activity in the ALDH domain.

Moving on to assays conducted with the full-length domain inactivated enzyme there are again three possible assays. The ALDH domain in an inactive ADH context can be measured both by NADH oxidation and CPM free CoA release. This provides a good internal control as activity determined by both methods can be directly compared. This also gives some insight into whether the state of the ADH domain has an impact on the function of the ALDH domain. Finally, the ADH domain in an inactive ALDH context is again assayed by directly supplying aldehyde substrates.

Lastly, the truncated ALDH domain again offers two possible assays (NADH oxidation and CPM free CoA release), and the ADH domain is assayed by NADH oxidation with directly supplied aldehyde substrates. A representative Michaelis–Menten curve of full-length wild type AdhE2 with butyryl-CoA as the substrate is show in *Figure 2.5*. A complete dataset was collected for all constructs with both butyryl- and acetyl-CoA as well as butyraldehyde and acetaldehyde. All assays were performed with the same preparation of enzyme.

The complete dataset is shown in (*Table 2.6*). On a surface level, the activity of full length wild type AdhE2 is largely as expected from the product profile of its native host, *C. acetobutylicum*[2], as well as the behavior of our heterologous butanol production pathway in *E. coli*[25]. We see that from NADH oxidation data reporting on both domains active at the same time there is a seven-fold higher k_{cat}/K_M for butyryl-CoA that acetyl-CoA. This is the result of a 10-fold lower K_M for the larger substrate, despite a 50% faster k_{cat} for the smaller substrate. Importantly, the K_M for acetyl-CoA is still well within the physiological range of around 0.5-1.0 mM[26]. It is likely that under steady state conditions AdhE2 is at saturation for both substrates and capable of producing considerable amounts of ethanol.

Enzyme	Substrate	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} s^{-1})
WT AdhE2				
NADH oxidation	butyryl-CoA	2.2 ± 0.1	23 ± 2	(9.6 ± 0.3) x 10^4
(both domains)	acetyl-CoA	3.3 ± 0.1	250 ± 15	(1.3 ± 0.1) x 10^4
CoA release	butyryl-CoA	1.2 ± 0.1	10 ± 1	(1.1 ± 0.1) x 10^5
(ALDH domain)	acetyl-CoA	1.3 ± 0.1	100 ± 10	(1.3 ± 0.2) x 10^4
Aldehyde substrate	butyraldehyde	2.9 ± 0.1	4000 ± 400	(7.0 ± 0.2) x 10^2
(ADH domain)	acetaldehyde	5.6 ± 0.1	4500 ± 300	(1.2 ± 0.1) x 10^3
Domain mutants				
ALDH inactive	butyraldehyde	18.7 ± 1.3	2500 ± 500	(7.5 ± 0.7) x 10^3
	acetaldehyde	19.8 ± 1.0	2800 ± 400	(7.1 ± 0.3) x 10^3
ADH inactive	butyryl-CoA	0.3 ± 0.1	4 ± 1	(9.0 ± 1.0) x 10^4
	acetyl-CoA	1.3 ± 0.1	70 ± 10	(1.9 ± 0.4) x 10^4
Domain truncations				
ALDH	butyryl-CoA	< 0.1		
ADH	butyraldehyde	0.2 ± 0.1	300 ± 50	(5.8 ± 0.4) x 10^2
	acetaldehyde	0.2 ± 0.1	2000 ± 700	(1.1 ± 0.2) x 10^2

Table 2.6 Complete in vitro kinetic characterization of AdhE2
Extensive in vitro kinetics of various AdhE2 constructs highlighted several key attributes. As expected, AdhE2 displays a seven-fold higher k_{cat}/K_M for butyryl-CoA compared to acetyl-CoA, but the K_M for acetyl-CoA is still well within the physiological range. Next, inactivation of the ALDH domain appears to have a stimulatory effect on the ADH domain in the full-length context, but the opposite is true (ALDH domain inhibition) when the ADH domain is inactivated, although the effect is only seen for butyryl-CoA. Finally, the isolated ALDH domain is inactive, while the isolated ADH domain displays a lower k_{cat} and K_M.

Activity of the ALDH domain in the full-length wild type enzyme (as measured independently by free CoA release) is approximately half the rate of AdhE2, which is roughly in line with the expectation of measuring only half of the enzyme's total activity. When measuring the activity of the ADH domain independently by supplying aldehyde

substrates we see a modestly higher rate, but dramatically higher K_M with minimal difference between acetaldehyde and butyraldehyde. This suggests that the ALDH domain is the rate-limiting step of the complete reaction, and that any preference for butanol over ethanol production is controlled by the ALDH domain.

Looking next at the activity of the full-length enzyme with a domain inactivated, there appears to be some level of activation in the ADH domain when the ALDH domain is mutated. A greater than 10-fold increase in k_{cat}/K_M is seen for both aldehyde substrates, and shifts in both k_{cat} and K_M are responsible. In contrast there may be a mild deactivation of the ALDH domain when the ADH domain is inactive, where a 4-fold lower k_{cat} is seen in the case of a butyryl-CoA substrate.

Finally, truncating either domains seems to have a largely detrimental effect. The ALDH domain showed no detectable activity at any substrate or enzyme concentration, while the ADH domain showed marginal activity with a 100-fold lower k_{cat}/K_M relative to the potentially activated full-length protein with an inactive ALDH domain. Relative to the wild type enzyme the defect is much more moderate but only for butyraldehyde; acetaldehyde still has a 10-fold k_{cat}/K_M defect.

Oligomerization state and activity profile

To explore the potential role of oligomerization on the activity of AdhE2 we set out to characterize the range of oligomers most commonly formed and to measure the activity of different oligomeric states. To begin we analyzed purified protein with an analytical scale Superose 6 Size Exclusion Chromatography column (*Figure 2.7*). The column has a very broad fractionation range from 5-5,000 kDa. Here we saw that the bulk of the protein is found in a range of 5-10 oligomers, corresponding to a complexes of around 500-1,000 kDa. A small amount of protein was seen as a monomer but relatively little was seen as dimers or trimers. A significant amount of protein continued to be seen at greater sizes, up to and including the exclusion limit of the column of 5 mDa, corresponding to ~50 subunits.

Due to the analytical size of the column collecting fractions for activity assays was not practical. Instead we turned to a preparatory scale Superdex 200 column with a fractionation range of 10-600 kDa (*Figure 2.8*). Although this range is substantially narrower and limited to smaller complexes, the exclusion limit of 600 kDa still captures a good deal of the size distribution observed with the Superose 6 column. Importantly, this range is ideal for separation of monomers and dimers from larger complexes, and allowed us to asses whether or not monomers and dimers are active. Upon separation we observed a similar trend as before, a small fraction of the protein exists as monomers and dimers, relatively few complexes of three or four monomers are seen, and most of the protein exists as five monomers or greater. Fractions were collected throughout separation and at the conclusion of the separation all fractions were assayed with the CPM free CoA release assay used previously. The activity was normalized to the protein concentration and from this we saw that indeed monomers and dimers are largely inactive. The relative activity of monomers and dimers was less than 10% of high-order oligomers. Additionally, the activity plateaued after three or four monomers, and did not increase substantially at the largest oligomer sizes.

Taken together is clear that Adhe2 preferentially exists as a broad range of oligomers, and that it is largely inactive as a monomer or dimer. The implications of this behavior are less clear however. It is tempting to speculate that oligomerization is another means of substrate channeling. The local concentration of ADH active sites would be substantially higher in the oligomeric state compared to monomers or dimers. This effect is even greater when compared to independent monofunctional aldehyde and alcohol dehydrogenases as opposed to bifunctional aldehyde alcohol dehydrogenases such as AdhE2. These oligomeric complexes could also allow for intersubunit domain coordination instead of or in addition to coordination between domains on the same polypeptide chain.

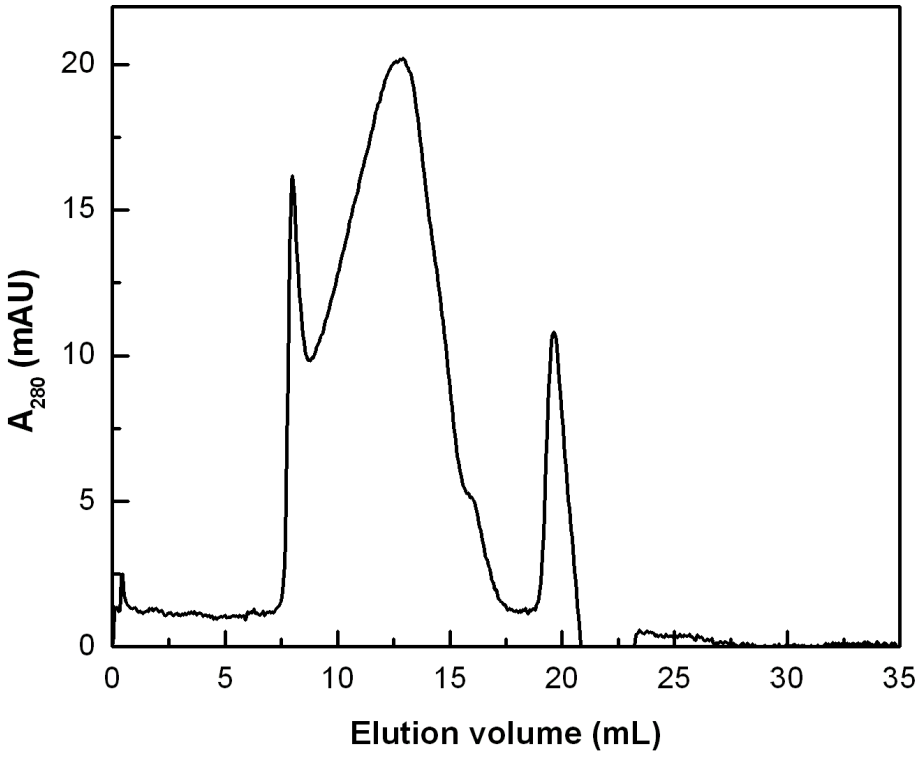


Figure 2.7 Broad oligomerization state of AdhE2
Analytical Size Exclusion Chromatography of AdhE2 demonstrates the very heterogeneous oligomerization state of AdhE2. Purified protein run over an analytical Superose 6 column (fractionation range: 5-5,000 kDa) elutes as a broad peak with the bulk of the protein in the 5-10 monomer range. A monomer peak is observed but relatively little protein exists as a dimer or trimer.

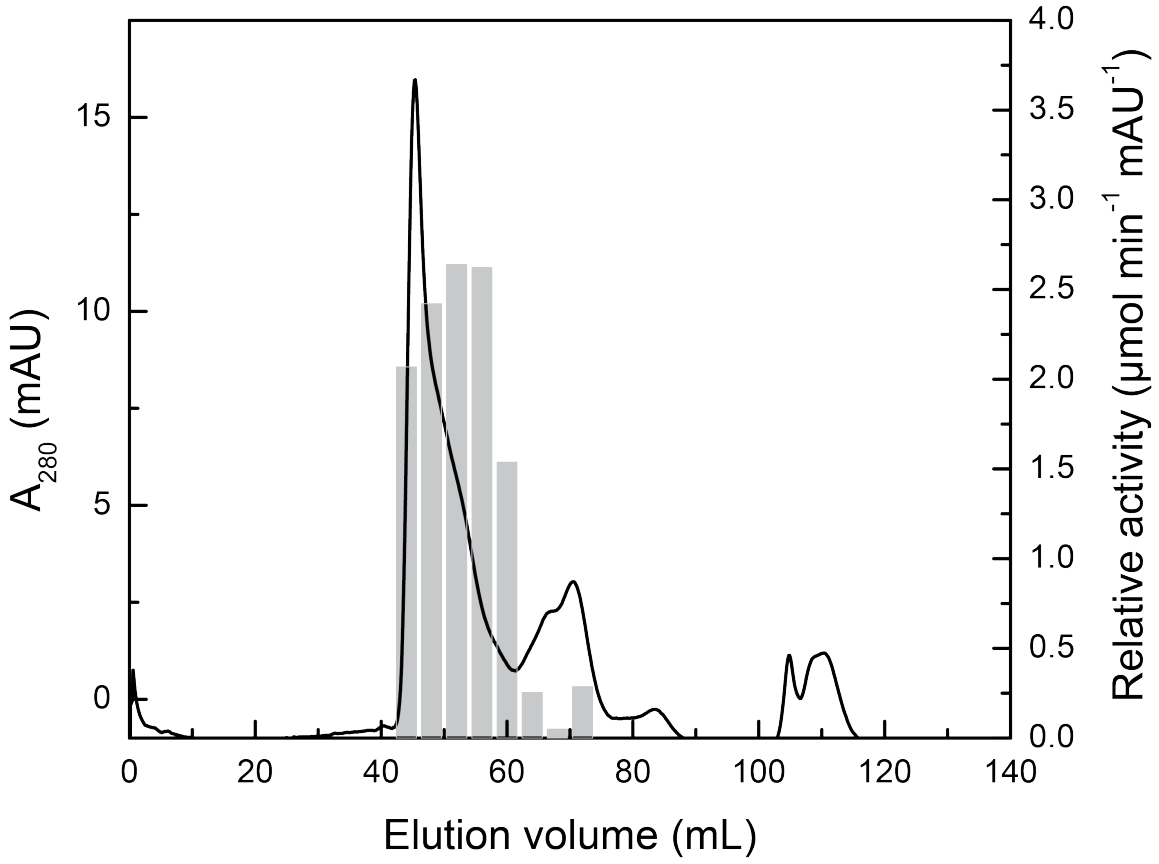


Figure 2.8 AdhE2 activity is oligomerization state dependent
To assess the activity of AdhE2 as a function of oligomerization state, purified protein was fractionated on a Superdex 200 SEC column (fractionation range: 10-600 kDa) to separate monomers and dimers from higher-order oligomers. Fractions were immediately assayed and normalized by protein content and revealed that only higher-order oligomers are significantly active.

In vivo butanol production with AdhE2

Turning to the behavior of Adhe2 *in vivo* with the goal of butanol production, it becomes clear from *Figure 2.9* how easily significant undesirable ethanol titer can be achieved. From the standpoint of redox requirements, under various conditions *E. coli* will produce 4 NADH per glucose consumed, resulting in 2 acetyl-CoA. Our desired butanol production pathway will recycle 4 NADH during the production of butanol from 2 acetyl-CoA. However, production of 2 ethanol from 2 acetyl-CoA will also consume 4 NADH, making each pathway equally redox balanced. Additionally, ethanol production may be advantageous as it is a faster route to turnover free CoA and may place less demand on the CoA pool.

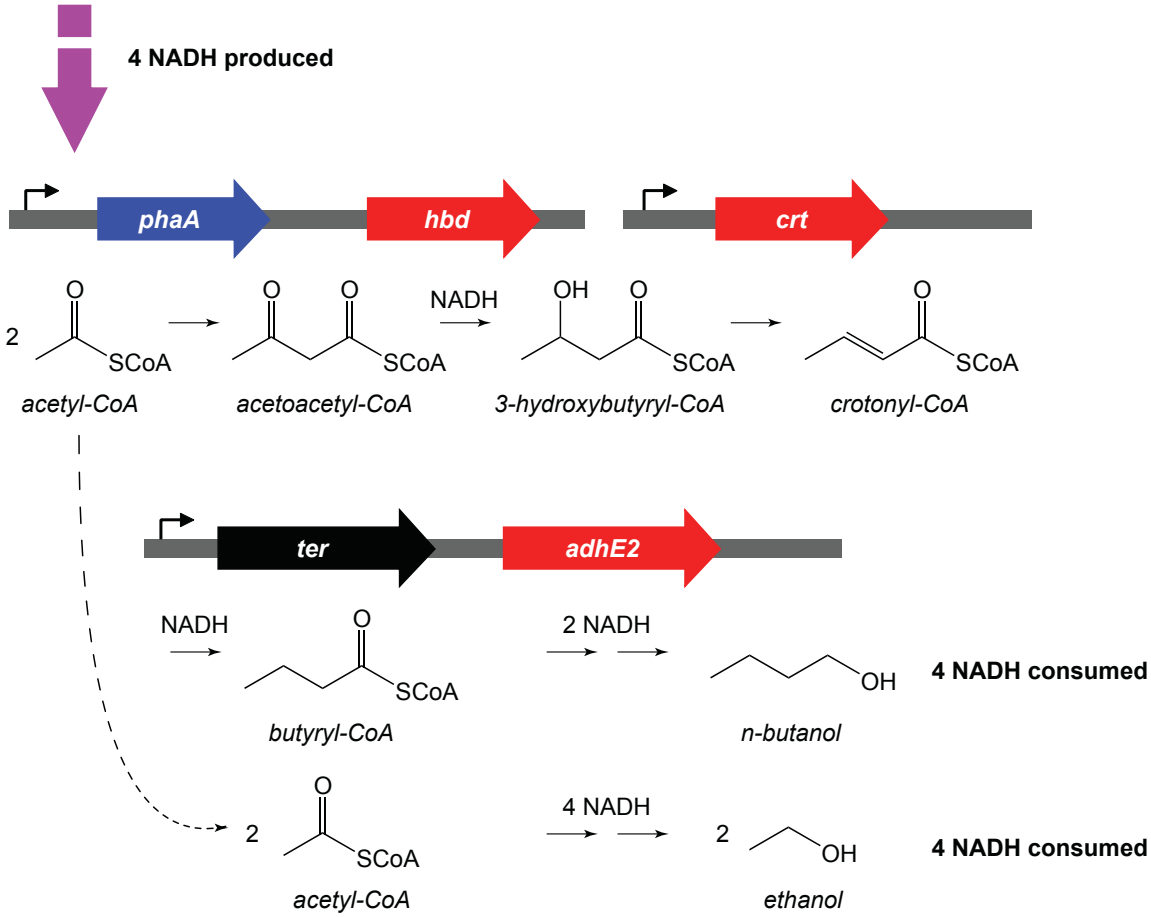


Figure 2.9 Pathway side products due to low substrate specificity
Low substrate specificity of AdhE2 enables unwanted ethanol production in addition to the desired production of butanol. Despite a seven-fold higher k_{cat}/ K_M for butyryl-CoA compared to acetyl-CoA, the K_M of acetyl-CoA is well within the physiological range and is likely to be at saturation for the enzyme. This pathway shortcut decreases carbon yield to butanol while still preserving redox balance and may benefit the cell through faster turnover of Coenzyme A.

To understand the magnitude of carbon losses to this side product, we cultured butanol production strains both aerobically and anaerobically. At that time our best production conditions were aerobic[25], but anaerobic conditions are usually preferred for large scale metabolite production. Under anaerobic conditions carbon flux to cell biomass is drastically limited. This is because ATP yield per glucose is very low during anaerobic growth without the benefit of oxidative phosphorylation using oxygen as a terminal electron acceptor. Very low ATP yield enforces two linked metabolic phenotypes: very low biomass accumulation and very high glucose consumption. With glycolysis serving as the only source of ATP for all cellular needs, making many new cells is simply outside the energy budget of the cell. This energy poverty also means that the redox requirements needed to balance large glycolytic flux must come from fermentation. Under anaerobic conditions it is the linkage of ATP yield, glycolytic flux, and fermentation redox balance that enables the high titer and high yield production of ethanol in yeast, which has been the biofuel gold standard for decades. Another factor favoring anaerobic production is the

fact that at industrial scales of fermentation it becomes a very challenging chemical engineering problem to deliver sufficient oxygen to a million liter fermenter filled with rapidly metabolizing cells.

As seen in *Figure 2.10*, our current pathway produces equivalent titers of butanol and ethanol under aerobic conditions. Switching to anaerobic growth yields even poorer results. Here butanol production is cut in half while ethanol production remains constant. This is largely a symptom of the metabolic requirement of low biomass accumulation under anaerobic conditions. Large amounts of biomass are derived from acetyl-CoA, which is also the entry point of our butanol production pathway. Anaerobically growing cells have limited acetyl-CoA availability; there is not sufficient energy to make building substantial new biomass a priority, and the vast majority of carbon is directed towards fermentation products not derived from acetyl-CoA. Under this low acetyl-CoA availability regime the butanol pathway is starved of reactants, and the relatively low K_M of AdhE2 for acetyl-CoA allows the side pathway to effectively compete for scarce resources.

As a first attempt at identifying alternative pathways that may not suffer from this poor substrate specificity, we again turned to native butanol producing *Clostridia*. Some *Clostridia* express independent monofunctional aldehyde and alcohol dehydrogenases in addition to the bifunctional aldehyde alcohol dehydrogenase AdhE2[27-29]. It is not fully understood to what extent and under what conditions these monofunctional dehydrogenases contribute to butanol production[3]. Furthermore, *Clostridia* characteristically produce a mixture of butanol and ethanol (as well as acetone), and some monofunctional aldehyde and alcohol dehydrogenases have been reported to be highly specific for one product or the other, in comparison to the promiscuous AdhE2. We initially assembled a butanol pathway employing a monofunctional aldehyde and alcohol dehydrogenase in place of AdhE2 (*Figure 2.11*), but this resulted in low titers. As *Clostridia* are strict anaerobes, some of their enzymes require strict anaerobic conditions to be functional[27,30]. Additionally, some *Clostridial* proteins may simply be insoluble when expressed outside their native host. Although these specific monofunctional aldehyde and alcohol dehydrogenases failed to support robust and specific butanol production, we sought to explore this approach more systematically, and this is discussed at length in Chapter 3.

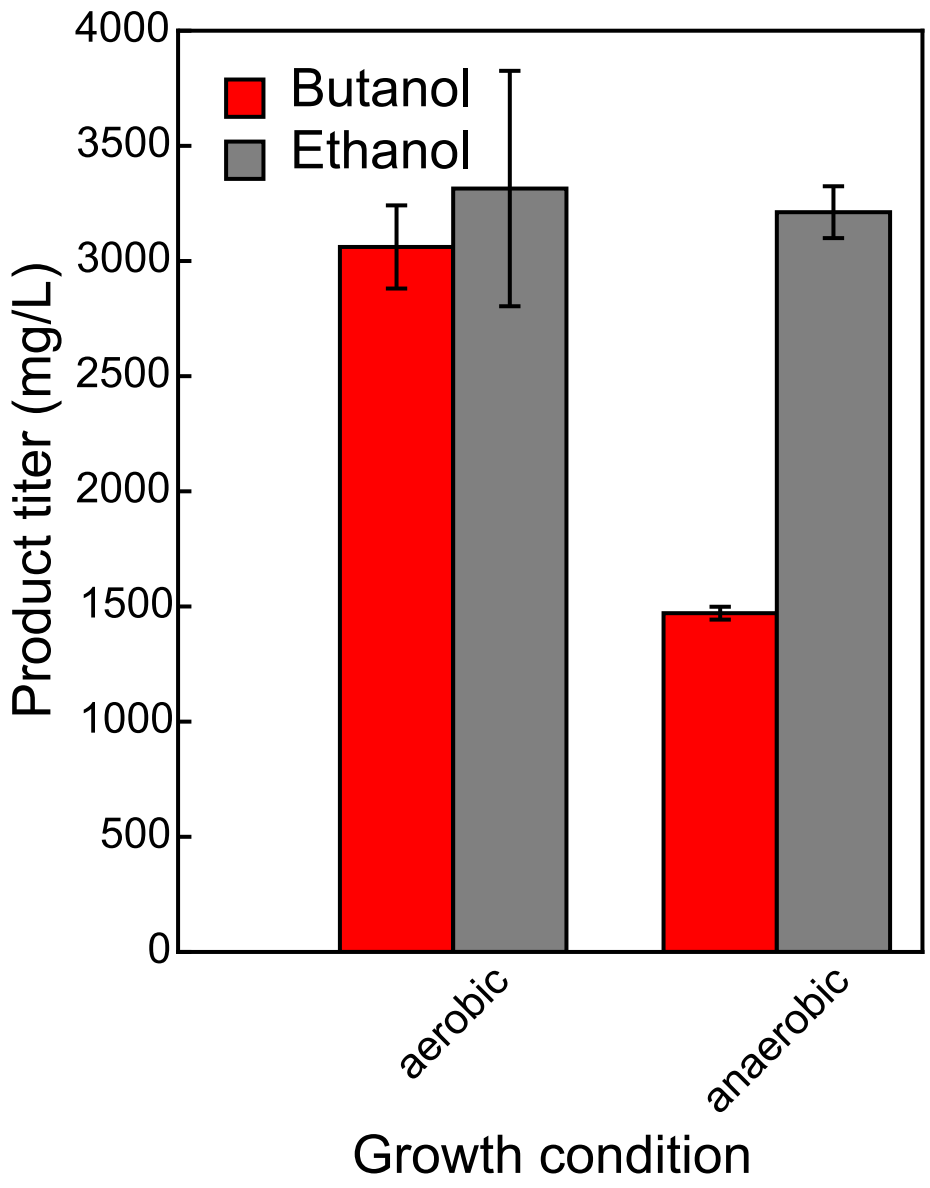


Figure 2.10 Butanol and ethanol production under aerobic and anaerobic conditions
The ratio of butanol and ethanol produced depends strongly on the growth condition. Under aerobic conditions acetyl-CoA availability is high and significant flux through the pathway allows butyryl-CoA to accumulate for reduction to butanol. However under anaerobic conditions acetyl-CoA availability is diminished; there remains sufficient acetyl-CoA for high-level ethanol production but flux to butyryl-CoA and butanol is limited. An aldehyde-alcohol dehydrogenase with greater specificity is needed to reduce butyryl-CoA while sufficiently discriminating against ubiquitous acetyl-CoA. Data are mean ± s.d. (n = 3).

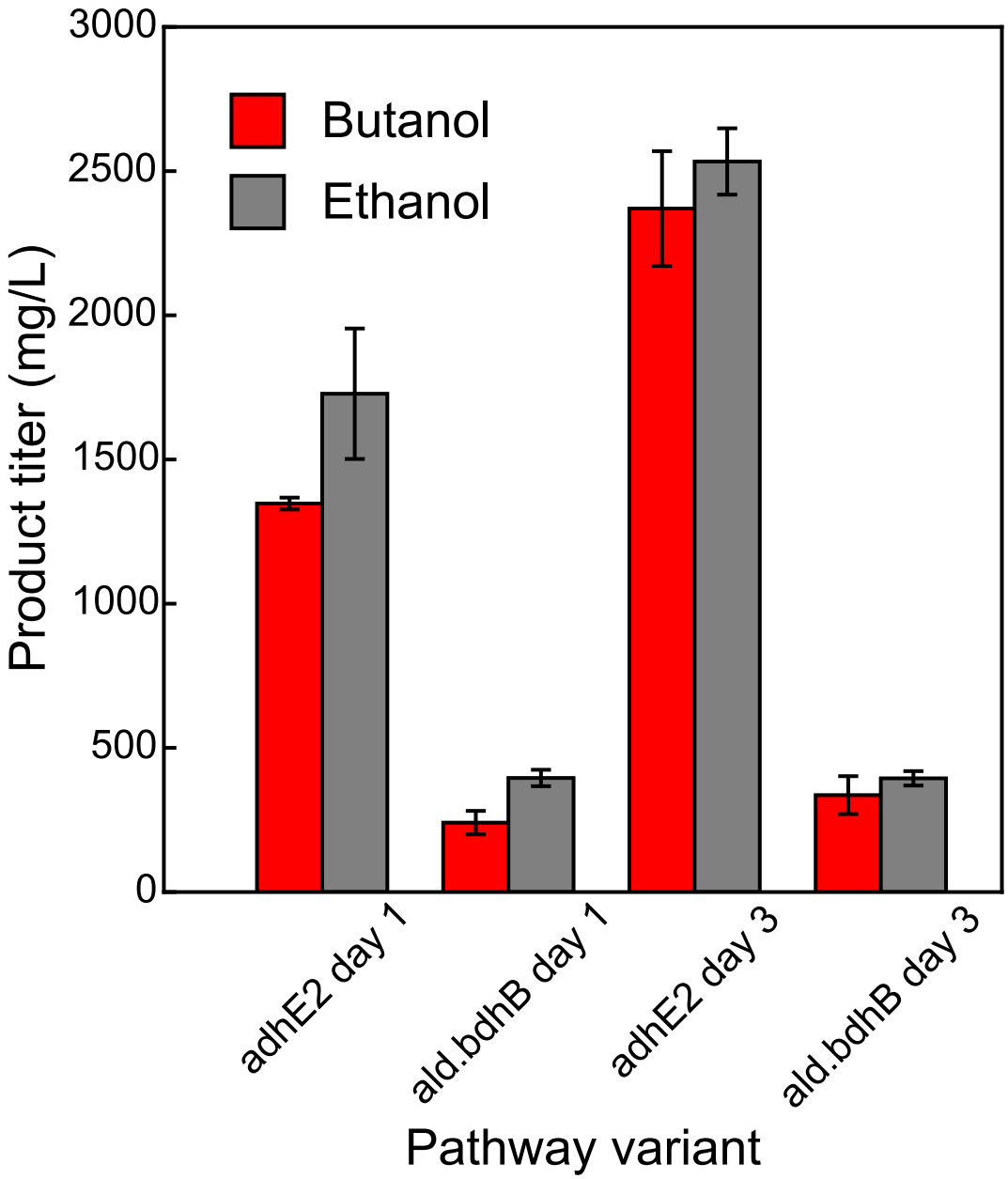


Figure 2.11 AdhE2 alternatives fail to support high butanol titer
In addition to bifunctional aldehyde-alcohol dehydrogenases, some native butanol-producing organisms express discrete monofunctional aldehyde and alcohol dehydrogenases. Some homologs of these dehydrogenases have been shown to poses considerable specificity for C4 substrates, but expression of these enzymes in E. coli resulted in very low butanol production. Some of these enzymes have been show to require strict anaerobic conditions for proper activity, and proper folding outside of their native organism may present a further challenge to functionally expressed protein. Data are mean ± s.d. (n = 3).

2.4 Conclusions

Through extended biochemical characterization we have determined the substrate specificity of AdhE2. As the *Clostridial* fermentation is adapted to producing a mixture of acetone, butanol, and ethanol it is no surprise the AdhE2 enzyme would be tailed to produce both alcohols. Although some *Clostridia* do express monofunctional ALDHs and ADHs that are thought to be more specific, it is not yet clear what their role is relative to AdheE2[2,27,29].

Biochemically AdhE2 presents a number of unique properties, chief among them its bifunctional nature which may provide a mechanism for substrate channeling of volatile and reactive aldehyde intermediates[20]. We observed some evidence of coordination between the domains by measuring the activity of one domain while the other was catalytically inactivated, but we did not conclusively prove domain coordination or substrate channeling. Attempts were made to capture any freely diffusing aldehydes with semicarbazide[8], but these were unsuccessful. Interestingly, if aldehydes did freely diffuse into solution before being bound in the ADH active site it would take > 10^{10} seconds for the ADH domain to reach steady state; this seems improbable.

In regard to the effect of oligomerization on activity, we demonstrated that monomers and dimers are inactive; trimers or larger structures were needed to observe activity and we did not observe significantly increased activity once that threshold had been passed. In attempts to alter the relationship between oligomerization state and activity we assessed mild detergents to break up complexes but this was inconclusive. We also investigated reports that truncation of 12 amino acids from the C-terminus eliminated oligomerization, but the effect on activity was not reported[5]. In our hands truncation of 8 or 12 residues did disrupt oligomerization as well as activity, while truncation of 4 residues showed wild type behavior.

Finally, we examined using phenanthroline to chelate iron from the media during expression or by adding it directly to purified protein[31]. The ADH domain of AdhE2 requires a bound iron atom, which could play a structural role as well as its importance for catalysis. However these studies were inconclusive.

Potentially fruitful future work on AdhE2 could focus on a thorough investigation of ways in which the two domains may be coordinated[18,32,33]. This could be achieved through mutational study of domain interfaces, exploration of conformational changes that may take place when either domain binds substrate or NADH[34], or attempting to lock one domain into a substrate bound state. The later could likely be achieved through synthesis of non-hydrolyzable acyl-CoA analogs and could prove very informative.

Structural information would be extremely interesting to obtain. The tertiary structure of the core ALDH or ADH domains is unlikely to be significantly different from monofunctional enzymes; instead it is the interface between the two domains that is more likely to be of interest. Indeed, structures of homologs of both the ALDH and ADH domains been solved but are unremarkable[35,36]. Solving a structure of the oligomeric form seems improbable given the very heterogeneous nature of the oligomer. A structure of the monomeric or dimeric form may prove feasible, but its relevance is questionable given that those forms are inactive. Given the advances in electron microscopy in recent

years a fresh round of EM studies may be warranted[37], but methods for dealing with the heterogeneous oligomeric state call this into question.

2.5 References

1. L. Fontaine, I. Meynial-Salles, L. Girbal, X. Yang, C. Croux and P. Soucaille, Molecular characterization and transcriptional analysis of adhE2, the gene encoding the NADH-dependent aldehyde/alcohol dehydrogenase responsible for butanol production in alcohologenic cultures of Clostridium acetobutylicum ATCC 824, *J. Bacteriol.* **2002**, *184*, 821–830.
2. S. Lee, J. Park, S. Jang, L. Nielsen, J. Kim and K. Jung, Fermentative butanol production by Clostridia, *Biotechnol. Bioeng.* **2008**, *101*, 209–228.
3. R. Gheshlaghi, J. M. Scharer, M. Moo-Young and C. P. Chou, Metabolic pathways of clostridia for producing butanol, *Biotechnol. Adv.* **2009**, *27*, 764–781.
4. L. B. Sánchez, Aldehyde Dehydrogenase (CoA-Acetylating) and the Mechanism of Ethanol Formation in the Amitochondriate Protist,Giardia lamblia, *Arch. Biochem. Biophys.* **1998**, *354*, 57–64.
5. A. Espinosa, The Bifunctional Entamoeba histolytica Alcohol Dehydrogenase 2 (EhADH2) Protein Is Necessary for Amebic Growth and Survival and Requires an Intact C-terminal Domain for Both Alcohol Dehydrogenase and Acetaldehyde Dehydrogenase Activity, *J. Biol. Chem.* **2001**, *276*, 20136–20143.
6. R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, *et al.*, Pfam: the protein families database, *Nucleic Acids Res.* **2014**, *42*, D222–30.
7. A. M. Kunjapur, Y. Tarasova and K. L. J. Prather, Synthesis and accumulation of aromatic aldehydes in an engineered strain of Escherichia coli, *J. Am. Chem. Soc.* **2014**, *136*, 11644–11654.
8. S. Gómez-Manzo, J. E. Escamilla, A. González-Valdez, G. López-Velázquez, A. Vanoye-Carlo, J. Marcial-Quino, I. de la Mora-de la Mora, I. Garcia-Torres, S. Enríquez-Flores, M. L. Contreras-Zentella, *et al.*, The oxidative fermentation of ethanol in Gluconacetobacter diazotrophicus is a two-step pathway catalyzed by a single enzyme: alcohol-aldehyde Dehydrogenase (ADHa), *Int. J. Mol. Sci.* **2015**, *16*, 1293–1311.
9. M. S. Patel, N. S. Nemeria, W. Furey and F. Jordan, The pyruvate dehydrogenase complexes: structure-based function and regulation, *J. Biol. Chem.* **2014**, *289*, 16615–16623.
10. M. F. Dunn, D. Niks, H. Ngo, T. R. M. Barends and I. Schlichting, Tryptophan synthase: the workings of a channeling nanomachine, *Trends Biochem. Sci.* **2008**, *33*, 254–264.

11. D. Kessler, W. Herth and J. Knappe, Ultrastructure and pyruvate formate-lyase radical quenching property of the multienzymic AdhE protein of Escherichia coli, *J. Biol. Chem.* **1992**, *267*, 18073–18079.

12. M. Chen, E. Li and S. L. Stanley Jr., Structural analysis of the acetaldehyde dehydrogenase activity of Entamoeba histolytica alcohol dehydrogenase 2 (EhADH2), a member of the ADHE enzyme family, *Mol. Biochem. Parasitol.* **2004**, *137*, 201–205.

13. R. Laurenceau, P. V. Krasteva, A. Diallo, S. Ouarti, M. Duchateau, C. Malosse, J. Chamot-Rooke and R. Fronzes, Conserved Streptococcus pneumoniae spirosomes suggest a single type of transformation pilus in competence, *PLoS Pathog.* **2015**, *11*, e1004835.

14. S. An, R. Kumar, E. D. Sheets and S. J. Benkovic, Reversible compartmentalization of de novo purine biosynthetic complexes in living cells, *Science* **2008**, *320*, 103–106.

15. R. Narayanaswamy, M. Levy, M. Tsechansky, G. M. Stovall, J. D. O'Connell, J. Mirrielees, A. D. Ellington and E. M. Marcotte, Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 10147–10152.

16. E. Puchulu-Campanella, H. Chu, D. J. Anstee, J. A. Galan, W. A. Tao and P. S. Low, Identification of the components of a glycolytic enzyme metabolon on the human red blood cell membrane, *J. Biol. Chem.* **2013**, *288*, 848–858.

17. P. Baker, D. Pan, J. Carere, A. Rossi, W. Wang and S. Y. K. Seah, Characterization of an Aldolase−Dehydrogenase Complex That Exhibits Substrate Channeling in the Polychlorinated Biphenyls Degradation Pathway, *Biochemistry* **2009**, *48*, 6551–6558.

18. C. E. Atreya, Kinetic Characterization of Bifunctional Thymidylate Synthase-Dihydrofolate Reductase (TS-DHFR) from Cryptosporidium hominis, *J. Biol. Chem.* **2004**, *279*, 18314–18322.

19. N. Nagradova, Interdomain Communications in Bifunctional Enzymes: How Are Different Activities Coordinated? *IUBMB Life* **2003**, *55*, 459–466.

20. M. F. Dunn, Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase bienzyme complex, *Arch. Biochem. Biophys.* **2012**, *519*, 154–166.

21. G. L. Ellman, Tissue sulfhydryl groups, *Arch. Biochem. Biophys.* **1959**, *82*, 70–77.

22. D. W. Pettigrew, Inactivation of Escherichia coli glycerol kinase by 5,5'-dithiobis(2-nitrobenzoic acid) and N-ethylmaleimide: evidence for nucleotide regulatory binding sites, *Biochemistry* **1986**, *25*, 4711–4718.

23. W. X. Tian, R. Y. Hsu and Y. S. Wang, Studies on the reactivity of the essential sulfhydryl groups as a conformational probe for the fatty acid synthetase of chicken liver. Inactivation by 5,5'-dithiobis-(2-nitrobenzoic acid) and intersubunit cross-linking of the inactivated enzyme, *J. Biol. Chem.* **1985**, *260*, 11375–11387.

24. C. C. Chung, K. Ohwaki, J. E. Schneeweis, E. Stec, J. P. Varnerin, P. N. Goudreau, A. Chang, J. Cassaday, L. Yang, T. Yamakawa, *et al.*, A fluorescence-based thiol quantification assay for ultra-high-throughput screening for inhibitors of coenzyme A production, *Assay. Drug Dev. Technol.* **2008**, *6*, 361–374.

25. B. B. Bond-Watts, R. J. Bellerose and M. C. Y. Chang, Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways, *Nat. Chem. Biol.* **2011**, *7*, 222–227.

26. A. Danchin, L. Dondon and J. Daniel, Metabolic alterations mediated by 2-ketobutyrate in Escherichia coli K12, *Mol. Gen. Genet.* **1984**, *193*, 473–478.

27. R. T. Yan and J. S. Chen, Coenzyme A-acylating aldehyde dehydrogenase from Clostridium beijerinckii NRRL B592, *Appl. Env. Microbiol.* **1990**, *56*, 2591–2599.

28. J. Toth, A. A. Ismaiel and J.-S. Chen, The ald Gene, Encoding a Coenzyme A-Acylating Aldehyde Dehydrogenase, Distinguishes Clostridium beijerinckii and Two Other Solvent-Producing Clostridia fromClostridium acetobutylicum, *Appl. Env. Microbiol.* **1999**, *65*, 4973–4980.

29. K. Walter, G. Bennett and E. Papoutsakis, Molecular characterization of two Clostridium acetobutylicum ATCC 824 butanol dehydrogenase isozyme genes, *J. Bacteriol.* **1992**, *174*, 7149–7158.

30. R. W. Welch, F. B. Rudolph and E. T. Papoutsakis, Purification and characterization of the NADH-dependent butanol dehydrogenase from Clostridium acetobutylicum (ATCC 824), *Arch. Biochem. Biophys.* **1989**, *273*, 309–318.

31. A. Espinosa, G. Perdrizet, G. Paz-y-Mino C, R. Lanfranchi and M. Phay, Effects of iron depletion on Entamoeba histolytica alcohol dehydrogenase 2 (EhADH2) and trophozoite growth: implications for antiamoebic therapy, *J. Antimicrob. Chemother.* **2009**, *63*, 675–678.

32. A. K. Bera, S. Chen, J. L. Smith and H. Zalkin, Interdomain signaling in glutamine phosphoribosylpyrophosphate amidotransferase, *J. Biol. Chem.* **1999**, *274*, 36498–36504.

33. P.-H. Liang and K. S. Anderson, Substrate Channeling and Domain−Domain Interactions in Bifunctional Thymidylate Synthase−Dihydrofolate Reductase †, *Biochemistry* **1998**, *37*, 12195–12205.

34. S. Chen, J. W. Burgner, J. M. Krahn, J. L. Smith and H. Zalkin, Tryptophan fluorescence monitors multiple conformational changes required for glutamine phosphoribosylpyrophosphate amidotransferase interdomain signaling and catalysis, *Biochemistry* **1999**, *38*, 11659–11669.

35. A. J. Stein, A. Weger, L. Volkart, M. Gu and A. Joachimiak, The Crystal Structure of the ACDH domain of an Alcohol Dehydrogenase from Vibrio parahaemolyticus to 2.25A, *Protein Data Bank* **2010**,

36. J. Extance, S. J. Crennell, K. Eley, R. Cripps, D. W. Hough and M. J. Danson, Structure of a bifunctional alcohol dehydrogenase involved in bioethanol generation in Geobacillus thermoglucosidasius, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 2104–2115.

37. J. L. S. Milne, M. J. Borgnia, A. Bartesaghi, E. E. H. Tran, L. A. Earl, D. M. Schauder, J. Lengyel, J. Pierson, A. Patwardhan and S. Subramaniam, Cryo-electron microscopy--a primer for the non-microscopist, *FEBS J.* **2013**, *280*, 28–45.

Chapter 3: *Improved butanol production through aldehyde-alcohol dehydrogenase screening and whole-genome mutagenesis*

3.1 Introduction

Following the initial failure of monofunctional ALDH/ADH replacements for AdhE2 we chose to take a more comprehensive approach. In collaboration with Calysta Biosystems (Menlo Park, CA) we pursued bioinformatic methods to identify AdhE2 homologs that may have greater substrate specificity[1]. These homologs were then screened in our butanol production pathway, as well as used as a source of diversity for mutations of AdhE2. Screening homologs has often proved useful an many contexts and it was employed successfully here. Other systematic methods of improving enzyme properties without the use of traditional high-throughput directed evolution have also proven successful, and is likely to be a key technique for improving engineered pathways in the future[2,3].

A related goal was to use a high specificity butanol production pathway as a platform for mutagenesis and selection of mutants producing more butanol. Despite the success of focused, lower-throughput approaches described above, the diversity of techniques available once a high-throughput screen or selection has been developed will always be attractive[4]. With such a screen or selection in hand our yield and titer could be dramatically improved, but just as importantly it could give us insight on regulatory mechanisms impacting central carbon flux. This selection was made possible by employing the high butanol specificity pathway in a fermentation knockout strain that can only grow anaerobically by producing butanol. Butanol production balances the redox requirements of glycolysis, which in this strain is the only source of ATP production, thus tying butanol production to energy generation and cell survival. Similar mutagenesis and selection strategies have been successfully used to interrogate complex phenotypes[5,6], and there is a rich history of industrial strain improvement for metabolite production[7-9].

This anaerobic growth selection was validated and then employed with whole genome mutagenesis for two successive rounds of selection. Mutant high production strains were sequenced which revealed a diverse set of mechanisms that could play a role in carbon flux decisions. In the future these mutations could be combinatorially explored to enable further pathway improvements[10,11], and the mechanism through which these mutations act could be elucidated to aide in further rational mutant design.

3.2 Materials and methods

Commercial materials. Luria-Bertani (LB) Broth Miller, LB Agar Miller, and Terrific Broth (TB) were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), isopropyl-β-D-thiogalactopyranoside (IPTG), phenylmethanesulfonyl fluoride (PMSF), tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), sodium chloride, dithiothreitol (DTT), kanamycin (Km), ethyl acetate and ethylene diamine tetraacetic acid disodium dihydrate (EDTA), were purchased from Fisher Scientific (Pittsburgh, PA). Coenzyme A trilithium salt (CoA), acetyl-CoA, nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), β-mercaptoethanol, sodium phosphate dibasic hepthydrate, and N,N,N',N'-tetramethyl-ethane-1,2-diamine (TEMED)

were purchased from Sigma-Aldrich (St. Louis, MO). Acrylamide/Bis-acrylamide (30%, 37.5:1), electrophoresis grade sodium dodecyl sulfate (SDS), Bio-Rad protein assay dye reagent concentrate and ammonium persulfate were purchased from Bio-Rad Laboratories (Hercules, CA). Restriction enzymes, T4 DNA ligase, Phusion DNA polymerase, T5 exonuclease, and Taq DNA ligase were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs) and Platinum Taq High-Fidelity polymerase (Pt Taq HF) were purchased from Invitrogen (Carlsbad, CA). PageRuler™ Plus prestained protein ladder was purchased from Fermentas (Glen Burnie, Maryland). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a stock concentration of 100 μM in 10 mM Tris-HCl, pH 8.5, and stored at either 4°C for immediate use or -20°C for longer term use. DNA purification kits and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). Amicon Ultra 10,000 centrifugal concentrators were purchased from EMD Millipore (Billerica, MA).

Bacterial strains. *E. coli* DH10B-T1^R and BL21(de3)T1^R were used for DNA construction and heterologous protein production, respectively. *E. coli* DH1 and DH1 ΔackA-pta ΔadhE ΔldhA ΔpoxB ΔfrdBC (MC1.24) were used for metabolite production.

Gene and plasmid construction. Restriction enzyme cloning, Gibson assembly, and Golden Gate assembly were used to carry out plasmid construction. All PCR amplifications were carried out with Phusion or Platinum Taq High Fidelity DNA polymerases. All constructs were verified by sequencing (Quintara Biosciences; Berkeley, CA).

Cell culture. *E. coli* strains were transformed by electroporation using the appropriate plasmids. A single colony from a fresh transformation was then used to seed an overnight culture grown in Terrific Broth (TB) (EMD Biosciences) supplemented with 1.5% (w/v) glucose and appropriate antibiotics at 37 °C in a rotary shaker (200 rpm). Antibiotics were used at a concentration of 50 μg ml^{-1} for strains with a single resistance marker. For strains with multiple resistance markers, kanamycin and chloramphenicol were used at 25 μg ml^{-1} and carbenicillin was used at 50 μg ml^{-1}.

***In vivo* production of alcohols.** Overnight cultures of freshly transformed *E. coli* strains were grown for 12–16 h in TB at 37 °C and used to inoculate TB (50 ml) with glucose replacing the standard glycerol supplement (1.5% (w/v) glucose for aerobic cultures and 2.5% (w/v) glucose for anaerobic cultures) and appropriate antibiotics to an optical density at 600 nm (OD_{600}) of 0.05 in a 250 mL-baffled flask or a 250 mL-baffled anaerobic flask. The cultures were grown at 37 °C in a rotary shaker (200 rpm) and induced with IPTG (1.0 mM) at OD_{600} = 0.35–0.45. At this time, the growth temperature was reduced to 30 °C, and the culture flasks were sealed with Parafilm M (Pechiney Plastic Packaging) to prevent product evaporation for aerobic cultures. Anaerobic cultures were sealed and the headspace was sparged with argon for 3 minutes immediately follow induction. Aerobic cultures were unsealed for 10 to 30 min every 24 h then resealed with Parafilm M, and additional glucose (1% (w/v)) was added 1 day post-induction. Samples were quantified after 3 d of cell culture.

Quantification of alcohols. Samples (2 ml) were removed from cell culture and cleared of biomass by centrifugation at 20,817g for 2 min using an Eppendorf 5417R

centrifuge. The supernatant or cleared medium sample was then mixed in a 9:1 ratio with an aqueous solution containing the isobutanol internal standard (10,000 mg l^{-1}). These samples were then analyzed on a Trace GC Ultra (Thermo Scientific) using an HP-5MS column (0.25 mm × 30 m, 0.25 μM film thickness, J & W Scientific). The oven program was as follows: 75 °C for 3 min, ramp to 300 °C at 45 °C min^{-1}, 300 °C for 1 min. Alcohols were quantified by flame ionization detection (FID) (flow: 350 ml min^{-1} air, 35 ml min^{-1} H2 and 30 ml min^{-1} helium). Samples containing n-butanol levels below 500 mg l^{-1} were requantified after extraction of the cleared medium sample or standard (500 μl) with toluene (500 μl) containing the isobutanol internal standard (100 mg l^{-1}) using a Digital Vortex Mixer (Fisher) for 5 min set at 2,000. The organic layer was then quantified using the same GC parameters with a DSQII single-quadrupole mass spectrometer (Thermo Scientific) using single-ion monitoring (m/z 41 and 56) concurrent with full scan mode (m/z 35–80). Samples were quantified relative to a standard curve of 2, 4, 8, 16, 31, 63, 125, 250, 500 mg l^{-1} n-butanol for MS detection or 125, 250, 500, 1,000, 2,000, 4,000, 8,000 mg l^{-1} n-butanol/ethanol for FID detection. Standard curves were prepared freshly during each run and normalized for injection volume using the internal isobutanol standard (100 or 1,000 mg l^{-1} for MS and FID, respectively).

Anaerobic growth enrichment validation. MC1.24 transformed with butanol production plasmids capable of a range of titers were mixed at various ratios and cultured anaerobically as described above. Flasks were sampled with a syringe to collect culture media supernatants for quantification of metabolites and to measure growth. Pelleted cells were used as template for qPCR of butanol plasmids to determine the relative abundance of different subpopulations and compared to a standard curve of purified plasmids.

EMS mutagenesis. MC1.24 transformed with pT5T33-phaA.HBD, pCWO.trc-ter-aldh46.adh, and pBBR2-aceE.F.lpd were grown 12–16 h in LB at 37 °C and used to inoculate LB (50 ml) cultures and appropriate antibiotics to an optical density at 600 nm (OD_{600}) of 0.05 in a 250 mL-baffled flask or a 250 mL-baffled anaerobic flask. The cultures were grown at 37 °C in a rotary shaker (200 rpm) to an OD_{600} of 0.4 and 2 mL of culture in triplicate was gently pelleted. Cells were washed three times in 1X PBS and resuspended in 2 mL 1X PBS to which 35 μL of EMS were added. A second triplicate of cultures was treated identically with the omission of EMS. Cells were incubated shaking at 37 °C for 45 minutes before washing three times with 1X PBS. Cells were resuspended in 50 mL LB in a 250 mL baffled anaerobic flask supplemented with 2.5 % glucose and appropriate antibiotics. Cultures were grown at 37 °C in a rotary shaker (200 rpm) for 2 hours before inducing with IPTG (1.0 mM). At this time, the flasks were sealed and the headspace was sparged with argon for 3 minutes. The growth temperature was reduced to 30 °C and the cultures were grown overnight.

Selection of mutagenized cultures. Following EMS mutagenesis cultures were serially transferred to fresh media every 24-72 hours to approximate continuous growth with limited time spent in stationary phase. The initial growth media was MOPS M9 minimal media supplemented with 10% LB, 2.5% glucose, 1 mM IPTG, and appropriate antibiotics. Culture OD_{600} was monitored daily and cultures were transferred when the majority of cultures were in late log-phase growth, usually OD_{600} 0.5-1.0 depending on media composition. Culture supernatant samples (2 mL) were collected for metabolite

quantification. All cultures were transferred simultaneously, the headspace was sparged with argon for 3 minutes, and growth was continued at 30 °C in a rotary shaker (200 rpm). Selections were continued for up to three weeks and the culture media LB supplement was tapered to 0% over time. Selections were terminated when EMS treated cultures ceased growth rate improvement or when the growth rate of untreated cultures began to improve. Final cultures were stored as 15% glycerol stocks at -80 °C in addition to being streaked on MOPS M9 1% glucose agar plates. Individual colonies were picked and cultured for butanol production in MOPS M9 or TB to confirm butanol production relative to wild type strains.

Whole genome sequencing of mutagenized strains. Genomic DNA was isolated from verified high production clones using Qiagen Genomic-tip 100/G kits according to the manufactures instructions. Genomic DNA was fragmented with Fragmentase (NEB) to approximately 400-700 bp fragments as determined by Bio-Rad Experion Electrophoresis System. Illumina libraries were prepared with the NEBNext DNA library prep master mix set following the manufactures instructions. MiSeq paired-end 250 base sequencing was performed by the UC Davis Genome Center (Davis, CA).

Identification of mutations in evolved strains. Illumina read data was processed with the Breseq pipeline[12] using the DH1 genome (CP001637) as the reference. Mutations were classified by cellular function and scored by relevance to metabolic pathways.

3.3 Results and discussion

Improving AdhE2 through incorporation of natural sequence diversity

We began by collaborating with Calysta Biosystems to bioinformatically identify AdhE2 homologs that may have altered substrate specificity. This was initiated by searching the literature for ALDH domains of characterized substrate specificity (*Table 3.1*).

The set of identified homologs consisted of three bifunctional AdhE2 homologs that were all characterized as C2 specific, four C4 specific monofunctional ALDHs, and one atypical C2 specific ALDH. The sequences were all of bacterial origin except for and AdhE2 homolog identified from the protozoan parasite *Giardia intestinalis.* Only the specificity of the ALDH domain was considered, as our previous characterization of AdhE2 showed that the ADH domain confers little specificity to the overall reaction, thus the ALDH domain is considered the gatekeeper and specificity improvement was focused there exclusively.

	Gene	Accession	Organism	Reference
C4 preference	AdhE2	NP_149199	Clostridium acetobutylicum	This study
	ALD	AAD31841	Clostridium beijerinckii	13
	ALDH	YP_001310903	Clostridium beijerinckii	14
	PduP	BAG26139	Lactobacilus reuteri	15
	ALDH	YP_003687877	Propionibacterium freudenreichii	16
C2 preference	AdhE	NP_415757	Escherichia coli	17
	AdhE	AAC47539	Giardia intestinalis	18
	AdhE	AAV66076	Leuconostoc mesenteroides	19
	DmpF	CAA43226	Pseudomonas sp. strain CF600	20

Table 3.1 AdhE2 homologs with C4 or C2 preference
The biochemical literature was surveyed for ALDH domains characterized to have preference for acyl-CoA substrates larger than acetyl-CoA. Both monofunctional and bifunctional ALDH domains were considered, including the atypical ALDH DmpF that is a fusion with an aldolase domain. In the case of bifunctional ALDHs only the specificity of the ALDH domain was considered.

These sequences were then placed in a phylogenetic tree such that the branching pattern was biased by their characterized substrate preference (*Figure 3.2 A)*. Next, the entire ALDH gene family[21] was assembled into a second phylogenetic tree with the branching pattern again biased by the characterized substrate specificity from the first tree (*Figure 3.2 B)*. The full family tree comprised greater than 1,200 sequences, of which approximately 33% were derived from bifunctional ALDH domains. This is not unexpected as the majority of ALDH domains characterized in the Pfam database are not found in bifunctional enzymes.

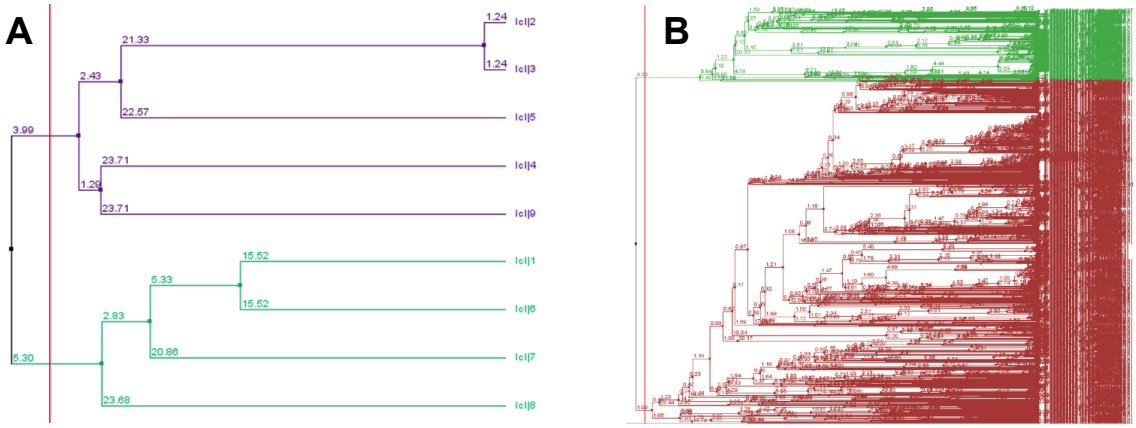


Figure 3.2 AdhE2 phylogenetic trees incorporating substrate specificity
(A) A phylogenetic tree incorporating biochemically characterized ALDH domains was assembled to identify homologs that may have greater C4 substrate specificity. Sequences in purple have higher k_{cat}/K_M for acyl-CoAs larger than acetyl-CoA, and sequences in green have higher k_{cat}/K_M for acetyl-CoA. (B) The branching pattern of the biochemically informed tree was applied to all sequences in the ALDH family. Green sequences denote ALDH domains of bifunctional enzymes and red sequences denote monofunctional ALDH domains.

From this point two approaches can be taken: the existing amino acid diversity present in the C4 specific branch can be used as a pool of mutations to be made in Adhe2[1], or wild type homologs can be directly sampled from the C4 specific branch. We began with the former, but both approaches were eventually used. Sampling mutations from existing sequences is an efficient means of generating a diverse and highly active library. This is in contrast to methods such as error-pronePCR which produce many variants with deleterious mutations or even stop codons[22]. Here all "mutations" are also the wild type sequence of a presumed functional homolog, and introducing this variant in a closely related sequence and is unlikely to generate a non-functional protein. Highly functional libraries incorporating focused diversity are essential in scenarios such as these where screening capacity is limited[23].

Ninety-six variants of AdhE2 were designed to incorporate approximately 40 mutations selected from the natural sequence diversity of AdhE2 homologs. Each variant contained 3-5 mutations and every mutation was present in multiple variants. This design ensures that each mutation is evaluated in multiple contexts so as to not discard positive mutations that were randomly paired with negative mutations. With each variant being evaluated in multiple independent contexts a linear regression can be performed to assign a contribution score to each mutation. Highly scored mutations can then be recombined at a greater frequency in successive rounds of design and screening to rapidly arrive at highly improved sequences[1]. This approach is analogous to iterative rounds of saturation mutagenesis[24] but has the considerable benefit that library sizes are kept small which enables low throughput assays to be performed. This approach has been used successfully in a number of cases.

AdhE2 variants were synthesized and cloned by Calysta, transformed into DH1 with the appropriate butanol production plasmids, and screened in 250 mL baffled flasks.

Around half of the variants produced minimal butanol (*Figure 3.3*), indicating that one or more mutations in that variant resulted in non-functional enzymes. The remaining variants produced a wide range of titers in comparison to the wild type sequence (far right). Some variants produced higher butanol titers, but upon followup production experiments exhibited only mild improvements in butanol:ethanol ratio, suggesting that overall activity was the primary improvement without dramatically shifting substrate specificity.

Despite modest results the linear regression was performed and identified 12 mutations for a second round of design and screening. Eight new variants incorporating these mutations were designed, cloned and tested but again the primary effect was a modest overall improvement in titer without considerable improvement to butanol:ethanol ratio. Post induction samples from a number of these variants were analyzed by SDS-PAGE, which revealed that many of these clones had greater levels of soluble AdhE2 expression. It is unsurprising that improved soluble expression would lead to higher titer as AdhE2 is expressed poorly and partially insolubly.

It is difficult to ascribe the failure of this approach to any one factor. A large confounding factor may be that all of the characterized bifunctional homologs other than AdhE2 used in construction of the tree were C2 specific. Although a number of AdhE2 homologs from other *Clostridia* exist[25,26], and these species are known to produce high butanol:ethanol ratios, their homologs have not been biochemically characterized and could not be included in the initial phylogenetic tree. This in turn may have compromised the breadth of information captured within the tree. Another contributing factor could be that mutations made to AdhE2 derived from C4 specific monofunctional ALDHs may be less beneficial or more disruptive than is otherwise expected when transferring mutations among more similar enzymes. Finally, it is likely that the large majority of AdhE2 homologs are simply more specific for acetyl-CoA, as will be discussed below. Given these results we next shifted to the second strategy for utilizing the phylogenetic trees: sampling of wild type homologs.

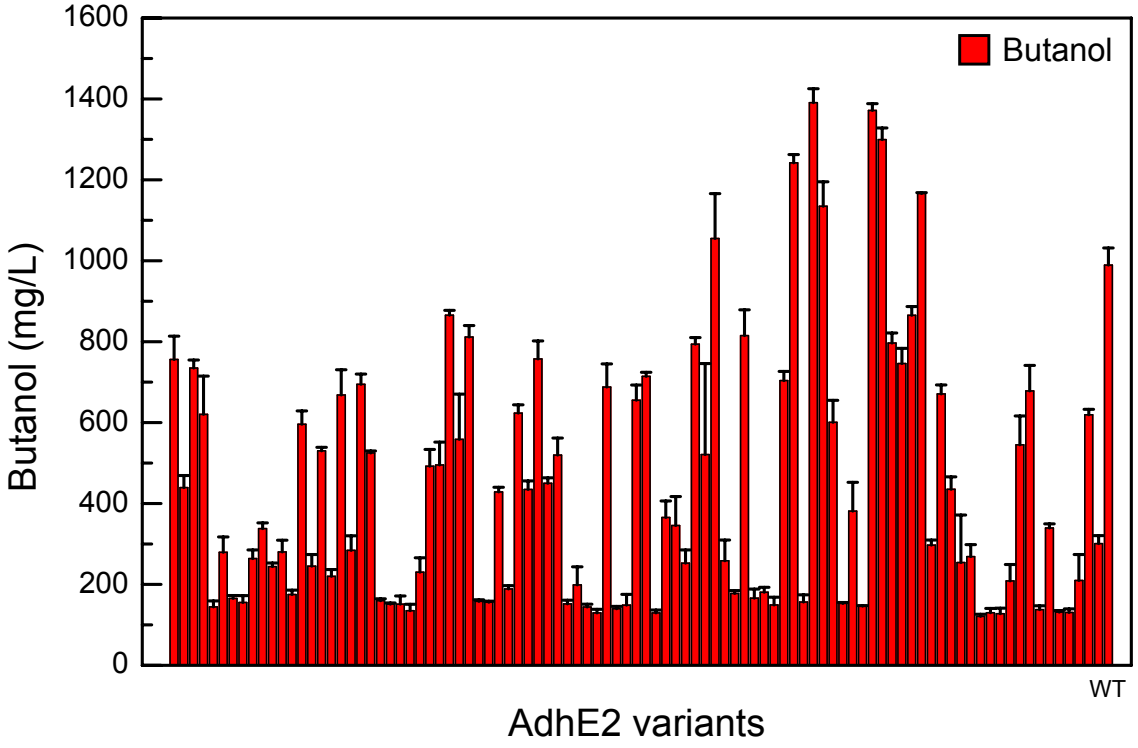


Figure 3.3 Butanol production with AdhE2 variants incorporating natural sequence diversity

Variants of AdhE2 were constructed to incorporate approximately 40 mutations selected from the natural sequence diversity of AdhE2 homologs. Each variant contained 3-5 mutations and every mutation was present in multiple variants. Around half of the variants produced minimal butanol, indicating that particular combinations of mutations resulted in non-functional enzymes. The remaining variants produced a wide range of titers in comparison to the wild type sequence (far right). Some variants produced higher butanol titers, but upon further inspection exhibited only mild improvements in butanol:ethanol ratio, suggesting the primary effect was an overall activity increase without shifting substrate specificity. Data are mean ± s.d. (n = 3).

Identification of C4 specific monofunctional aldehyde dehydrogenases

In this approach the C4 specific branch of the tree was widely sampled to incorporate the full diversity of the branch in a small number of sequences. This comprised 15 bifunctional AdhE2 homologs as well as 3 monofunctional ALDHs homologous to the ALDH domain of AdhE2. These sequences were synthesized, cloned, and screened as in the previous experiments. Here we observed that all sequences save for one produced worse butanol:ethanol ratios compared to AdhE2 (*Figure 3.4*). The lone outlier was sequence 46, a monofunctional ALDH from *Clostridium beijerinckii NCIMB 8052*. Amusingly, this sequence (aldh46 hereafter) differed from the ALDH shown to be inactive in *Figure 2.11* by only 8 out of 450 residues.

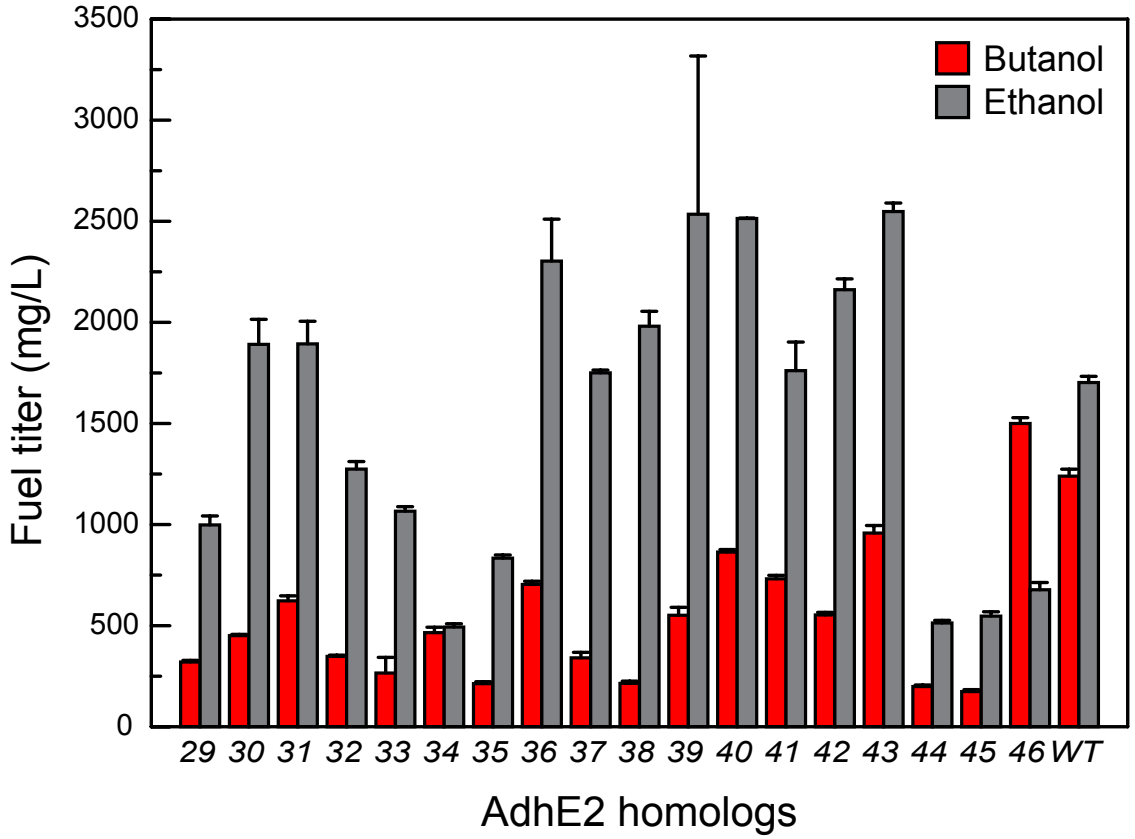


Figure 3.4 Identification of a C4 specific monofunctional aldehyde dehydrogenase
Additional wild type AdhE2 homologs were screened, but as with AdhE2 (far right) all were found to have poor butanol:ethanol ratios. However, one sequence (46) was observed to produce more butanol than AdhE2 with considerably less ethanol. Intriguingly, this sequence is a monofunctional aldehyde dehydrogenase. Data are mean ± s.d. (n = 3).

We next examined whether monofunctional ALDHs might be a more fruitful source of C4 specific enzymes. We again returned to the phylogenetic tree and widely sampled exclusively monofunctional ALDHs which were synthesized, cloned, and screened as before. We were delighted to find that 15 of 16 monofunctional ALDHs produced more butanol than ethanol, and one sequence appeared to be inactive (*Figure 3.5*).

At this point we had identified a class of monofunctional C4 specific ALDHs, but several new questions arose. Some monofunctional ALDH production experiments at this point included the ADH domain of AdhE2 (as an artifact of the cloning method used to make these plasmids) to catalyze the final reduction of butyraldehyde to butanol, but other experiments did not include this ADH yet still produced reasonable titers. We reasoned that a native *E. coli* broad specificity alcohol dehydrogenase was completing the pathway, so we next moved to identify this ADH as well as optimize the expression of new monofunctional ALDH.ADH based pathways.

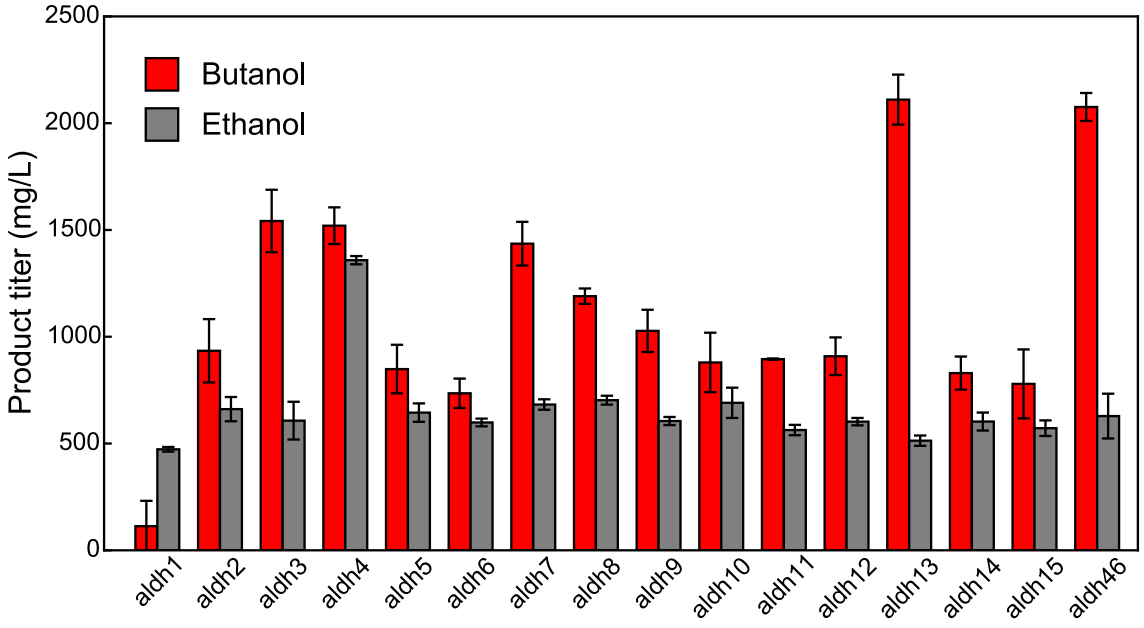


Figure 3.5 Identification of a family of C4 specific monofunctional ALDHs
To further assess the ability of monofunctional aldehyde dehydrogenases to support robust and specific butanol production additional homologs of aldh46 were screened. The majority of homologs displayed at least modest specificity for butanol production, and several homologs were seen to be highly specific. Data are mean ± s.d. (n = 3).

To search for the native *E. coli* ADH complementing the ALDH based pathways, we transformed our production plasmids into strains from the Keio collection[27] containing knockouts of annotated ADHs (*Figure 3.6*). Six strains were selected including 5 monofunctional ADHs as well as the *E. coli* homolog of AdhE2. The anticipated result was that all cultures would make butanol except for one or several cultures that would have diminished butanol titer due to the complementing ADH being knocked out in that strain. Surprisingly we observed the exact opposite; all cultures produced almost exclusively ethanol except for one, ΔadhE, which produced the expected butanol titer and minimal ethanol. Upon further consideration, combined with the fact that adhE is the major source of ethanol production in *E. coli*[28], the most likely explanation is that any strain expressing adhE would outcompete the unoptimized butanol pathway for acetyl-CoA availability, thus diverting all available substrate to ethanol production. Only in the ΔadhE strain is this competition relieved such that expected butanol production is restored. To identify which *E. coli* ADH is complementing butanol production it would be necessary to make the adhE knockout in each of the other strains. This experiment was not performed as the ALDH based pathway was successfully optimized using the ADH domain of AdhE2 and identifying permissive ADHs to optimize production was no longer a priority. After this work was completed a butanol pathway utilizing the *E. coli* ADH yqhD was published[29]. YqhD was previously known to be highly expressed and relatively promiscuous in substrate selection and is likely to have contributed to butanol production in our strains[30-32].

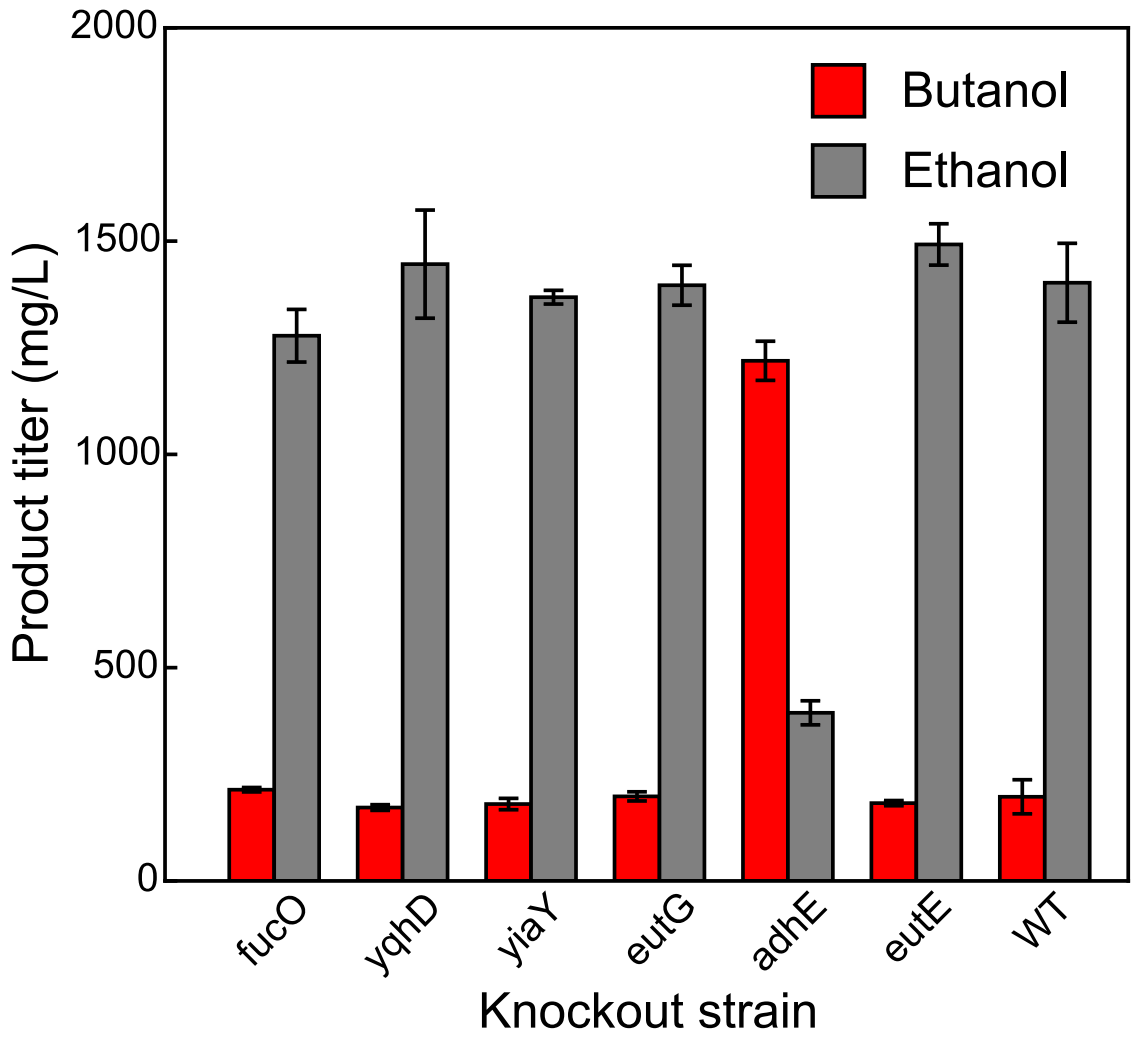


Figure 3.6 Native E. coli alcohol dehydrogenases complement monofunctional ALDH pathways

Multiple knockout strains were examined to identify which native E. coli alcohol dehydrogenase may contribute to butanol production in pathways employing monofunctional ALDHs. Surprisingly, all strains except ΔadhE (the E. coli homolog of AdhE2) produced large amounts of ethanol and very little butanol. The E. coli AdhE is specific for ethanol production and is the major source of ethanol in fermentation of wild type E. coli. Any strain containing AdhE would efficiently consume acetyl-CoA and severely limit flux to butanol, thus masking any losses in butanol production caused by the deletion of an alcohol dehydrogenase that can complement a monofunctional ALDH pathway. Deletion of adhE removes this acetyl-CoA consumption and restores the expected butanol production phenotype. Double knockouts of ADHs in the ΔadhE background are required to identify native ADHs contributing to butanol production. Data are mean ± s.d. (n = 3).

With a family of C4 specific monofunctional ALDHs identified the next goal was to optimize the performance of this pathway equal to or above AdhE2 production (*Figure 3.7*). To a pathway including only aldh46 we supplemented the ADH domain of AdhE2, which more than doubled titer with zero increase to ethanol production. Butanol production was then improved beyond AdhE2 production by optimizing the expression of aldh46.adh under a stronger Trc promoter.

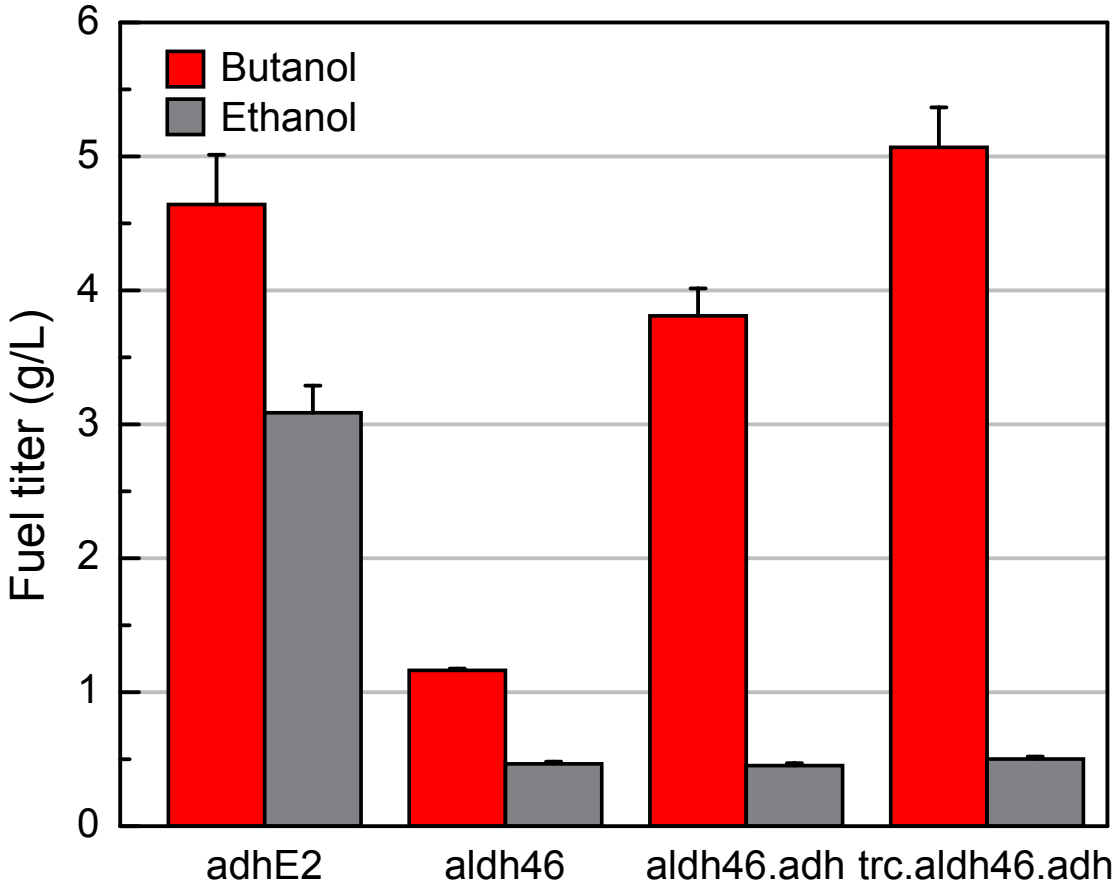


Figure 3.7 Optimization of a monofunctional ALDH/ADH butanol production pathway
Although replacement of AdhE2 with the C4 specific monofunctional ALDH46 does result in improved butanol:ethanol ratios, total butanol titer is substantially lower. Expression of a monofunctional ADH (the ADH domain of AdhE2) restores the majority of butanol titer, and increasing expression levels of both enzymes surpasses the butanol production of AdhE2 without any additional ethanol production. Data are mean ± s.d. (n = 3).

Development of a genetic selection for butanol production

With a highly specific high titer butanol production pathway now developed, we moved on to the longer-term goal of developing a genetic selection for improved butanol production. We initially explored using a butanol responsive transcription factor, BmoR from *Thauera butanivorans*[33], to drive expression of a reporter gene in response to

increasing butanol concentrations. Similar transcription factor based approaches have been successful in some scenarios[34-36], however in our hands this approach proved too inconsistent and had very narrow dynamic range which drastically limited utility. Instead we implemented an anaerobic growth selection in a fermentation pathway knockout strain.

This knockout strain, DH1 ΔackA-pta ΔadhE ΔldhA ΔpoxB ΔfrdBC (hereafter MC1.24), has every major *E. coli* fermentation pathway removed. This results in a strain that grows acceptably under aerobic conditions, but does not grow under anaerobic conditions. This is because under anaerobic growth oxidative phosphorylation with oxygen as a terminal electron acceptor is not possible. This severely limits ATP yield to 2 ATP per glucose derived during glycolysis. Glycolysis also generates 2 NADH per glucose, which must be recycled to allow glycolysis and ATP production to continue. In wild type *E. coli* this role is filled by a number of fermentation pathways (*Figure 3.8 A*), primarily those producing acetate, lactate, and ethanol[37]. With all of these pathways removed in MC1.24 there is no NADH recycling capacity and glycolysis cannot continue, terminating ATP production in the process.

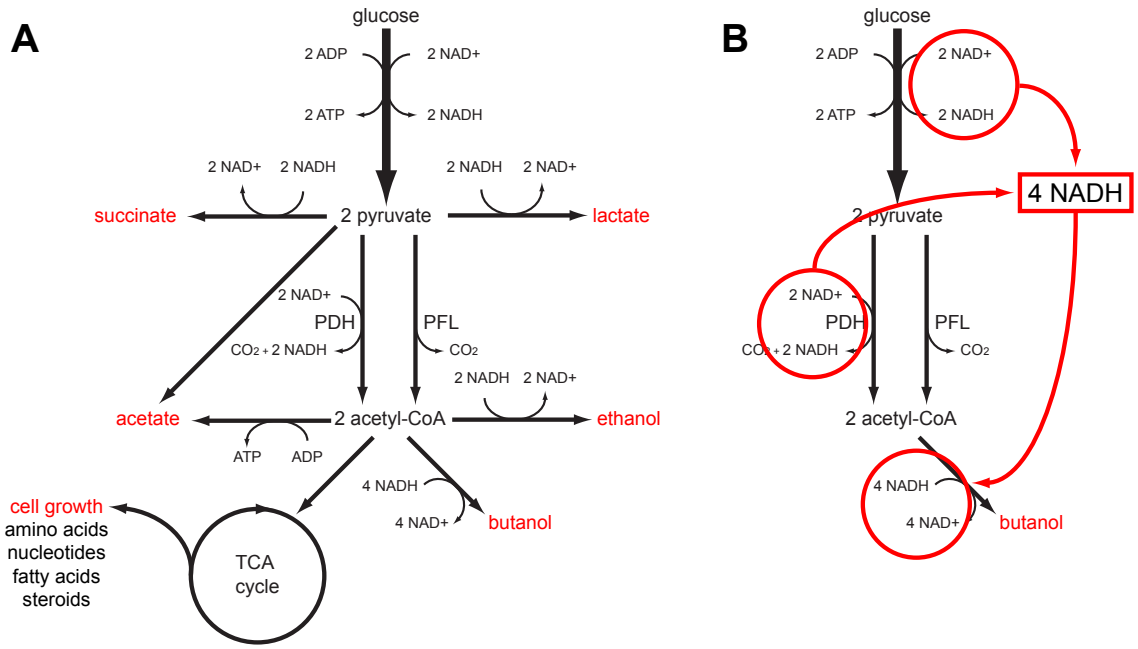


Figure 3.8 Redox balance and ATP production in wild type and fermentation knockout strains

The growth condition and fermentation pathways available to a cell have an enormous impact on the ATP yield, growth rate, and carbon fate decisions of a cell. (A) Wild type E. coli has multiple fermentation pathways available to meet redox requirements and generate ATP. (B) All major native fermentation pathways have been deleted from a quintuple knockout strain (MC1.24) leaving butanol production as the only fermentation pathway available to maintain redox balance and allow glycolytic ATP production to continue.

However, a strain carrying a synthetic fermentation pathway, producing butanol or another metabolite, can complement the fermentation knockouts by restoring NADH recycling and ultimately ATP generation and growth. Our butanol production pathway recycles 4 NADH per glucose[38] and is thus appropriately balanced with the 2 NADH per glucose produced by glycolysis in addition to the 2 NADH produced by PDHc during conversion of pyruvate to acetyl-CoA (*Figure 3.8 B*).

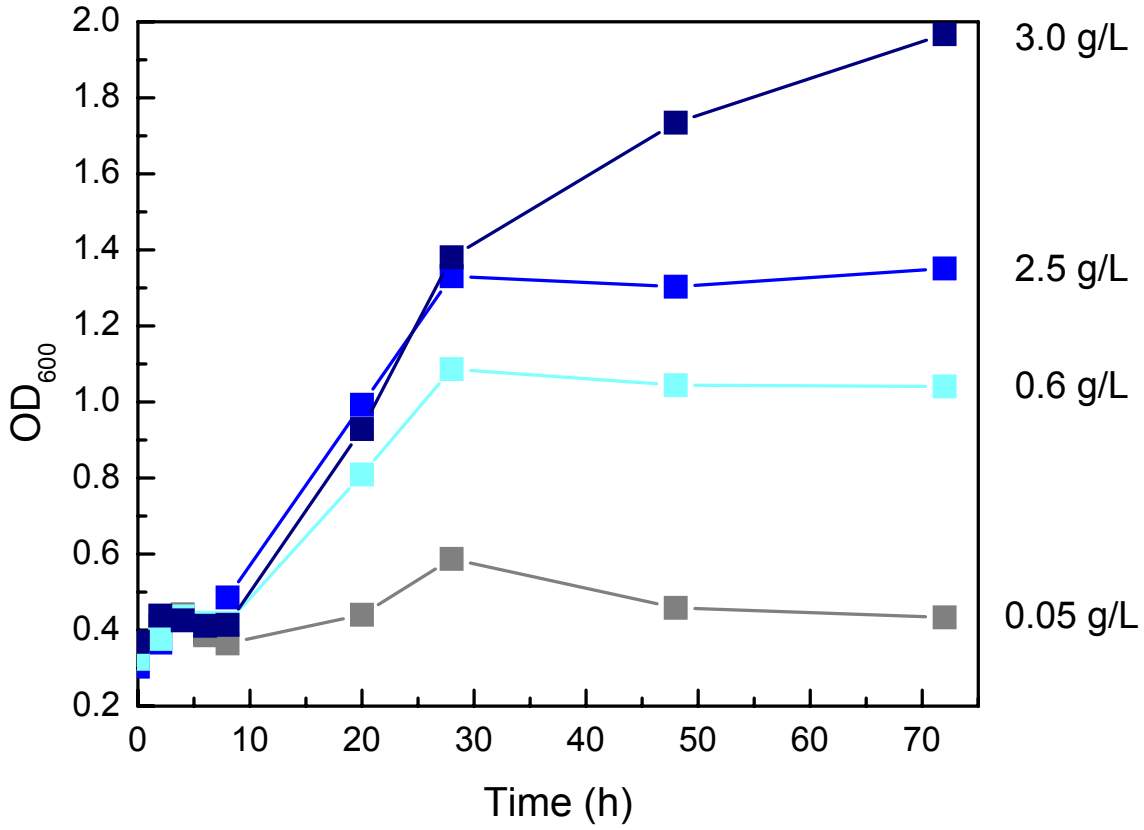


Figure 3.9 Butanol production rescues growth in fermentation deficient cells
A fermentation-compromised strain (MC1.24) can be rescued by expression of a functional butanol production pathway. Strains expressing poor performing butanol pathways (butanol titer indicated to the right) grow minimally under anaerobic conditions. In contrast, strains expressing high performing butanol pathways grow robustly. Anaerobic growth is tightly linked to butanol production.

The extent to which a synthetic fermentation pathway is able to rescue anaerobic growth is dependent upon the pathway's capacity to recycle NADH. In MC1.24 complemented with butanol production pathways capable of a range of titers, growth is highly correlated with pathway titer (*Figure 3.9*). Strains complemented with a very low titer pathway do not grow significantly, if at all, while strains complemented with robust pathways grow to high OD_{600}.

To explore the feasibility of this anaerobic growth complementation for the purpose of a selection, we mixed cocultures comprised of 1% or 0.1% medium production strains with 99% or 99.9% low production strains. Throughout the course of extended anaerobic growth we observed a significant lag phase as only a miniscule fraction of the population was able to grow at an appreciable rate (*Figure 3.10*). Over time however this small fraction was enriched and eventually grew to dominate the culture and produce significant butanol.

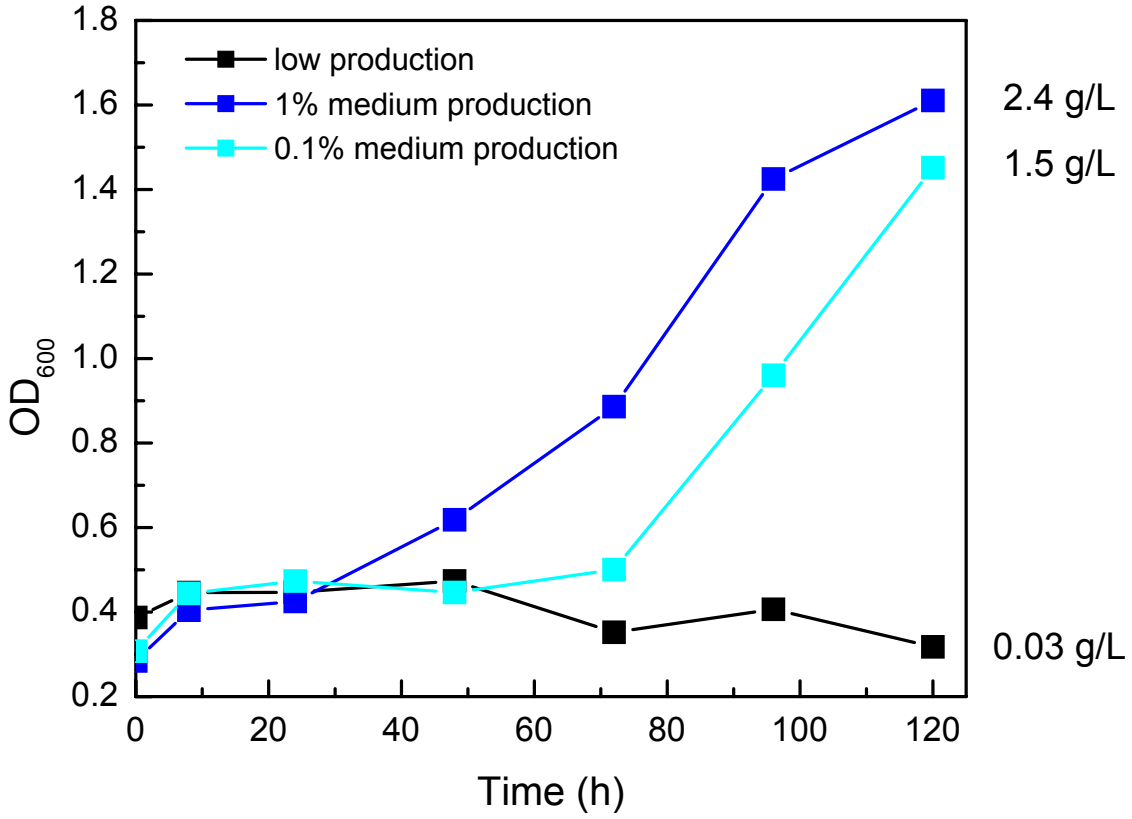


Figure 3.10 Anaerobic growth can enrich for high production strains
A small fraction of medium production strains were mixed in a large excess of low production strains to simulate a mutagenized library in which most mutations are neutral or deleterious. Through extended culturing under anaerobic growth conditions, the small fraction of medium production cells can proliferate and dominate the culture. A lag in culture growth correlates with the initial abundance of medium production cells. Over the course of five days the culture attained high OD_{600} and butanol tither.

This experiment emulated enrichment of a diverse mutant population where a large majority of the cells contain neutral or deleterious mutations and thus have no growth advantage, but a small fraction of the population carries beneficial mutations that confer increased butanol production and therefore growth. This result encouraged us that our selection scheme would be effective for identifying high production mutants following mutagenesis.

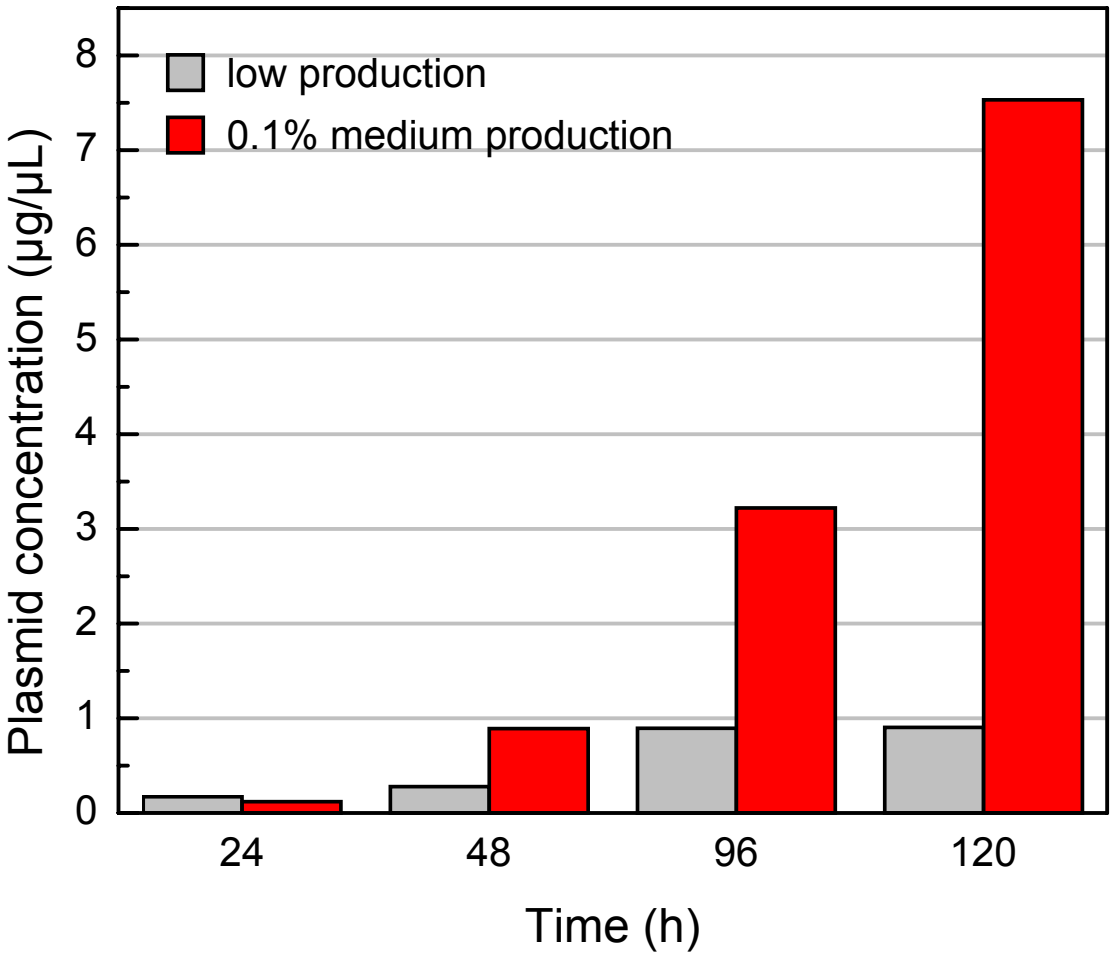


Figure 3.11 Abundance of culture subpopulations can be tracked during enrichment
The plasmid abundance of low and medium production subpopulations was monitored throughout the course of enrichment. Low and medium production strains carried distinct plasmids and the abundance of each was monitored by qPCR. The low production population remained relatively static over time due to severely compromised growth, while the medium production population expanded over 40-fold over the course of five days.

As a further validation of the simulated selection presented in *Figure 3.10*, we also tracked the abundance of the two subpopulations throughout the selection. The low and medium production strains used differed in the makeup of one of the butanol production plasmids. This difference allowed us to specifically quantify the abundance of each population through time by qPCR. In agreement with the growth curves, the abundance of the low production strain was largely static through the enrichment, but the abundance of the medium production strain was multiplied greater than 40-fold over the course of the experiment (*Figure 3.11*). This further confirmed our selection strategy was sound.

Whole-genome mutagenesis to improve butanol production

Although our selection strategy is generalizable to any mutagenesis method or target, we first sought to mutate the *E. coli* host genome as this gives us an opportunity to explore unknown regulatory mechanisms influencing the performance of our pathway[6]. MC1.24 transformed with our top butanol production plasmids was subjected to a moderate level of the mutagenic alkylating agent ethyl methanesulfonate (EMS) expected to result in 50-60 mutations per cell[39]. Although the mutational profile of EMS is limited in scope (generating almost exclusively G/C to A/T transitions) relative to mutagenesis methods that generate much larger phenotypes such as transposons, it is very straightforward to use and has a proven track record[8]. Additional mutagenesis methods including transposons and UV irradiation have also been explored but will not be detailed here.

Following mutagenesis of triplicate cultures (as well as triplicate mock mutagenesis controls), cultures were grown in MOPS M9 minimal media supplemented with 2.5% glucose, 1 mM IPTG, and appropriate antibiotics. During early rounds of culture growth the media was also supplemented with 10% LB, as DH1 based strains grow poorly in minimal media. As the cultures grew they were repeatedly serially transferred and the LB supplement was tapered to 1%. After only three dilutions EMS treated cultures had surpassed untreated cultures in both growth (*Figure 3.12*) and butanol production (*Figure 3.13*). That the data look very similar whether plotting growth or titer further speaks to the tight linkage between production and growth in this strain.

At the conclusion of the selection glycerol stocks were made and the cultures were streaked onto MOPS M9 agar plates. Clones were picked and recultured to asses the diversity of phenotypes remaining in the population. The top clone from the initial selection was subjected to a second round of mutagenesis to drive further improvements. This selection was carried out identically except that the LB supplement was eliminated to further increase the pressure on glycolytic flux.

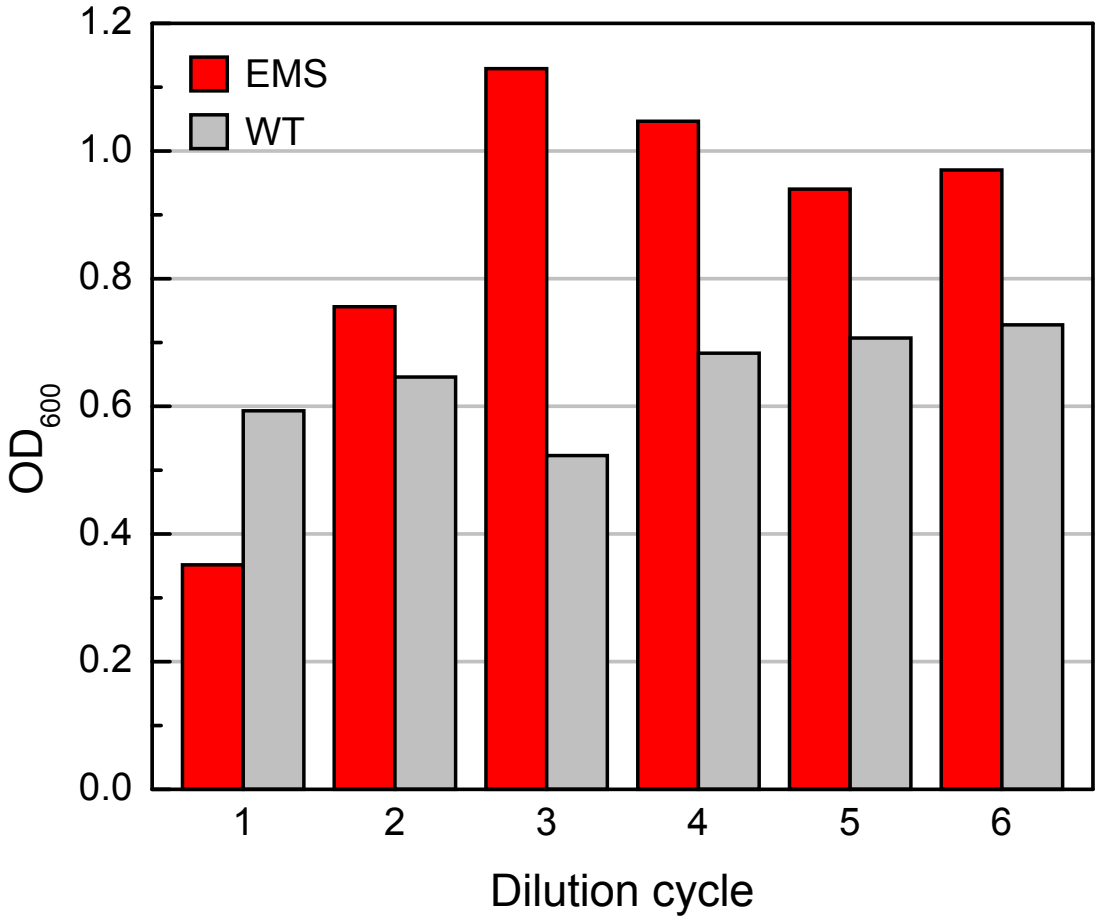


Figure 3.12 Growth improvement of EMS treated cells during anaerobic selection
MC1.24 carrying the optimized aldh46 butanol pathway was mutagenized with EMS and serially transferred in minimal media under anaerobic conditions to select for higher performing strains. At initial dilutions EMS treated cultures had a growth defect as the majority of the population carried an excess of deleterious mutations. At later time points growth improved above untreated cultures as these populations were enriched for strains with a growth advantaged conferred by higher butanol production. Data are representative cultures from biological triplicates.

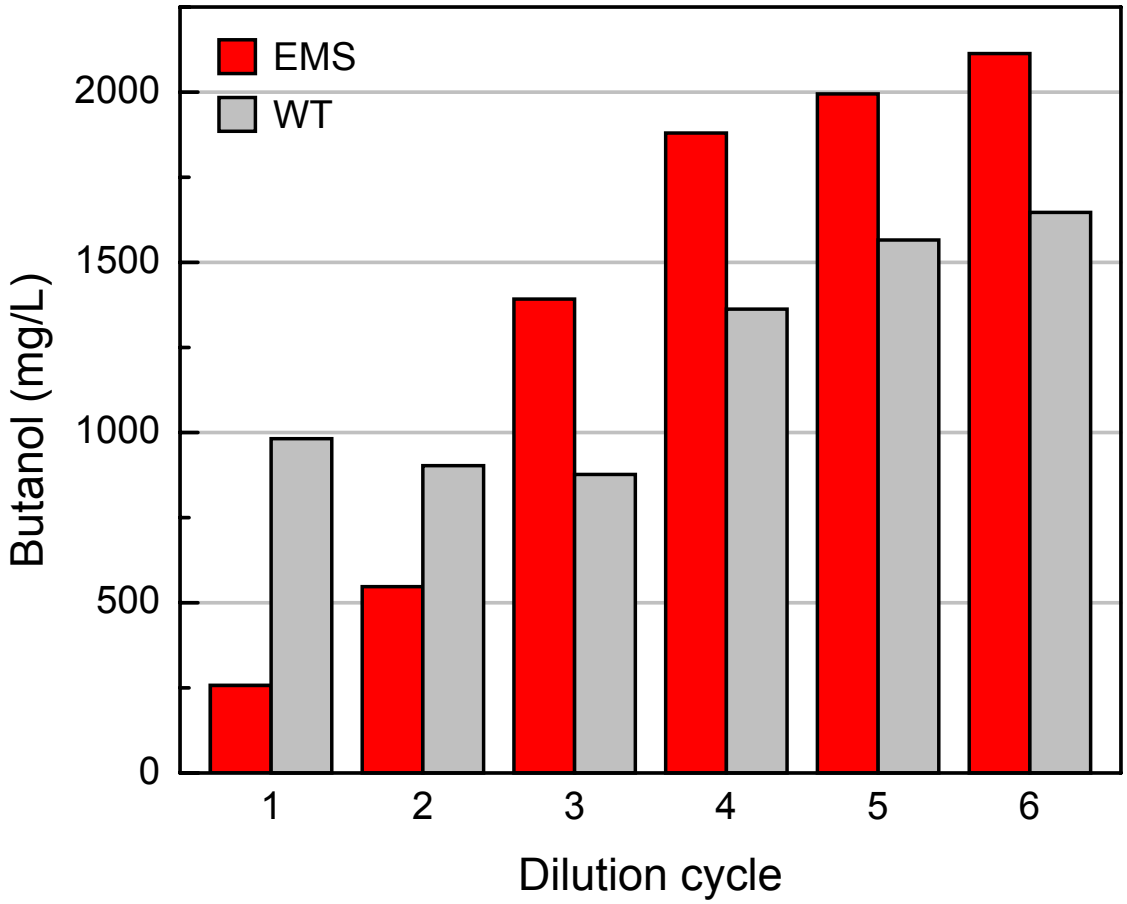


Figure 3.13 Butanol titer improvement of EMS treated cells during anaerobic selection
MC1.24 carrying the optimized aldh46 butanol pathway was mutagenized with EMS and serially transferred in minimal media under anaerobic conditions to select for higher performing strains. As in Figure 3.12, during initial dilutions EMS treated cultures had a production defect as the small fraction of the population with improved phenotypes had not yet become enriched within the culture. At later time points mutagenized cultures again outperformed untreated cultures. The similarity of this trajectory as compared to the trajectory of growth improvements over time highlights the close linkage of growth and butanol production in MC1.24 under anaerobic conditions. Data are representative cultures from biological triplicates.

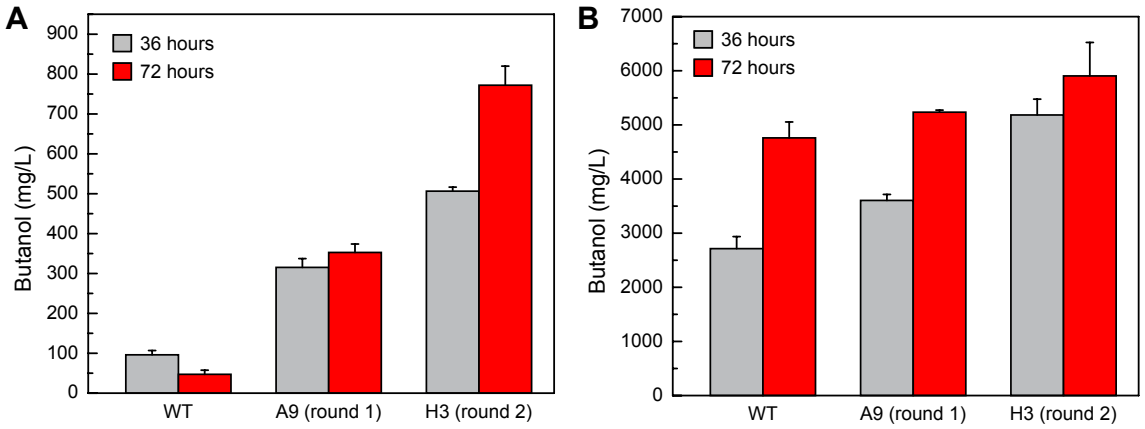


Figure 3.14 Evolved strains have higher titer and productivity
EMS mutagenesis followed by anaerobic growth selection was carried out for two rounds, and followed by clone picking and verification of production at the end of each round of selection. Evolved strains display overall higher titer in both minimal (A) and rich media (B) as well as substantially higher productivity per time in rich media. Data are mean ± s.d. (n = 3).

Following the second round of selection clones were again picked and characterized to confirm improved phenotypes. In all cases the round 2 strains demonstrated the highest titer and fastest productivity (*Figure 3.14*). It is possible, and not unexpected, that mutant strains could pass the selection by means other than improving butanol production such as activating cryptic fermentation pathways or rerouting flux to previously low flux pathways[40,41]. Although this may be occurring to some extent we did not observe large increases in fermentation products other than butanol, and all strains that were screened produced more butanol than the parent strain.

After the second selection clones from each round were sequenced to identify the mutations acquired during mutagenesis. Illumina libraries were prepared and MiSeq 250 base paired-end reads were generated at the UC Davis Genome Center (Davis, CA). Read data was compared to the reference genome using the Breseq pipeline[12] which confirmed the expected genomic distribution (*Figure 3.15*) and mutational profile (*Table 3.16*). Mutations were classified by cellular function and fell into a broad range of categories. As it is expected that only a fraction of the accumulated mutations contribute to the butanol production phenotype, extensive characterization and strain construction are required to confidently assign a causal relationship for a given mutation. This is especially true of mutations influencing phenotype in unusual or previously unknown ways.

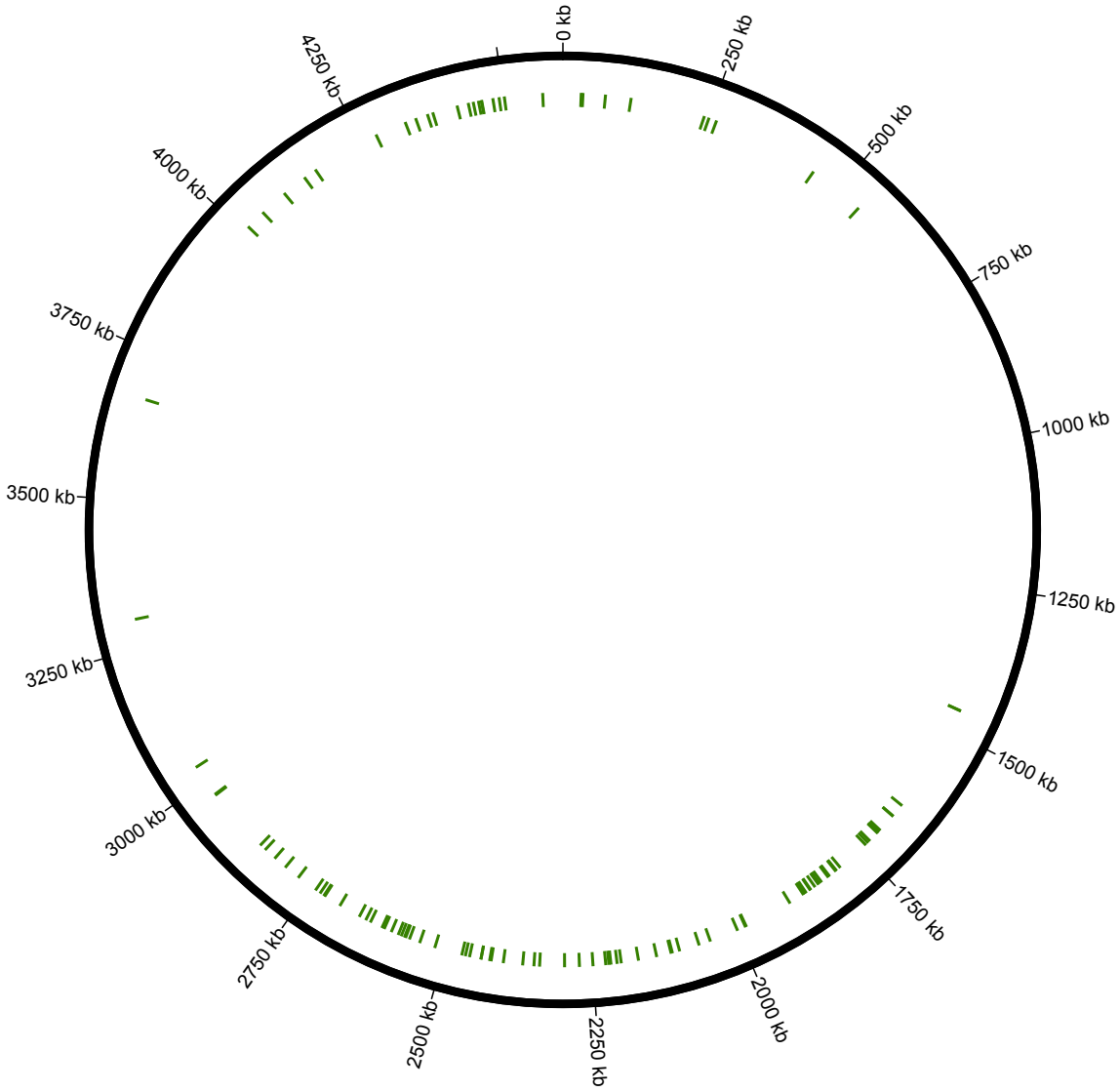


Figure 3.15 Genomic location of mutations in the round two evolved strain
A clone was picked following the second round of selection, confirmed to have higher production relative to its parent strain, and its genome was sequenced. After comparison to the reference genome 107 mutations were identified. Mutations were almost exclusively G/C to A/T transitions as expected with EMS mutagenesis, and mutations were distributed mostly randomly throughout the genome with some local heterogeneity as has been noted in previous studies. Figure generated with Circos[42].

Total mutations		Mutation type		Cellular function	
Round 1	62	*G→A*	47	*Metabolism*	29
Round 2	107	*C→T*	56	*Membrane*	20
		A→G	1	*Housekeeping*	17
		G→T	1	*Regulatory*	13
		Insertion	1	*Intergenic*	9
		Deletion	1	*Other*	9
				Amino acid	6
				Unknown	6

Table 3.16 Mutations discovered in evolved strains
Clonal strains from both rounds of selection were sequenced and compared to the reference genome. Each round produced 50-60 mutations as expected given the EMS dose, and mutations were almost exclusively G/C to A/T transitions. Mutations were classified by cellular function and fell into a broad range of categories.

However if we limit our analysis to only mutations with plausible or previously known mechanisms we can still observe a wide variety of mechanisms in play. Coding sequence mutations in native enzymes with catalytic functions similar to the acetoacetyl-CoA reductase (HBD) and the crotonase (crt) used in the butanol pathway could change the activity of these enzymes such that they now contribute to butanol production[43] (*Figure 3.17 A*). A premature stop codon inserted in the malate dehydrogenase dmlA eliminates activity that could be syphoning pyruvate away from acetyl-CoA[44] (*Figure 3.17 B*). Mutation of a surface exposed residue on the binding interface of the anti-repressor mtfA could increase its sequestration of the global transcription factor mlc (*Figure 3.17 C*). Mlc would no longer repress expression of ptsG[45,46], a subunit of the glucose uptake machinery, therefore increasing glucose uptake. A premature stop codon in the dispensable C-terminus of the essential gene RNaseE could slow its activity in mRNA degradation (*Figure 3.17 D*). Similar mutations have already been shown to limit degradation of ptsG mRNA, again leading to higher glucose uptake[47,48]. Decreased mRNA degradation could also increase expression of heterologous butanol pathway genes.

These examples illustrate the power of large libraries and efficient selections: these specific mutations are unlikely to have been rationally selected in a forward-engineering approach, and yet the law of large numbers and the efficiency of evolution will always bring effective solution to the fore.

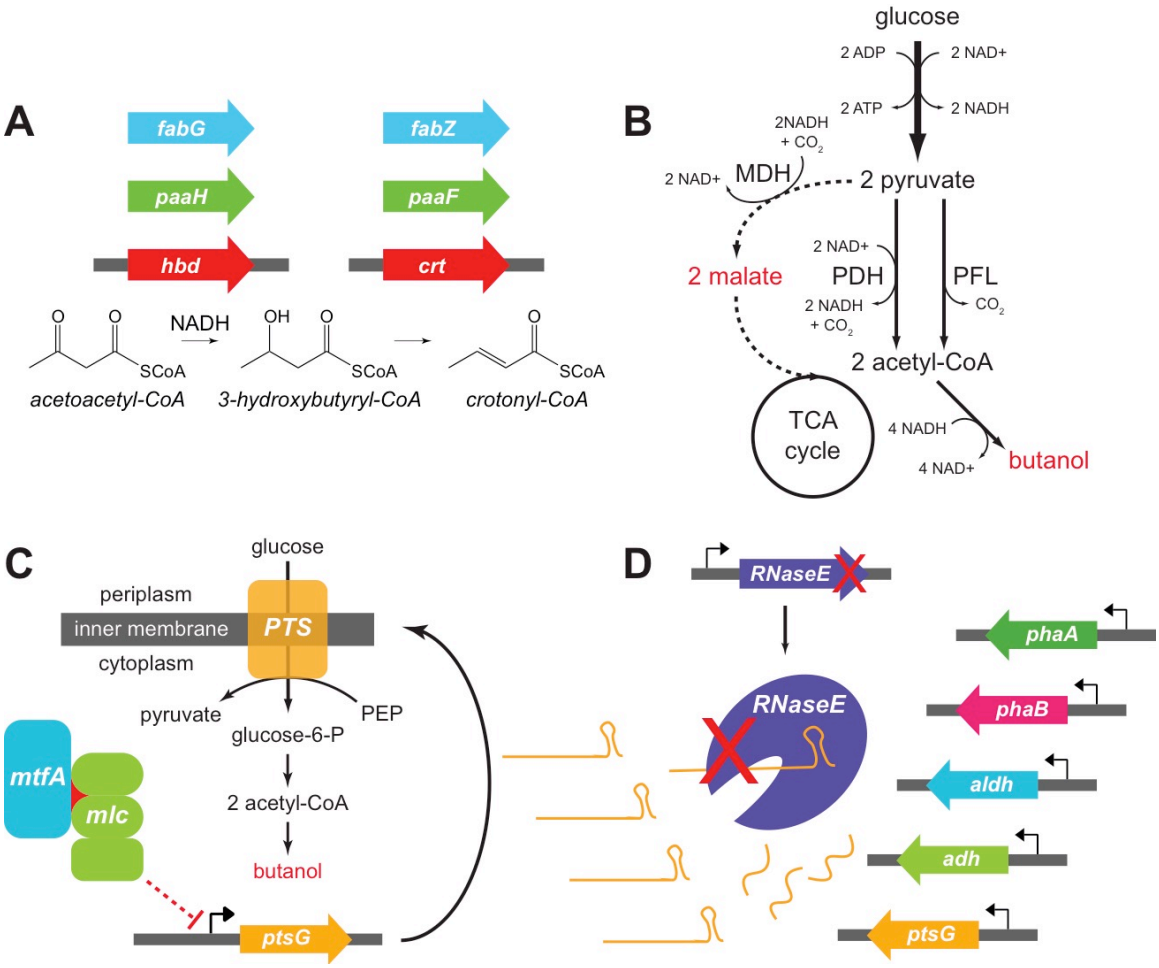


Figure 3.17 Regulatory mechanisms impacting butanol production
(A) Mutations within the coding sequence of metabolic enzymes that carry out similar functions to butanol pathway enzymes may alter substrate specificity to allow a native enzyme to contribute to flux. (B) A stop codon inserted in the malate dehydrogenase dmlA leads to a truncated and inactive protein that can no longer siphon away pyruvate flux to acetyl-CoA. (C) A mutation in the C-terminal binding interface of mtfA may cause it to sequester global transcription factor mlc away from the ptsG operator site, allowing greater expression of ptsG and higher glucose uptake. (D) A stop codon inserted in the C-terminus of the essential RNaseE its mRNA degradation activity, allowing higher expression levels of known targets including ptsG and general targets such as the highly expressed butanol pathway genes.

3.4 Conclusions

Here we have used bioinformatics approaches to identify and screen ALDHs (both bifunctional and monofunctional) with a preference for C4 substrates. This approach utilized the tremendous amount of sequence information available to evaluate more diverse groups of sequences, thus improving the chance of identifying a desirable sequence.

Although an initial strategy of using these diverse sequences as a mutational pool for AdhE2 was unsuccessful, the complimentary strategy of sampling homologs proved very

successful, although not quite in the manner expected. We were unable to identify any bifunctional aldehyde alcohol dehydrogenases that produced more butanol than ethanol, but we were able to find a large set of monofunctional aldehyde dehydrogenases that proved quite specific for butanol production. Pathways based on these monofunctional ALDHs were easily optimized by inclusion of the ADH domain of AdhE2, followed by expression level tuning. The high titer and specificity of this improved pathway opened the possibility to our long-term goal of developing a selection for butanol production.

The selection hinges on the need for cells to maintain redox balance when growing anaerobically. An engineered strain that cannot maintain this balance on its own is unable to grow unless it is complemented with a synthetic fermentation pathway for butanol production. With butanol production tightly linked to growth we were able to mutate the genome of *E. coli* and rapidly select for mutants with improved production. This selection was successful in two rounds of mutagenesis and lead to an evolved strain with 107 mutations.

Further study will be required to determine which mutations are causative and which are merely hitchhikers. Several mutations with easily rationalizable mechanisms were presented, but focusing only on the most easily rationalized mutations will always run the risk of overlooking unusual or completely novel mechanisms. A number of fruitful approaches exist for continuing this work. First and foremost should be following up on the identified mutations and beginning to asses their contribution to the overall phenotype. Until very recently the process of making each mutation in a clean background strain would have been quite daunting[10,49]. However today it appears that the latest cas9 genome editing techniques could make the thought of generating over 100 point mutant strains considerably more palatable[50]. However the feasibility of such an approach should be tempered by the potential reward (or lack thereof) on the time invested. If a small minority of the mutations (20% or fewer seems a reasonable guess) are responsible for the majority of the phenotype, then a considerable amount of time will have been spent reconstructing hitchhiker mutations. Also there is the possibility that the effect of some mutations may only be seen in concert with others; these mutations would be lost in a sequential search.

Several techniques have been employed for mapping genotype:phenotype linkages in scenarios precisely like this. These techniques have been shown to effectively identify causative mutations from diverse backgrounds[51-54], but the approaches are not straightforward and would require considerable effort and fine-tuning.

Aside from mapping genotype to phenotype, these evolved strains are likely to benefit from the application of additional “omics” techniques. The genomes of these strains have been sequenced, but the barrier to entry in collecting other large datasets continues to decrease over time. RNAseq of *E. coli* is relatively routine[55], and powerful proteomics tools are being developed rapidly[56,57]. A very well developed toolkit specifically for measuring protein abundance of nearly all *E. coli* metabolic enzymes has been presented[58] and the instrumentation required is readily available.

In contrast to pouring over mutants that have already been developed, it is likely to be just as profitable to evolve additional mutant strains. EMS mutagenesis was employed due to its simplicity, and transposon mutagenesis and UV irradiation should be revisited.

Transposon mutagenesis in particular is promising as each mutation is likely to have a large phenotype compared to the typical EMS point mutation. Mutations would be much easier to validate and every endpoint strain would generally contain a single mutation.

In the medium term there may also be utility in developing additional mutagenesis techniques. Two methods that have been explored are cas9 combinatorial knockins/knockouts[50] and P1-phage mediated genome shuffling[59]. Both techniques offer considerable power but are likely to require equally considerable development effort.

Regardless of which mutagenesis methods are employed or how mutation validation is prioritized, the selection developed here will continue to be useful in addressing the goal of understanding complex biological networks so that they may be more easily re-engineered with a purpose.

3.5 References

1. J. Liao, M. K. Warmuth, S. Govindarajan, J. E. Ness, R. P. Wang, C. Gustafsson and J. Minshull, Engineering proteinase K using machine learning and synthetic genes, *BMC Biotechnol.* **2007**, *7*, 16.
2. N. J. Turner, Directed evolution drives the next generation of biocatalysts, *Nat. Methods* **2009**, *5*, 567–573.
3. R. J. Fox, S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam, S. Muley, *et al.*, Improving catalytic function by ProSAR-driven enzyme evolution, *Nat. Biotechnol.* **2007**, *25*, 338–344.
4. J. A. Dietrich, A. E. McKee and J. D. Keasling, High-Throughput Metabolic Engineering: Advances in Small-Molecule Screening and Selection, *Annu. Rev. Biochem.* **2010**, *79*, 563–590.
5. H. S. Girgis, Y. Liu, W. S. Ryu and S. Tavazoie, A Comprehensive Genetic Characterization of Bacterial Motility, *PLoS Genet.* **2007**, *3*, e154.
6. H. Goodarzi, B. D. Bennett, S. Amini, M. L. Reaves, A. K. Hottes, J. D. Rabinowitz and S. Tavazoie, Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in E. coli, *Mol. Syst. Biol.* **2010**, *6*, 378.
7. I. D. Normansell, Strain improvement in antibiotic-producing microorganisms, *J. Chem. Tech. Biotechnol.* **2007**, *32*, 296–303.
8. R. T. Rowlands, Industrial strain improvement: Mutagenesis and random screening procedures, *Enzyme Microb. Technol.* **1984**, *6*, 3–10.
9. S. Parekh, V. A. Vinci and R. J. Strobel, Improvement of microbial strains and fermentation processes, *Appl. Microbiol. Biotechnol.* **2000**, *54*, 287–301.
10. H. Wang, F. Isaacs, P. Carr, Z. Sun, G. Xu, C. Forest and G. Church, Programming cells by multiplex genome engineering and accelerated evolution, *Nature* **2009**, *460*, 894–898.

11. N. R. Sandoval, J. Y. H. Kim, T. Y. Glebes, P. J. Reeder, H. R. Aucoin, J. R. Warner and R. T. Gill, Strategy for directing combinatorial genome engineering in Escherichia coli, *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 10540–10545.

12. D. E. Deatherage and J. E. Barrick, Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq, *Methods Mol. Biol.* **2014**, *1151*, 165–188.

13. J. Toth, A. A. Ismaiel and J.-S. Chen, The ald Gene, Encoding a Coenzyme A-Acylating Aldehyde Dehydrogenase, Distinguishes Clostridium beijerinckii and Two Other Solvent-Producing Clostridia fromClostridium acetobutylicum, *Appl. Env. Microbiol.* **1999**, *65*, 4973–4980.

14. R. T. Yan and J. S. Chen, Coenzyme A-acylating aldehyde dehydrogenase from Clostridium beijerinckii NRRL B592, *Appl. Env. Microbiol.* **1990**, *56*, 2591–2599.

15. L. H. Luo, J.-W. Seo, J.-O. Baek, B.-R. Oh, S.-Y. Heo, W.-K. Hong, D.-H. Kim and C. H. Kim, Identification and characterization of the propanediol utilization protein PduP of Lactobacillus reuteri for 3-hydroxypropionic acid production from glycerol, *Appl. Microbiol. Biotechnol.* **2010**, *89*, 697–703.

16. N. Hosoi, C. Ozaki, Y. Kitamoto and Y. Ichikawa, Purification and properties of aldehyde dehydrogenase (acylating) from propionibacterium freudenreichii, *J Ferment. Technol.* **1979**, *57*, 418–427.

17. S. Atsumi, A. F. Cann, M. R. Connor, C. R. Shen, K. M. Smith, M. P. Brynildsen, K. J. Y. Chou, T. Hanai and J. C. Liao, Metabolic engineering of Escherichia coli for 1-butanol production, *Metab. Eng.* **2008**, *10*, 305–311.

18. L. B. Sánchez, Aldehyde Dehydrogenase (CoA-Acetylating) and the Mechanism of Ethanol Formation in the Amitochondriate Protist,Giardia lamblia, *Arch. Biochem. Biophys.* **1998**, *354*, 57–64.

19. O. K. Koo, D.-W. Jeong, J. M. Lee, M. J. Kim, J.-H. Lee, H. C. Chang, J. H. Kim and H. J. Lee, Cloning and characterization of the bifunctional alcohol/acetaldehyde dehydrogenase gene (adhE) in Leuconostoc mesenteroides isolated from kimchi, *Biotechnol. Lett.* **2005**, *27*, 505–510.

20. J. Powlowski, L. Sahlman and V. Shingler, Purification and properties of the physically associated meta-cleavage pathway enzymes 4-hydroxy-2-ketovalerate aldolase and aldehyde dehydrogenase (acylating) from Pseudomonas sp. strain CF600, *J. Bacteriol.* **1993**, *175*, 377–385.

21. R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, *et al.*, Pfam: the protein families database, *Nucleic Acids Res.* **2014**, *42*, D222–30.

22. M. Camps, J. Naukkarinen, B. P. Johnson and L. A. Loeb, Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9727–9732.

23. R. D. Gupta and D. S. Tawfik, Directed enzyme evolution via small and effective neutral drift libraries, *Nat. Methods* **2008**, *5*, 939–942.

24. M. T. Reetz and J. D. Carballeira, Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes, *Nat. Protoc.* **2007**, *2*, 891–903.

25. S. Lee, J. Park, S. Jang, L. Nielsen, J. Kim and K. Jung, Fermentative butanol production by Clostridia, *Biotechnol. Bioeng.* **2008**, *101*, 209–228.

26. R. Gheshlaghi, J. M. Scharer, M. Moo-Young and C. P. Chou, Metabolic pathways of clostridia for producing butanol, *Biotechnol. Adv.* **2009**, *27*, 764–781.

27. T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner and H. Mori, Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection, *Mol. Syst. Biol.* **2006**, *2*, 2006.0008.

28. D. Kessler, I. Leibrecht and J. Knappe, Pyruvate-formate-lyase-deactivase and acetyl-CoA reductase activities of Escherichia coli reside on a polymeric protein particle encoded by adhE, *FEBS Lett.* **1991**, *281*, 59–63.

29. S. Atsumi, T.-Y. Wu, E.-M. Eckl, S. D. Hawkins, T. Buelter and J. C. Liao, Engineering the isobutanol biosynthetic pathway in Escherichia coli by comparison of three aldehyde reductase/alcohol dehydrogenase genes, *Appl. Microbiol. Biotechnol.* **2010**, *85*, 651–657.

30. Microbial conversion of glycerol to 1,3-propanediol by an engineered strain of Escherichia coli, *Appl. Env. Microbiol.* **2009**, *75*, 1628–1634.

31. S. Bastian, X. Liu, J. T. Meyerowitz, C. D. Snow, M. M. Y. Chen and F. H. Arnold, Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in Escherichia coli, *Metab. Eng.* **2011**, *13*, 345–352.

32. C. Dellomonaco, J. M. Clomburg, E. N. Miller and R. Gonzalez, Engineered reversal of the β-oxidation cycle for the synthesis of fuels and chemicals, *Nature* **2011**, *476*, 355–359.

33. E. G. Kurth, D. M. Doughty, P. J. Bottomley, D. J. Arp and L. A. Sayavedra-Soto, Involvement of BmoR and BmoG in n-alkane metabolism in 'Pseudomonas butanovora', *Microbiology* **2008**, *154*, 139–147.

34. J. A. Dietrich, D. L. Shis, A. Alikhani and J. D. Keasling, Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis, *ACS Synth. Biol.* **2013**, *2*, 47–58.

35. F. Zhang, J. M. Carothers and J. D. Keasling, Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids, *Nat. Biotechnol.* **2012**, *30*, 354–359.

36. F. Zhang and J. Keasling, Biosensors and their applications in microbial metabolic engineering, *Trends Microbiol.* **2011**, *19*, 323–329.

37. D. P. Clark, The fermentation pathways of Escherichia coli, *FEMS Microbiol. Rev.* **1989**, *5*, 223–234.

38. B. B. Bond-Watts, R. J. Bellerose and M. C. Y. Chang, Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways, *Nat. Chem. Biol.* **2011**, *7*, 222–227.

39. D. Parkhomchuk, V. Amstislavskiy, A. Soldatov and V. Ogryzko, Use of high throughput sequencing to observe genome dynamics at a single cell level, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 20830–20835.

40. S. S. Fong, A. Nanchen, B. O. Palsson and U. Sauer, Latent Pathway Activation and Increased Pathway Capacity Enable Escherichia coli Adaptation to Loss of Key Metabolic Enzymes, *J. Biol. Chem.* **2006**, *281*, 8024–8033.

41. D. Segre, D. Vitkup and G. M. Church, Analysis of optimality in natural and perturbed metabolic networks, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15112–15117.

42. M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones and M. A. Marra, Circos: an information aesthetic for comparative genomics, *Genome Res.* **2009**, *19*, 1639–1645.

43. C. T. Nomura, K. Taguchi, Z. Gan, K. Kuwabara, T. Tanaka, K. Takase and Y. Doi, Expression of 3-ketoacyl-acyl carrier protein reductase (fabG) genes enhances production of polyhydroxyalkanoate copolymer from glucose in recombinant Escherichia coli JM109, *Appl. Env. Microbiol.* **2005**, *71*, 4297–4306.

44. H. Lukas, J. Reimann, O. B. Kim, J. Grimpo and G. Unden, Regulation of aerobic and anaerobic D-malate metabolism of Escherichia coli by the LysR-type regulator DmlR (YeaT), *J. Bacteriol.* **2010**, *192*, 2503–2511.

45. A.-K. Becker, T. Zeppenfeld, A. Staab, S. Seitz, W. Boos, T. Morita, H. Aiba, K. Mahr, F. Titgemeyer and K. Jahreis, YeeI, a novel protein involved in modulation of the activity of the glucose-phosphotransferase system in Escherichia coli K-12, *J. Bacteriol.* **2006**, *188*, 5439–5449.

46. A.-K. Göhler, A. Staab, E. Gabor, K. Homann, E. Klang, A. Kosfeld, J.-E. Muus, J. S. Wulftange and K. Jahreis, Characterization of MtfA, a novel regulatory output signal protein of the glucose-phosphotransferase system in Escherichia coli K-12, *J. Bacteriol.* **2012**, *194*, 1024–1035.

47. K. Kimata, Y. Tanaka, T. Inada and H. Aiba, Expression of the glucose transporter gene, ptsG, is regulated at the mRNA degradation step in response to glycolytic flux in Escherichia coli, *EMBO J.* **2001**, *20*, 3587–3595.

48. T. Morita, H. Kawamoto, T. Mizota, T. Inada and H. Aiba, Enolase in the RNA degradosome plays a crucial role in the rapid decay of glucose transporter mRNA in the response to phosphosugar stress in Escherichia coli, *Mol. Microbiol.* **2004**, *54*, 1063–1075.

49. D. Yu, H. M. Ellis, E.-C. Lee, N. A. Jenkins, N. G. Copeland and D. L. Court, An efficient recombination system for chromosome engineering in Escherichia coli, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5978–5983.

50. Y. Jiang, B. Chen, C. Duan, B. Sun, J. Yang and S. Yang, Multigene Editing in the Escherichia coli Genome via the CRISPR-Cas9 System, *Appl. Env. Microbiol.* **2015**, *81*, 2506–2514.

51. H. Goodarzi, A. K. Hottes and S. Tavazoie, Global discovery of adaptive mutations, *Nat. Methods* **2009**, *6*, 581–583.

52. E. M. Quandt, D. E. Deatherage, A. D. Ellington, G. Georgiou and J. E. Barrick, Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in Escherichia coli, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 2217–2222.

53. M. K. Applebee, M. J. Herrgard and B. O. Palsson, Impact of Individual Mutations on Increased Fitness in Adaptively Evolved Strains of Escherichia coli, *J. Bacteriol.* **2008**, *190*, 5087–5094.

54. J. E. Barrick, D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski and J. F. Kim, Genome evolution and adaptation in a long-term experiment with Escherichia coli, *Nature* **2009**, *461*, 1243–1247.

55. G. Giannoukos, D. M. Ciulla, K. Huang, B. J. Haas, J. Izard, J. Z. Levin, J. Livny, A. M. Earl, D. Gevers, D. V. Ward, *et al.*, Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes, *Genome Biol.* **2012**, *13*, R23.

56. P. Picotti and R. Aebersold, Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions, *Nat. Methods* **2012**, *9*, 555–566.

57. T. S. Batth, J. D. Keasling and C. J. Petzold, Targeted proteomics for metabolic pathway optimization, *Methods Mol. Biol.* **2012**, *944*, 237–249.

58. T. S. Batth, P. Singh, V. R. Ramakrishnan, M. M. L. Sousa, L. J. G. Chan, H. M. Tran, E. G. Luning, E. H. Y. Pan, K. M. Vuu, J. D. Keasling, *et al.*, A targeted proteomics toolkit for high-throughput absolute quantification of Escherichia coli proteins, *Metab. Eng.* **2014**, *26C*, 48–56.

59. Y.-X. Zhang, K. Perry, V. A. Vinci, K. Powell, W. P. C. Stemmer and S. B. del Cardayre, Genome shuffling leads to rapid phenotypic improvement in bacteria, *Nature* **2002**, *415*, 644–646.

Chapter 4: *Development of C4 fuel and commodity chemical pathways with diverse aldehyde and alcohol dehydrogenases*

4.1 Introduction

As burgeoning synthetic biology and renewable fuel companies have learned repeatedly over the last decade, competing with staggeringly cheap fossil fuels is exceedingly difficult. Taking this reality to heart, many in academia and industry have explored the production of commodity chemicals as a (relatively) comfortable middle ground between very low margin but very high volume fuels and very high margin but very low volume fine chemicals. Industrial disclosures such as those for the development of 1,3-propandediol[1] and 1,4-butanediol[2] give insight into where metabolic engineering can be a successful and economically viable strategy.

These products and many others can be produced at a cost that can withstand a range of market forces and are not dependent upon record high oil prices paired with record low sugar prices. Additionally some or these products are "bio-advantaged", meaning that biological routes of production are superior to petrochemical routes of production, either for technical or economic reasons. Depending on the pathway used for productions, these products including malonate[3]and succinate[4] can have greater than 100% theoretical yield from sugar through carbon fixation. This makes production considerably less sensitive to the price of sugar and lowers the fraction of theoretical yield at which a product must be made to be economical. Consequently processes can be optimized more rapidly, and years of R&D are not required to push yield from 83% to 87%.

In this vein we sought to explore potential applications of the diverse set of ALDHs we had at our disposal from earlier screening efforts. Our butanol production pathway has served as a strong base for this effort. By reconfiguring the upstream portion of the pathway as well as exploring diverse sequence families of downstream ALDHs and ADHs, we have been successful in adapting our butanol production strains for production of 1,3-butanediol and 4-hydroy-2-butanone[5], both of which are useful for rubber production from butadiene. Extensive screening and optimization has been performed including DNA shuffling and saturation mutagenesis, and this work is ongoing.

4.2 Materials and methods

Commercial materials. Luria-Bertani (LB) Broth Miller, LB Agar Miller, and Terrific Broth (TB) were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), isopropyl-β-D-thiogalactopyranoside (IPTG), phenylmethanesulfonyl fluoride (PMSF), tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), sodium chloride, dithiothreitol (DTT), kanamycin (Km), ethyl acetate and ethylene diamine tetraacetic acid disodium dihydrate (EDTA), were purchased from Fisher Scientific (Pittsburgh, PA). Coenzyme A trilithium salt (CoA), acetyl-CoA, nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), β-mercaptoethanol, sodium phosphate dibasic hepthydrate, and N,N,N',N'-tetramethyl-ethane-1,2-diamine (TEMED) were purchased from Sigma-Aldrich (St. Louis, MO). Acrylamide/Bis-acrylamide (30%, 37.5:1), electrophoresis grade sodium dodecyl sulfate (SDS), Bio-Rad protein assay dye reagent concentrate and ammonium persulfate were purchased from Bio-Rad

Laboratories (Hercules, CA). Restriction enzymes, T4 DNA ligase, Phusion DNA polymerase, T5 exonuclease, and Taq DNA ligase were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs) and Platinum Taq High-Fidelity polymerase (Pt Taq HF) were purchased from Invitrogen (Carlsbad, CA). PageRuler™ Plus prestained protein ladder was purchased from Fermentas (Glen Burnie, Maryland). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a stock concentration of 100 μM in 10 mM Tris-HCl, pH 8.5, and stored at either 4°C for immediate use or -20°C for longer term use. DNA purification kits and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). Amicon Ultra 10,000 centrifugal concentrators were purchased from EMD Millipore (Billerica, MA).

Bacterial strains. *E. coli* DH10B-T1^R and BL21(de3)T1^R were used for DNA construction and heterologous protein production, respectively. *E. coli* DH1 and DH1 ΔackA-pta ΔadhE ΔldhA ΔpoxB ΔfrdBC (MC1.24) were used for metabolite production.

Gene and plasmid construction. Restriction enzyme cloning, Gibson assembly, and Golden Gate assembly were used to carry out plasmid construction. All PCR amplifications were carried out with Phusion or Platinum Taq High Fidelity DNA polymerases. All constructs were verified by sequencing (Quintara Biosciences; Berkeley, CA).

Bioinformatics search for alcohol dehydrogenases. The Fe-ADH sequence family (PF00465) was filtered using cd-hit (http://www.bioinformatics.org/cd-hit/) to remove sequences greater than 90% identical. The remaining sequences were blasted all-vs-all using BLAST and the resulting sequence similarity network was visualized in Cytoscape at various E-value cutoffs. Alcohol dehydrogenases of known substrate specificity were overlaid on the network and sequences were randomly sampled from adjacent sequence clusters.

Expression of His-tagged proteins. TB (1 L) containing carbenicillin (50 μg/mL) in a 2.8 L Fernbach baffled shake flask was inoculated to OD_{600} = 0.05 with an overnight TB culture of freshly transformed *E. coli* containing the appropriate overexpression plasmid. The cultures were grown at 37°C at 200 rpm to OD_{600} = 0.6 to 0.8 at which point cultures were cooled on ice for 20 min, followed by induction of protein expression with 1 mM IPTG and overnight growth at 16°C. Cell pellets were harvested by centrifugation at 9,800 × *g* for 7 min and resuspended at 20 mL/L of culture with Buffer A (50 mM sodium phosphate, 300 mM sodium chloride, 20 mM imidazole, 0.5 mM EDTA, pH 8.0) supplemented with 2 mg/mL lysozyme and 2 uL/50 mL final volume Benzonase and frozen at -80°C.

Purification of His-tagged proteins. Frozen cell suspensions were thawed and frozen twice before finally thawing and adding 0.5 mM PMSF as a 50 mM stock solution in ethanol dropwise. The cell suspension was lysed at with a Misonix 3000 probe sonicator at full power with a 15 second on, 60 second off cycle for a total sonication time of 2.5 minutes. The lysate was centrifuged at 15,300 × *g* for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction by addition of 1% streptomycin sulfate as a 20% w/v stock solution added dropwise. The precipitated DNA was removed by centrifugation at 15,300 × *g* for 20 min at 4°C. The lysate was loaded onto a Ni-NTA agarose column (Qiagen, 1 mL resin/L expression

culture) by gravity flow. The column was washed with 20 column volumes Buffer A. The protein was then eluted with 250 mM imidazole in Buffer A.

Fractions containing the target protein were pooled by $A_{280\ nm}$ and supplemented with 100 mM DTT to 1 mM final. TEV protease (QB3 Macrolab) was added at a 1:20 ratio w/w. Protein was then placed in 10 kDa MWCO dialysis tubing in 1.8 L Buffer A with 1 mM DTT and dialyzed overnight at 4°C.

Dialyzed protein was loaded onto the previous Ni-NTA agarose column equilibrated with Buffer A and the flow through was collected. This procedure was repeated two times and the column was washed with 1 column volume of buffer A. The pooled flow through was concentrated in an Amicon Ultra 10,000 MWCO concentrator to a final volume of 2 mL. Concentrated protein was loaded on a Superdex 200 SEC column (GE Healthcare; Piscataway, NJ) connected to an ÄKTApurifier FPLC (1 mL/min; GE Healthcare). Fractions containing ALDH protein by A_{280} were pooled and concentrated in an Amicon Ultra 10,000 MWCO concentrator. Concentrated protein was supplemented with glycerol to 10% v/v and stored at -80°C.

Crystallization and Structure Determination of GA-ALDH3 and GA-ALDH16. Protein crystals were obtained using the sitting drop vapor diffusion method by combining equal volumes of a 10 mg/mL protein solution and a reservoir solution [0.2 M tri-sodium citrate (pH 7.5) and 20% (w/v) polyethylene glycol 3350]. Crystals grew within 2 days and were cryoprotected by being briefly soaked in a solution containing 75% reservoir solution and 25% ethylene glycol followed by flash-freezing in liquid nitrogen. Data were collected at Beamline 8.3.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory, Berkeley, CA). Data sets for native crystals were collected at a wavelength of 1.116 Å. Data sets were processed and merged with XDS and XSCALE. Phases were determined by molecular replacement using Phenix AutoMR and AutoBuild to build a near-complete chain trace of each crystal. Iterative cycles of Phenix AutoRefine and manual refinement in Coot32 were used to generate the final model.

Expression of Strep-tagged proteins. TB (1 L) containing carbenicillin (50 μg/mL) in a 2.8 L Fernbach baffled shake flask was inoculated to OD_{600} = 0.05 with an overnight TB culture of freshly transformed *E. coli* containing the appropriate overexpression plasmid. The cultures were grown at 37°C at 200 rpm to OD_{600} = 0.6 to 0.8 at which point cultures were cooled on ice for 20 min, followed by induction of protein expression with 1 mM IPTG and overnight growth at 16°C. Cell pellets were harvested by centrifugation at 9,800 × *g* for 7 min and resuspended at 20 mL/L of culture with Buffer W (100 mM Tris-HCl, 150 mM sodium chloride, 1 mM EDTA, pH 8.0) supplemented with 2 mg/mL lysozyme and 2 uL/50 mL final volume Benzonase and frozen at -80°C.

Purification of Strep-tagged proteins. Frozen cell suspensions were thawed and frozen twice before finally thawing and adding 0.5 mM PMSF as a 50 mM stock solution in ethanol dropwise. The cell suspension was lysed at with a Misonix 3000 probe sonicator at full power with a 15 second on, 60 second off cycle for a total sonication time of 2.5 minutes. The lysate was centrifuged at 15,300 × *g* for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction by addition of 0.5% polyethylenimine as a 15% v/v stock solution added dropwise. The

precipitated DNA was removed by centrifugation at 15,300 × *g* for 20 min at 4°C. The lysate was loaded onto a Strep-tactin Superflow High Capacity column (IBA, 1 mL resin/L expression culture) by gravity flow. The column was washed with 20 column volumes Buffer W. The protein was then eluted with 2.5 mM desthiobiotin in Buffer W. Fractions containing ALDH protein by A_{280} were pooled and concentrated in an Amicon Ultra 10,000 MWCO concentrator. Concentrated protein was supplemented with glycerol to 10% v/v and stored at -80°C.

Enzyme assays. Activity of ALDH proteins was measured by monitoring the oxidation of NADH at 340 nm at 25 °C. The assay mixture (400 µL) contained 100 µM NADH in 100 mM Tris 1 mM DTT pH 7.5. The reaction was initiated by the addition of substrate. Kinetic parameters (k_{cat}, K_M) were determined by fitting the data using Microcal Origin to the equation: $v_o = v_{max}[S] / (K_M + [S])$, where v is the initial rate and [S] is the substrate concentration. Data are reported as mean ± s.e. (n = 3) unless otherwise noted with standard error derived from the nonlinear curve fitting. Error bars on graphs represent mean ± s.d. (n = 3). Error in k_{cat}/K_M is calculated by propagation of error from the individual kinetic parameters.

Cell culture. *E. coli* strains were transformed by electroporation using the appropriate plasmids. A single colony from a fresh transformation was then used to seed an overnight culture grown in Terrific Broth (TB) (EMD Biosciences) supplemented with 1.5% (w/v) glucose and appropriate antibiotics at 37 °C in a rotary shaker (200 rpm). Antibiotics were used at a concentration of 50 µg ml^{-1} for strains with a single resistance marker. For strains with multiple resistance markers, kanamycin and chloramphenicol were used at 25 µg ml^{-1} and carbenicillin was used at 50 µg ml^{-1}.

***In vivo* production of alcohols.** Overnight cultures of freshly transformed *E. coli* strains were grown for 12–16 h in TB at 37 °C and used to inoculate TB (50 ml) with glucose replacing the standard glycerol supplement (1.5% (w/v) glucose for aerobic cultures and 2.5% (w/v) glucose for anaerobic cultures) and appropriate antibiotics to an optical density at 600 nm (OD_{600}) of 0.05 in a 250 mL-baffled flask or a 250 mL-baffled anaerobic flask. The cultures were grown at 37 °C in a rotary shaker (200 rpm) and induced with IPTG (1.0 mM) at OD_{600} = 0.35–0.45. At this time, the growth temperature was reduced to 30 °C, and the culture flasks were sealed with Parafilm M (Pechiney Plastic Packaging) to prevent product evaporation for aerobic cultures. Anaerobic cultures were sealed and the headspace was sparged with argon for 3 minutes immediately follow induction. Aerobic cultures were unsealed for 10 to 30 min every 24 h then resealed with Parafilm M, and additional glucose (1% (w/v)) was added 1 day post-induction. Samples were quantified after 3 d of cell culture.

Quantification of n-butanol. Samples (2 ml) were removed from cell culture and cleared of biomass by centrifugation at 20,817g for 2 min using an Eppendorf 5417R centrifuge. The supernatant or cleared medium sample was then mixed in a 9:1 ratio with an aqueous solution containing the isobutanol internal standard (10,000 mg l^{-1}). These samples were then analyzed on a Trace GC Ultra (Thermo Scientific) using an HP-5MS column (0.25 mm × 30 m, 0.25 µM film thickness, J & W Scientific). The oven program was as follows: 75 °C for 3 min, ramp to 300 °C at 45 °C min^{-1}, 300 °C for 1 min. n-Butanol was quantified by flame ionization detection (FID) (flow: 350 ml min^{-1} air, 35

ml min^{-1} H2 and 30 ml min^{-1} helium). Samples containing n-butanol levels below 500 mg l^{-1} were requantified after extraction of the cleared medium sample or standard (500 µl) with toluene (500 µl) containing the isobutanol internal standard (100 mg l^{-1}) using a Digital Vortex Mixer (Fisher) for 5 min set at 2,000. The organic layer was then quantified using the same GC parameters with a DSQII single-quadrupole mass spectrometer (Thermo Scientific) using single-ion monitoring (m/z 41 and 56) concurrent with full scan mode (m/z 35–80). Samples were quantified relative to a standard curve of 2, 4, 8, 16, 31, 63, 125, 250, 500 mg l^{-1} n-butanol for MS detection or 125, 250, 500, 1,000, 2,000, 4,000, 8,000 mg l^{-1} n-butanol for FID detection. Standard curves were prepared freshly during each run and normalized for injection volume using the internal isobutanol standard (100 or 1,000 mg l^{-1} for MS and FID, respectively).

Quantification of crotyl alcohol. Samples (2 ml) were removed from cell culture and cleared of biomass by centrifugation at 20,817g for 2 min using an Eppendorf 5417R centrifuge. The cleared medium sample or standard (500 µl) was extracted with toluene (500 µl) containing the isobutanol internal standard (100 mg l^{-1}) using a Digital Vortex Mixer (Fisher) for 5 min set at 2,000. The organic layer was then analyzed on a Trace GC Ultra (Thermo Scientific) using an HP-5MS column (0.25 mm × 30 m, 0.25 µM film thickness, J & W Scientific). The oven program was as follows: 75 °C for 4 min, ramp to 300 °C at 45 °C min^{-1}, 300 °C for 2 min. Crotyl alcohol was detected with a DSQII single-quadrupole mass spectrometer (Thermo Scientific) using single-ion monitoring (m/z 29, 41, 43, and 57) concurrent with full scan mode (m/z 37-58). Samples were quantified relative to a standard curve of 2, 4, 8, 16, 31, 63, 125, 250, 500 mg l^{-1} crotyl alcohol for MS detection. Standard curves were prepared freshly during each run and normalized for injection volume using the internal isobutanol standard (100 mg l^{-1}).

Quantification of 1,3-butanediol. Samples (2 ml) were removed from cell culture and cleared of biomass by centrifugation at 20,817g for 2 min using an Eppendorf 5417R centrifuge. The cleared medium samples, or standards prepared in TB medium, were diluted 1:100 into water and filtered through a 0.22 µm filter (EMD Millipore MSGVN2210). The samples were analyzed on an Agilent 1290 HPLC (Agilent) using a Rezex ROA-Organic Acid H+ (8%) column (150 x 4.6 mm, Phenomenex) with isocratic elution using 0.5% formic acid (0.3 mL/min, 55°C). Samples were detected with an Agilent 6460C triple quadrupole MS with Jet Stream ESI source (Agilent), operating in positive MRM mode (91-73 transition, fragmentor 50 V, collision energy 0 V, cell accelerator voltage 7 V, delta EMV +400). Samples were quantified relative to a standard curve of 31, 63, 125, 250, 500, 1000, 2000, 4000 mg l^{-1} 1,3-butanediol.

Quantification of 4-hydroxy-2-butanone. Samples (2 ml) were removed from cell culture and cleared of biomass by centrifugation at 20,817g for 2 min using an Eppendorf 5417R centrifuge. The cleared medium samples, or standards prepared in TB medium, were diluted 1:100 into water and filtered through a 0.22 µm filter (EMD Millipore MSGVN2210). The samples were analyzed on an Agilent 1290 HPLC (Agilent) using a Rezex ROA-Organic Acid H+ (8%) column (150 x 4.6 mm, Phenomenex) with isocratic elution using 0.5% formic acid (0.3 mL/min, 55°C). Samples were detected with an Agilent 6460C triple quadrupole MS with Jet Stream ESI source (Agilent), operating in positive MRM mode (89-71 transition, fragmentor 50 V, collision energy 0 V, cell

accelerator voltage 7 V, delta EMV +400). Samples were quantified relative to a standard curve of 31, 63, 125, 250, 500, 1000, 2000, 4000 mg l^{-1} 1,3-butanediol.

DNA shuffling of aldehyde dehydrogenases. ALDH PCR products were treated with variable units of DNaseI for increasing time intervals from 30 seconds to 7 minutes to generate fragments of a desired size range; typically 2 or 4 U of DNaseI treatment for 1-3 minutes at 30 °C yielded fragments centered around 400 bp. After purification of digested fragments in the desired size range by gel-extraction, chimeric reassembly is achieved by PCR without primers. Reassembly of small fragments or fragments with low homology was generally difficult and required optimal reassembly conditions to be determined by varying the number of PCR cycles, annealing temperature, amount of template, and polymerase type. Here full length reassembled products were achieved using 1 μL of template for 30 cycles and 0.1 μL of template for 25 or 30 cycles with Phusion polymerase. Reassembled full length products were cloned into production plasmids by Golden Gate assembly and transformed into commercial electrocompetent cels (NEB). Pooled transformations were recovered in 50 mL LB for 1 hour at 37 °C, then dilutions were plated on agar plates and appropriate antibiotics were added to the liquid culture for growth overnight. The next day the plasmid library was recovered by miniprep and colonies were counted to determine total library size, typically greater than 1×10^6.

Saturation mutagenesis of aldehyde dehydrogenases. Six regions of three residues each were chosen for NNK saturation mutagenesis. The ALDH expression plasmid was cloned with an RFP dropout cassette interupting the ALDH gene such that Golden Gate cloning of degenerate oligo cassettes would restore the open reading frame and target the desired residues for NNK saturation. Oligos were phosphorylated with T4 PNK (NEB), annealed and slowly cooled in a thermalcycler, and used for Golden Gate cloning with the ALDH plasmid. Ligated plasmid was transformed into commercial electrocompetent cels (NEB). Pooled transformations were recovered in 50 mL LB for 1 hour at 37 °C, then dilutions were plated on agar plates and appropriate antibiotics were added to the liquid culture for growth overnight. The next day the plasmid library was recovered by miniprep and colonies were counted to determine total library size, typically greater than 1×10^7.

Anaerobic growth selection of aldehyde dehydrogenase mutants. Following mutagenesis cultures were serially transferred to fresh media every 24-72 hours to approximate continuous growth with limited time spent in stationary phase. Growth media was TB with 2.5% glucose, 1 mM IPTG, and appropriate antibiotics. Culture OD_{600} was monitored daily and cultures were transferred when the majority of cultures were in late log-phase growth, usually OD_{600} 1.5-2.0. Culture supernatant samples (2 mL) were collected for metabolite quantification. All cultures were transferred simultaneously, the headspace was sparged with argon for 3 minutes, and growth was continued at 30 °C in a rotary shaker (200 rpm). Selections were continued for up to three weeks and were terminated when mutant cultures ceased growth rate improvement or when the growth rate of wild type cultures began to improve. Final cultures were stored as 15% glycerol stocks at -80 °C in addition to being streaked on LB agar plates. Individual colonies were picked and cultured for metabolite production in TB to confirm butanediol and hydroxybutanone production relative to wild type strains.

4.3 Results and discussion

Exploration of C4 commodity chemical production

To explore alternatives and expansions to the butanol pathway, upstream enzymatic steps were removed such that the diverse aldehyde dehydrogenases would catalyze reactions on former pathway intermediates. The initial target pathways are diagramed in *Figure 4.1*. Crotyl alcohol can be produced by removing the trans-enoyl-CoA reductase (TER) from the pathway such that the ALDH.ADH pair reduces crotonyl-CoA to crotonaldehhye and crotyl alcohol (*Figure 4.1 B*). 1,3-butanediol can be produced by additionally removing the crotonase (crt) such that the ALDH.ADH pair reduces 3-hydroxybutyryl-CoA to 3-hydroxybutyraldehyde and 1,3-butanediol (*Figure 4.1 C*). Results from an initial screen are presented in *Figure 4.2*. The pathways depicted in *Figure 4.1* were cloned with each ALDH in our library, transformed into DH1, and metabolite production was quantified. In general, butanol and butanediol titers ranged from several 100 mg/L to 1.7 g/L. In contrast crotyl alcohol production was limited to merely 6 mg/L. Potential causes for this low titer include low steady-state concentration of crotonyl-CoA or poor acceptance of this substrate by the enzymes tested. Regardless of the reason, crotyl alcohol production was not examined further. As for butanol and butanediol production, some ALDHs displayed little preference for one product or the other, while other ALDHs showed a strong preference. From this starting point we decided to focus on butanediol production.

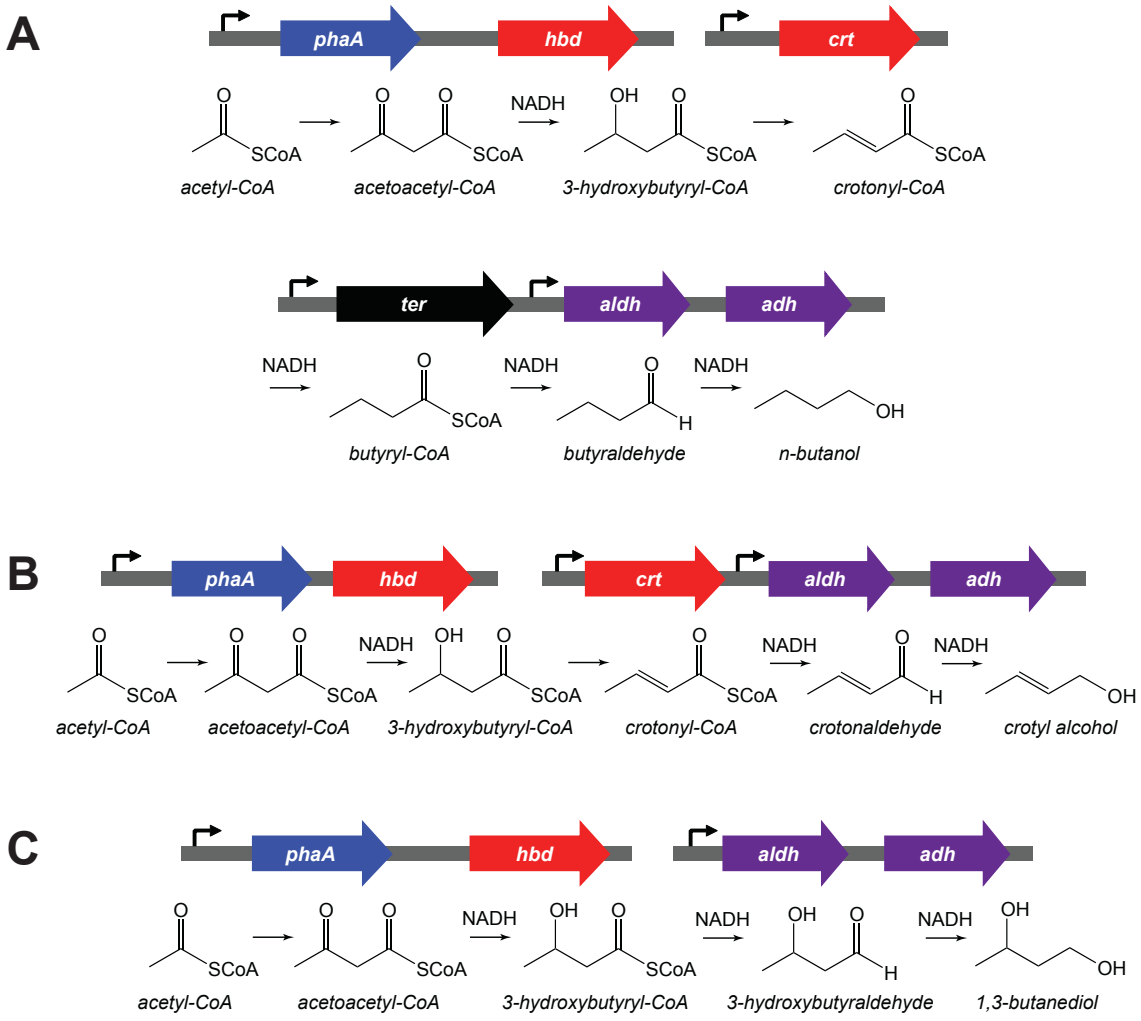


Figure 4.1 C4 fuel and commodity chemical pathways employing diverse aldehyde and alcohol dehydrogenases

Building upon the proven butanol production pathway, we sought to expand the number of compounds we could produce by leveraging a diverse collection of aldehyde and alcohol dehydrogenases. By removing intermediate enzymes from the butanol production pathway (A) it is possible to produce crotyl alcohol (B) and 1,3-butanediol (C). Both are commodity chemicals with a variety of applications, the largest being conversion to butadiene for rubber manufacturing.

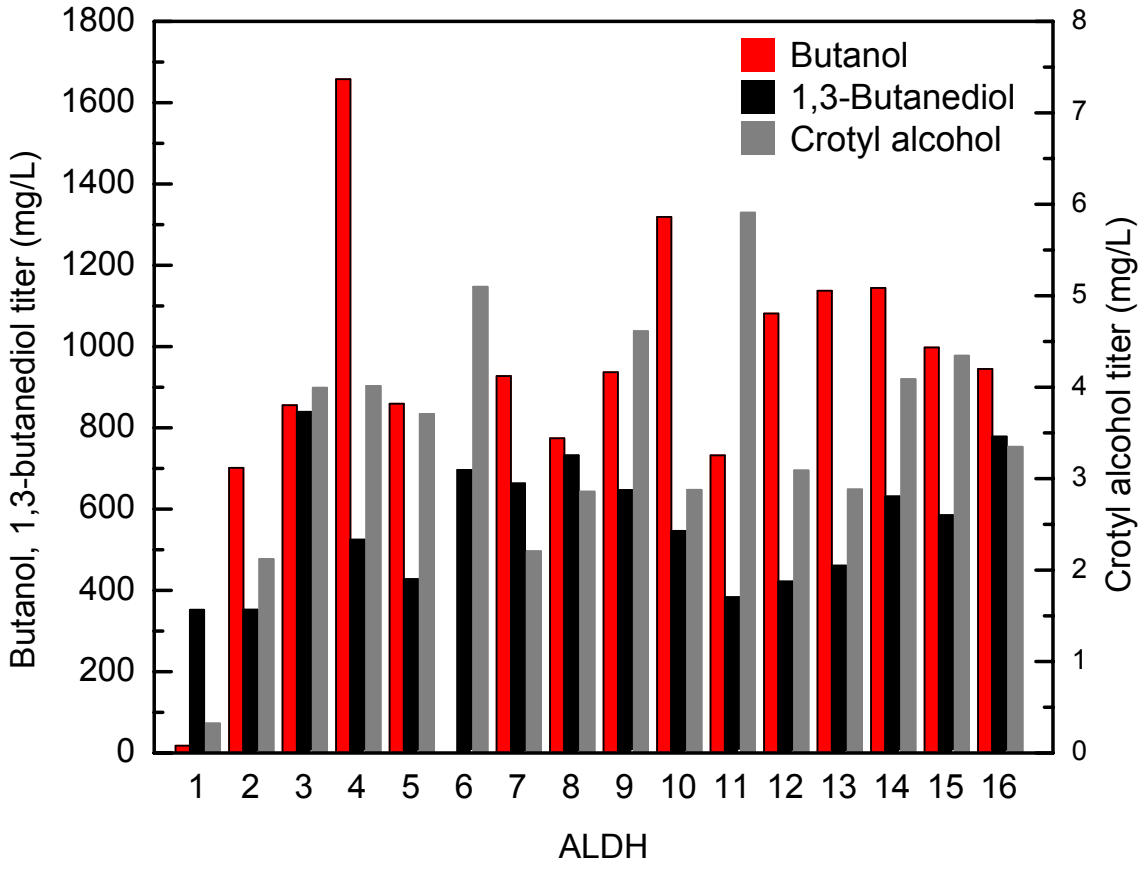


Figure 4.2 Production of diverse alcohols using aldehyde dehydrogenases
The pathways depicted in Figure 4.1 were cloned with each ALDH in our library, transformed into DH1, and metabolite production was quantified. In general, butanol and butanediol titers ranged from several 100 mg/L to 1.7 g/L. In contrast crotyl alcohol production was limited to merely 6 mg/L. Potential causes for this low titer include low steady-state concentration of crotonyl-CoA or poor acceptance of this substrate by the enzymes tested. As for butanol and butanediol production, some ALDHs display little preference for one product or the other, while other ALDHs show a strong preference.

Butanediol pathway optimization

With our initial screen confirming that significant butanediol production was possible from a number of ALDHs, we next characterized its response to the stereochemistry of 3-hydroxybutyryl-CoA. Our standard butanol production pathway utilizes HBD which produces (*S*)-3-hydroxybutyryl-CoA, but an alternative pathway using phaB is equally as effective at producing (*R*)-3-hydroxybutyryl-CoA[6]. When using these pathways in the context of butanediol production with only an ALDH we noticed almost no difference in production based on the two substrates (*Figure 4.3*). We found it somewhat surprising that except for two small exceptions, all of the ALDHs tested appeared to have no preference for the stereochemistry of the 3-hyddroxyl group.

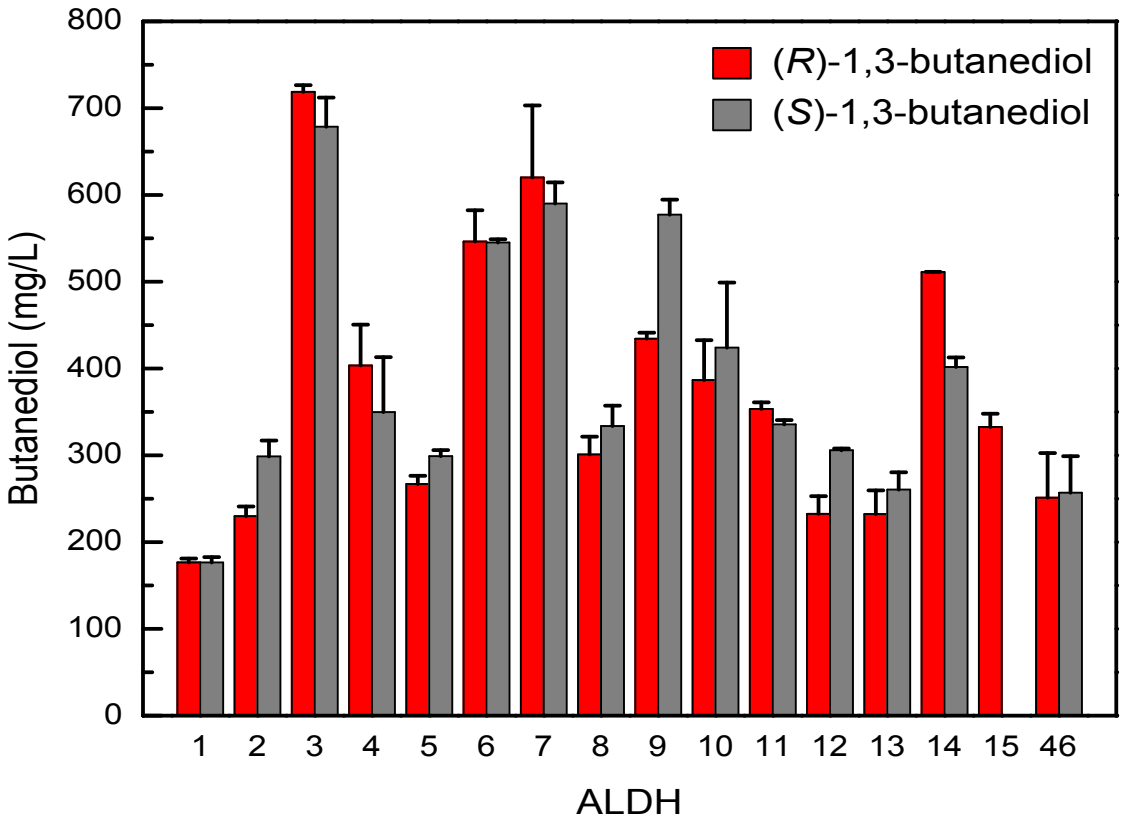


Figure 4.3 Screening ALDH library for production of (R) and (S)-1,3-butanediol
The upstream butanediol pathway can produce either (R)-3-hydroxybutyryl-CoA or (S)-3-hydroxybutyryl-CoA, depending on the use of phaB or HBD as the acetoacetyl-CoA reductase respectively. Both pathways have been show to enable significant butanol production, but ALDH specificity of these enantiomers was unknown. Screening the ALDH library with both upstream pathways revealed that most ALDHs have little preference for one substrate over another. Data are mean ± s.d. (n = 3).

With this initial confirmation that many of our ALDHs would support at least moderate production of butanediol, we next sought to improve our butanediol production pathway in much the same way as we approached improving the butanol pathway. We sought out diverse sets of alcohol dehydrogenases that might be able to accommodate 1,3-hydroxybutyraldehyde. To this point we were again relying on native *E.* coli ADHs with suitable promiscuity to complete the final reduction of 3-hydroxybutyraldehyde to butanediol. To identify alcohol dehydrogenases that would efficiently reduce 3-hydroxybutyraldehyde to 1,3-butanediol, we generated a sequence similarity network[7] of the ADH family. Sequence similarity networks (SSNs) are a methodology used to sort through large families of sequences with the goal of identifying subfamilies that may have properties such as substrate specificity unique to that subfamily[8]. These networks have commonly been used to identify such subfamilies within large superfamilies[9]. Although the overall effect of dividing sequence space at increasing similarity between sequences is similar to that accomplished by phylogenetic trees, the visual representation

and ability to dynamically change the stringency make SSNs very useful for broadly exploring a family of sequences.

To generate the SSN all ADH sequences in the Pfam database[10] were filtered to remove sequences of greater than 90% identity[11]. This decreases the number of sequences to a size that can be searched by blast in hours instead of days without materially decreasing the diversity of the collection. Filtered sequences were blasted against each other and then clustered using Cytoscape[12] at increasingly stringent e-values, such that subfamilies become apparent (*Figure 4.4*). In SSNs each dot (node) represents a sequence, and each line between nodes (edge) represents a percent identity between two sequences that is above the cutoff. As the stringency of the percent identity is increased edges between nodes are removed and large clusters of sequences begin to separate into smaller subfamilies. This network was then overlaid with ADHs of known substrate specificity[13-16] as a frame of reference, and the stringency was increased such that these known enzymes were reasonably well separated. Sequences from the network were then broadly sampled to maximize diversity and increase the likelihood of identifying a highly active ADH.

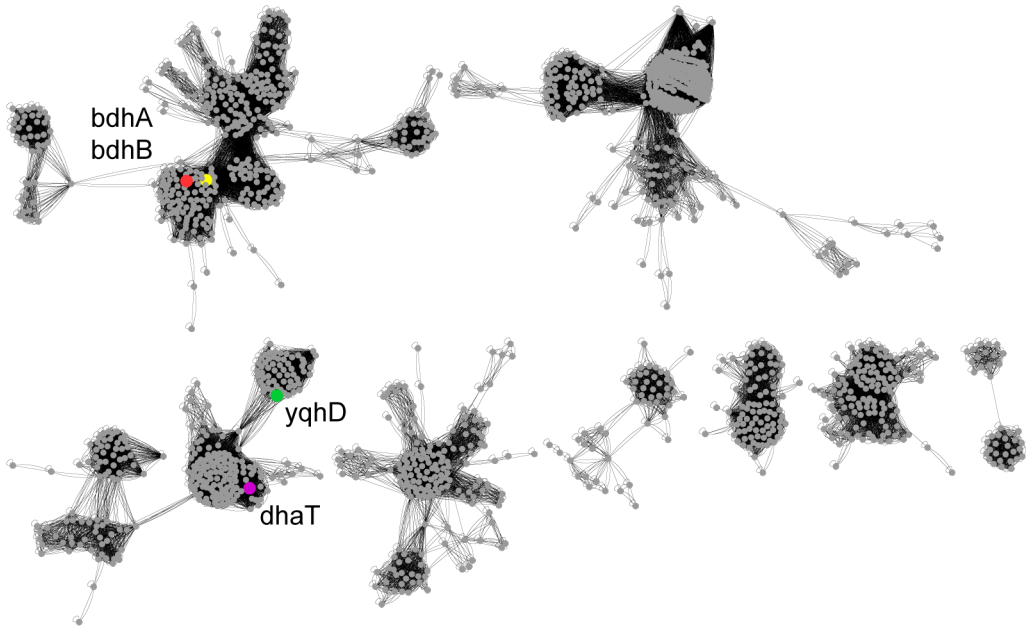


Figure 4.4 Sequence similarity network of monofunctional alcohol dehydrogenases
To identify alcohol dehydrogenases that would efficiently reduce 3-hydroxybutyraldehyde to 1,3-butanediol, we generated a sequence similarity network of the ADH family. ADH sequences were blasted against each other and then clustered at increasingly stringent e-values, such that subfamilies become apparent. This network was then overlaid with ADHs of known substrate specificity as a frame of reference. Sequences from the network were then broadly sampled to maximize diversity and increase the likelihood of identifying a highly active ADH.

Gene	Accession	Organism
adh1	B6YQP9_AZOPC	Azobacteroides pseudotrichonymphae genomovar. CFP2
adh2	A0RQF7_CAMFF	Campylobacter fetus subsp. fetus (strain 82-40)
adh3	G5F136_9ACTN	Olsenella sp. oral taxon 809 str. F0356
adh4	B1C7G7_9FIRM	Anaerofustis stercorihominis DSM 17244
adh5	YUGK_BACSU	Bacillus subtilis (strain 168)
adh6	A8SGI9_9FIRM	Faecalibacterium prausnitzii M21/2
adh7	E2SQ66_9FIRM	Erysipelotrichaceae bacterium 3_1_53
adh8	E1QYZ8_OLSUV	Olsenella uli (strain ATCC 49627
adh9	F5X0G1_STRG1	Streptococcus gallolyticus (strain ATCC 43143 / F-1867)
adh10	E6W4G5_DESIS	Desulfurispirillum indicum (strain ATCC BAA-1389 / S5)
adh11	E6K7W2_9BACT	Prevotella buccae ATCC 33574
adh12	B1C4Z8_9FIRM	Clostridium spiroforme DSM 1552
adh13	G4L3E3_TETHN	Tetragenococcus halophilus (strain DSM 20338
adh14	E8LLW8_9GAMM	Succinatimonas hippei YIT 12066
dhaT2	E4RKV2_HALSL	Halanaerobium hydrogeniformans (Halanaerobium sp)
dhaT3	Q15G22_CITFR	Citrobacter freundii
dhaT4	A0PY50_CLONN	Clostridium novyi (strain NT)
dhaT5	Q3A1K9_PELCD	Pelobacter carbinolicus (strain DSM 2380 / Gra Bd 1)
dhaT6	A5D4X5_PELTS	Pelotomaculum thermopropionicum (strain DSM 13744)
dhaT7	B1V2D9_CLOPF	Clostridium perfringens D str. JGS1721
dhaT8	E3H9G9_ILYPC	Ilyobacter polytropus (strain DSM 2926 / CuHBu1)
adh15	Q1JYE4_DESAC	Desulfuromonas acetoxidans DSM 684
adh16	B5YIE2_THEYD	Thermodesulfovibrio yellowstonii (strain ATCC 51303)
adh17	D2BSS7_DICD5	Dickeya dadantii (strain Ech586)
adh18	F0ERB1_HAEPA	Haemophilus parainfluenzae ATCC 33392
adh19	G5IQ05_9ENTE	Enterococcus saccharolyticus 30_1
adh20	B2V5D0_CLOBA	Clostridium botulinum (strain Alaska E43 / Type E3)
adh21	E2SME8_9FIRM	Erysipelotrichaceae bacterium 3_1_53
adh22	B0NYL0_9CLOT	Clostridium sp. SS2/1

Table 4.5 Bioinformatically identified alcohol dehydrogenases

Alcohol dehydrogenases identified with a sequence similarity network (Figure 4.4) are from a diverse set of bacterial species, and none have been previously characterized. The top set was initially screened (Figure 4.6) for (R) and (S)-1,3-butanediol production with aldh46, and top performing ADHs were cloned combinatorially with top performing ALDHs (Figure 4.7). Upon identifying the formation of 4-hyddroxy-2-butanone as a side-product, additional ADHs similar to adh2, 8, and 12 were sampled from the network (bottom set) and combinatorially screened (Figure 4.10) with top ALDHs for high butanediol production and improved product ratio.

Alcohol dehydrogenases identified with a sequence similarity network (*Table 4.5*) are from a diverse set of bacterial species. The top portion of the list was initially screened (*Figure 4.6*) for (*R*) and (*S*)-1,3-butanediol production with aldh46, which we had previously shown to be very competent for butanediol production and to show no preference for the stereochemistry of the 3-hydroxybutyryl-CoA substrate. This ensured that any enantiomeric excess observed would be due to the ADH. Several ADHs did markedly improve production, but to our surprise they were all highly specific for (*R*)-1,3-butanediol production.

ADHs are generally thought to be somewhat promiscuous, so it was unexpected that the ADHs displayed a preference while the ALDHs did not. Regardless of the substrate specificity, we had now identified several ADHs that significantly increased production above what was achieved relying on the native *E. coli* ADH. No search for the *E. coli* ADH was performed.

Having identified a number of ALDHs and ADHs competent for high titer butanediol production, we screened the combinatorial set of candidate enzymes to find optimal combinations (*Figure 4.7*). This screen identified aldh7.adh2 as the best overall performer, with several other combinations performing similarly well.

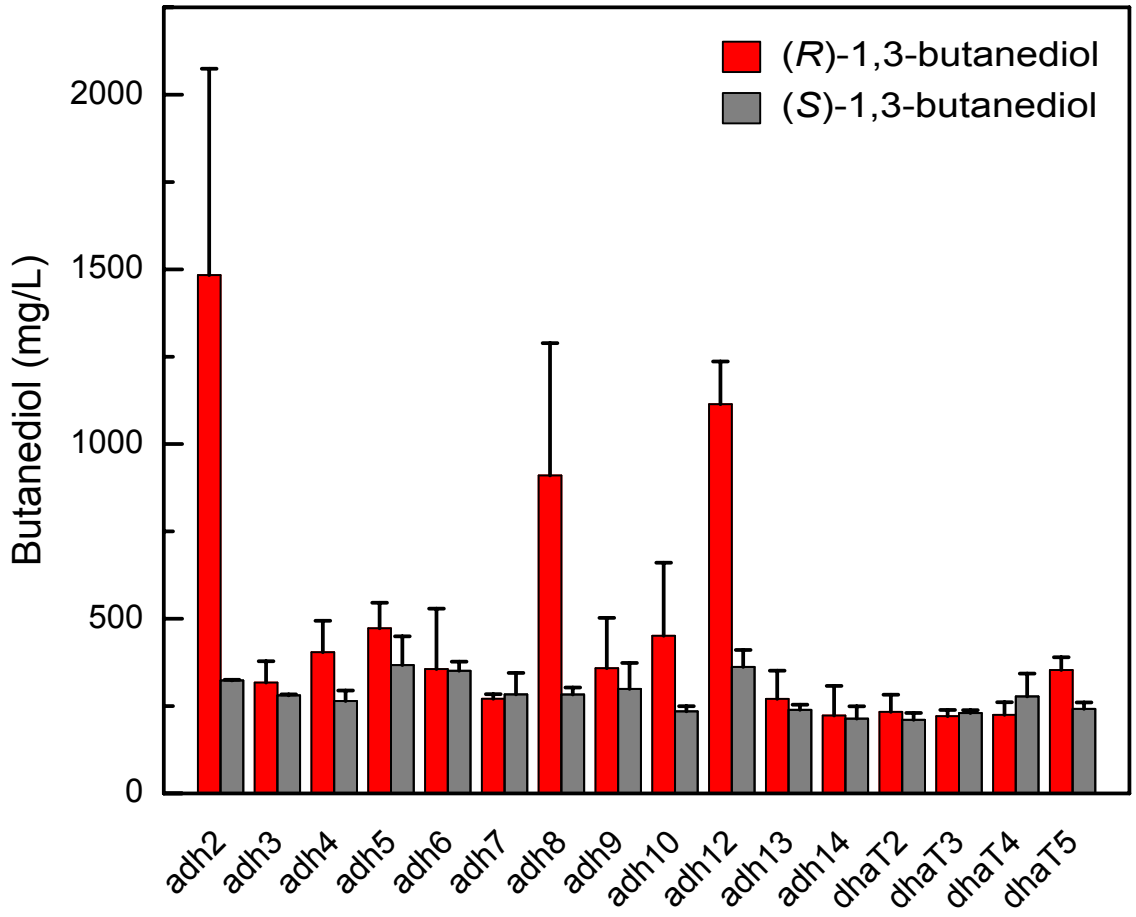


Figure 4.6 Screening ADH library for production of (R) and (S)-1,3-butanediol
The bioinformatically selected alcohol dehydrogenases were cloned into pathways for the production of (R) and (S)-1,3-butanediol using phaB and HBD respectively. ALDH46 was used as it is one of the top performing enzymes and showed no preference for substrate stereochemistry, thus any enantiomeric excess in butanediol production could be attributed to the specificity of the ADH. Surprisingly, no ADHs contributed to substantial production of (S)-1,3-butanediol, but three ADHs conferred significant titers of (R)-1,3-butanediol. This was unexpected as alcohol dehydrogenases are generally thought to be somewhat promiscuous enzymes. Data are mean ± s.d. (n = 3).

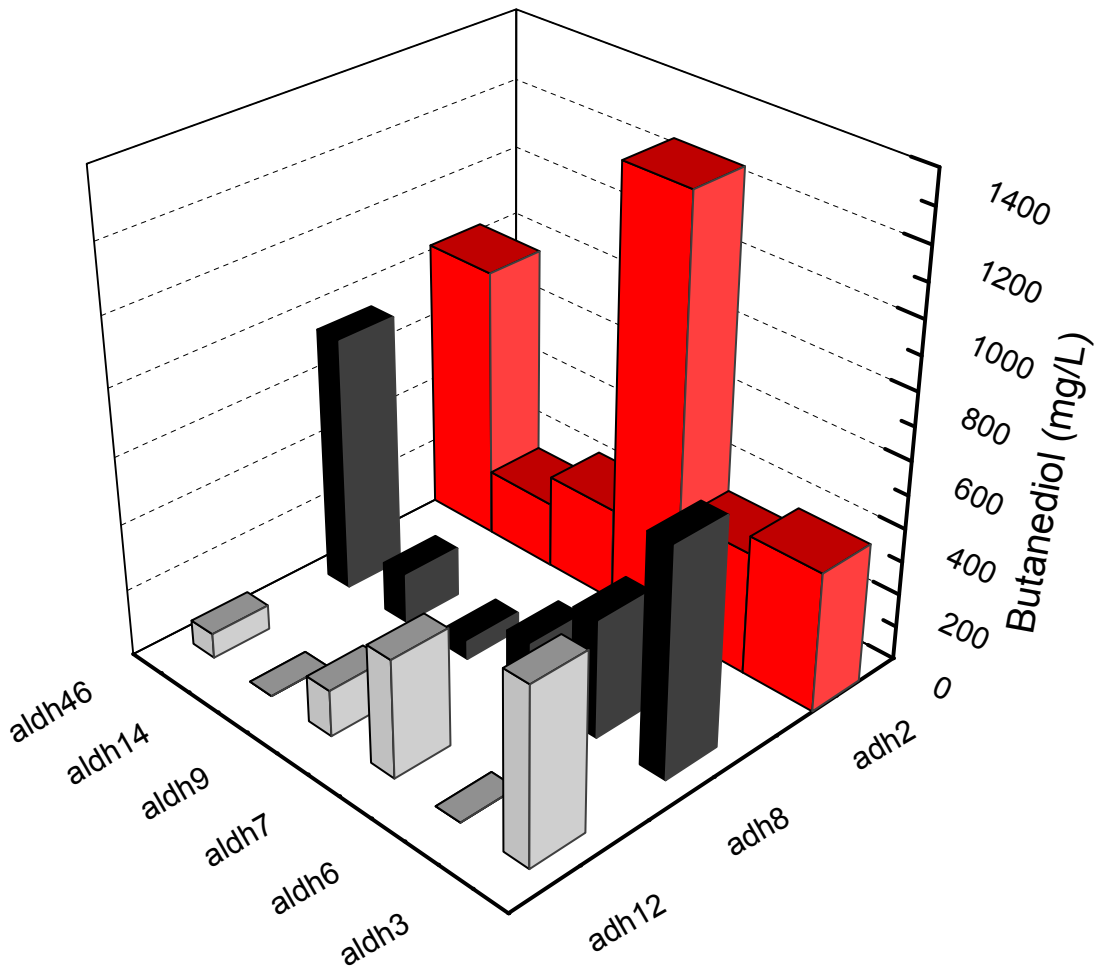


Figure 4.7 Combinatorial screening of ALDH.ADH pairs for (R)-1,3-butanediol production
Having identified a number of ALDHs and ADHs competent for high titer butanediol production, we screened the combinatorial set of candidate enzymes to find optimal combinations. This screen identified aldh7.adh2 as the best overall performer, with several other combinations performing similarly well. Data are mean (n = 3).

Identification and control of off-pathway products

While analyzing the results of this combinatorial screen we noted the appearance of an unexpected peak in GC-MS quantification of butanediol production. Closer examination identified this compound as 4-hydroxy-2-butanone, which appeared to be a significant side-product present in some cultures. Hydroxybutanone may be produced by reduction of an earlier pathway intermediate, acetoacetyl-CoA, by an ALDH, followed by subsequent reduction of acetoacetaldehyde by an ADH (*Figure 4.8*). This phenomenon had not been witnessed in previous production experiments.

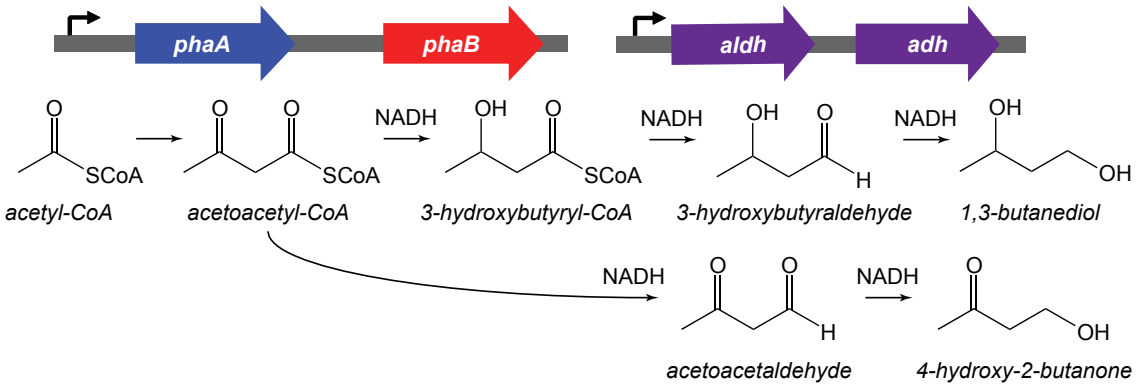


Figure 4.8 Discovery of pathway side-products resulting from a promiscuous aldh.adh pair *Stemming from appearance of an unexpected peak in GC-MS quantification of butanediol production, we identified 4-hydroxy-2-butanone as a significant side-product present in some cultures. Hydroxybutanone may be produced by reduction of an earlier pathway intermediate, acetoacetyl-CoA, by an ALDH, followed by subsequent reduction of acetoacetaldehyde by an ADH. This phenomenon had not been witnessed in previous production experiments.*

Upon discovering the unexpected production of hydroxybutanone as a side product of butanediol production, we reanalyzed media supernatant samples from the combinatorial screen of ALDH.ADH pairs (*Figure 4.9*). This revealed that hydroxybutanone production is highly specific to the aldh7.adh2 pair. Even more surprisingly, hydroxybutanone production of nearly 1.2 g/L was observed, equal to the titer of butanediol produced by the same ALDH.ADH pair. Strains carrying this pathway produced up to 2.5 g/L of mixed C4 metabolites.

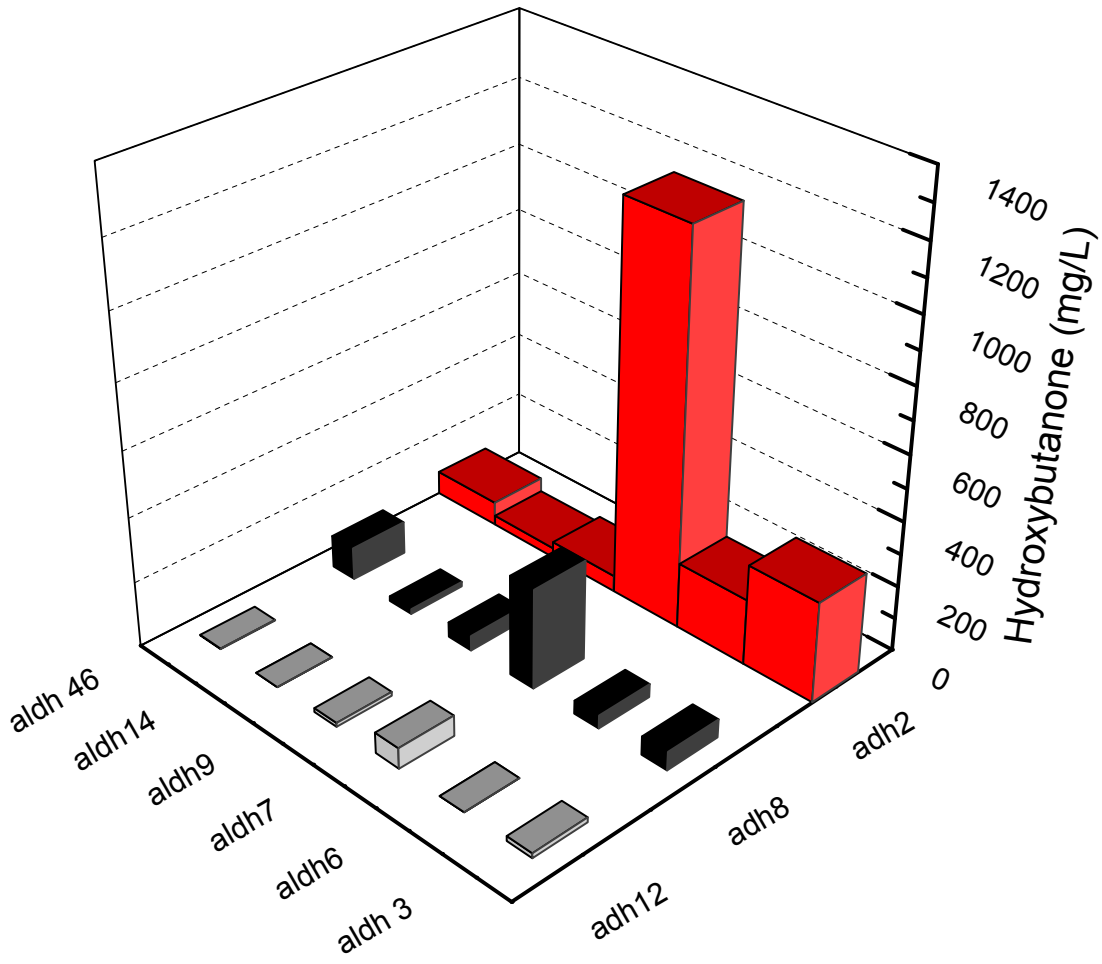


Figure 4.9 Combinatorial screening of ALDH.ADH pairs 4-hydroxy-2-butanone production
Upon discovering the unexpected production of hydroxybutanone as a side product of butanediol production, we reanalyzed media supernatant samples from a combinatorial screen of aldh.adh pairs. This revealed that hydroxybutanone production is highly specific to the aldh7.adh2 pair. Even more surprisingly, hydroxybutanone production of nearly 1.2 g/L was observed, equal to the titer of butanediol produced by the same aldh.adh pair. Strains carrying this pathway can produce up to 2.5 g/L of mixed C4 metabolites. Data are mean (n = 3).

To attempt to alter butanediol and hydroxybutanone product profiles, additional ADHs from the sequence similarity network (*Figure 4.4*) were sampled to identify enzymes with greater specificity that would not enable hydroxybutanone production. The subfamilies containing adh2, 8, and 12 were sampled at greater depth as these ADHs were shown to be most active in the initial screen.

The second set of ADHs (*Table 4.5*) was again cloned combinatorially with high performing ALDHs, including the only ALDH capable of supporting significant hydroxybutanone production, aldh7. This screen identified a new optimal pair, aldh3.adh22, capable of capturing a large fraction of the C4 product pool as butanediol, and producing 3 g/L of total products (*Figure 4.10*).

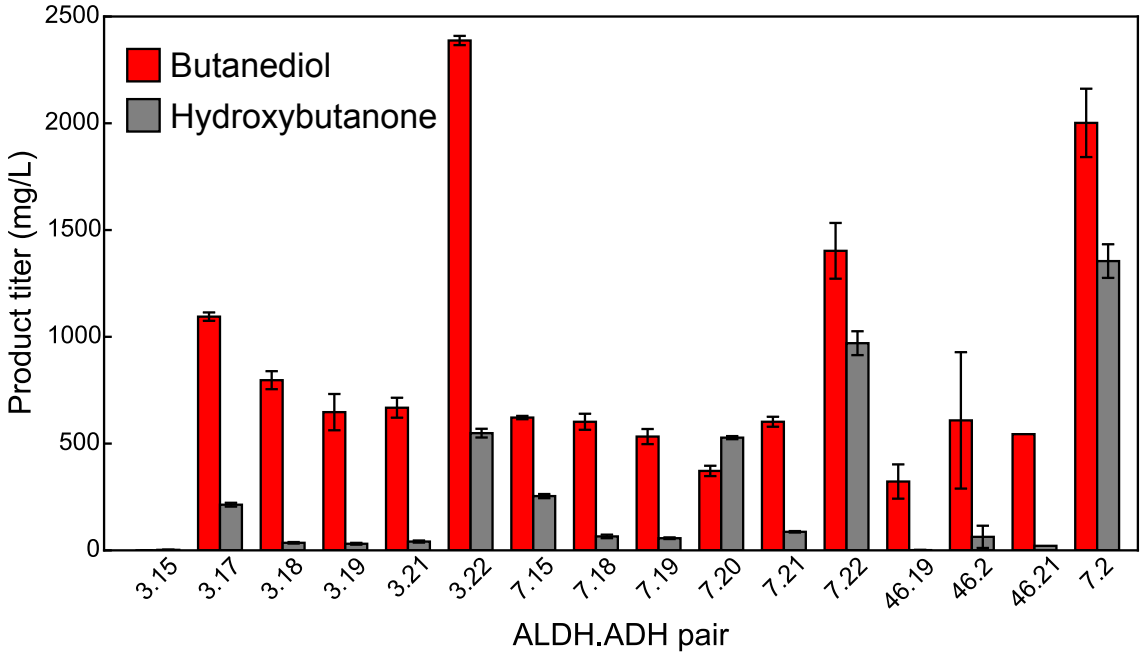


Figure 4.10 Additional alcohol dehydrogenase screening to identify higher specificity enzymes

To attempt to alter butanediol and hydroxybutanone product profiles, additional ADHs from the sequence similarity network (Figure 4.4) were sampled to identify enzymes with greater specificity that would not enable hydroxybutanone production. The subfamilies containing adh2, 8, and 12 were sampled at greater depth as these ADHs were shown to be most active in the initial screen. The second set of ADHs was again cloned combinatorially with high performing ALDHs, including the only ALDH capable of supporting significant hydroxybutanone production, aldh7. This screen identified a new optimal pair, aldh3.adh22, capable of capturing a large fraction of the C4 product pool as butanediol, and producing 3 g/L of total products. Data are mean ± s.d. (n = 3).

Although the approach of increasing specificity at the ADH step in the pathway did increase production of butanediol, this is not the optimal step in the pathway to enforce specificity. Any acetoacetyl-CoA that is reduced by a permissive ALDH to acetoacetaldehyde may become trapped in a pathway using an ADH that does not accept this substrate. Acetoacetaldehyde could eventually be converted back to acetoacetyl-CoA by the ALDH performing the backwards reaction, but it is more likely that the aldehyde would be lost to the supernatant or react with something else in the cell. This potential loss of substrate would be small when using ALDHs with minimal activity on acetoacetyl-CoA, but we still sought to address this possibility in another way.

As an alternative strategy for altering the ratio of butanediol and hydroxybutanone that would not preclude off-pathway acetoacetaldehyde from conversion to butanediol, we designed a pathway that would accept acetoacetaldehyde, reduce it to 4-hydroxy-2-butanone, and then further reduce it to 1,3-butanediol. The net product of this pathway would ultimately be butanediol, but some carbon would be channeled through 3-hydroxybutyraldehyde and some carbon would be channeled through acetoacetaldehyde.

To implement this pathway the biochemical literature was thoroughly surveyed to identify secondary alcohol dehydrogenases (SADHs) either reported to reduce 4-hydroxy-2-butanone to 1,3-butanediol or reported to have broad specificity for similar substrates. A substantial number of these enzymes have been reported in bacteria, yeast, and parasitic protozoa[5]. These enzymes are generally classified as zinc or iron-alcohol dehydrogenases and maximum percent identity within the sequences represented here range from 27-76% (*Table 4.11*).

Gene	Accession	Organism	Reference
SADH1	KGK36767.1	Pichia kudriavzevii	17
SADH2	WP_011011186.1	Pyrococcus furiosus DSM 3638	18,19
SADH3	WP_011614641.1	Cupriavidus necator	20
SADH4	P14941.1	Thermoanaerobacter brockii	21
SADH5	AAA23199.2	Clostridium beijerinckii	15
SADH6	XP_455102.1	Kluyveromyces lactis NRRL Y-1140	22
SADH7	AAP39869.1	Phytomonas sp. ADU-2003	23
SADH8	Q0KDL6.1	Ralstonia eutropha H16	24
SADH9	XP_001580601.1	Trichomonas vaginalis G3	25,26
SADH10	AJP52792.1	Pseudomonas fluorescens	27
SADH11	WP_011835462.1	Lactococcus lactis	28
SADH12	AAC04974.1	Saccharomyces cerevisiae	29
SADH13	WP_000374004.1	Escherichia coli	30
SADH14	BAD32689.1	Zygoascus ofunaensis	31
SADH15	BAA24528.1	Candida parapsilosis	32,33
SADH16	BAN45671.1	Cyberlindnera jadinii	34
SADH17	CAD36475.1	Rhodococcus ruber	35

Table 4.11 Identification of secondary alcohol dehydrogenases for reduction of hydroxybutanone to butanediol
The biochemical literature was thoroughly surveyed to identify secondary alcohol dehydrogenases either reported to reduce 4-hydroxy-2-butanone to 1,3-butanediol or reported to have broad specificity for similar substrates. These enzymes have been reported in bacteria, yeast, and parasitic protozoa. These enzymes are generally classified as zinc or iron-alcohol dehydrogenases and maximum percent identity within the sequences represented here range from 27-76%.

The identified SADHs were cloned into pathways with aldh7.adh2 (which consistently produced an even mixture of butanediol and hydroxybutanone), cultured, and metabolite production was quantified. As hoped, many SADHs shifted the product profile compared to the aldh7.adh2 control; at least four SADHs enabled butanediol production of 2 g/L with hydroxybutanone production limited to 250 mg/L or less (*Figure 4.12*). This pathway design appears preferable to enforcing specificity through an ADH that will not accept acetoacetaldehyde; acetoacetaldehyde is no longer a dead end product and can still be channeled to butanediol production.

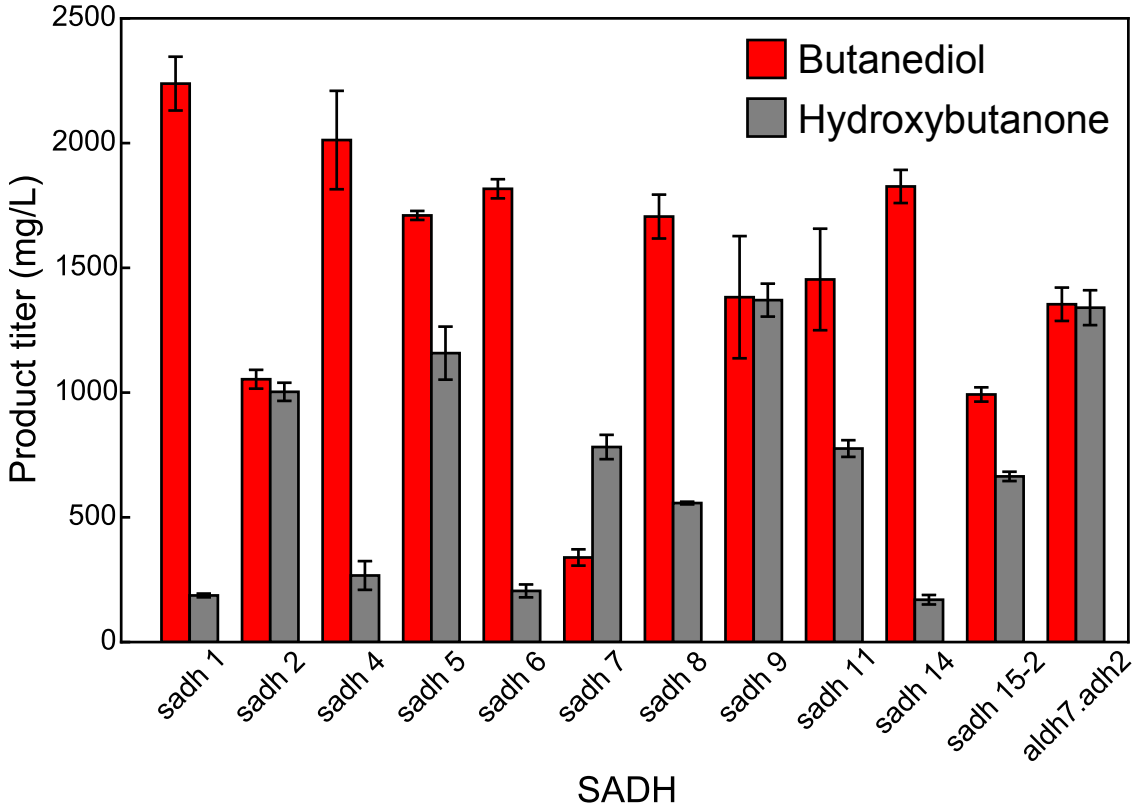


Figure 4.12 Screen of secondary alcohol dehydrogenases for reduction of hydroxybutanone to butanediol

The identified SADHs were cloned into pathways with aldh7.adh2 (50:50 product profile), cultured, and metabolite production was quantified. Encouragingly, many SADHs shifted the product profile compared to the aldh7.adh2 control; at least four SADHs enabled butanediol production of 2 g/L with hydroxybutanone production limited to 250 mg/L or less. Data are mean ± s.d. (n = 3).

Extensive screening of candidate ALDHs and (S)ADHs and pathway optimization enables tight control of the butanediol:hydroxybutanone product profile (*Figure 4.13*). Maximum hydroxybutanone production is achieved with a pathway that does not express an acetoacetyl-CoA reductase and thus can only supply acetoacetyl-CoA to aldh7.adh2. An even mixture of products can be achieved when an acetoacetyl-CoA reductase is added, thus allowing aldh7.adh2 to reduce both acetoacetyl-CoA and 3-hydroxybutyrl-CoA. Finally maximum butanediol titer can be achieved when the pathway is equipped with sadh1, which yields a two-tier pathway where half of the flux proceeds through 3-hydroxybutyryl-CoA to butanediol, and half of the flux proceeds through hydroxybutanone. The ability to deliver an arbitrary product profile through balancing expression level of these enzymes affords a great deal of control, and opens the door to applications where tunable product profile is desired, such as catalytic upgrading to longer-chain compounds[36].

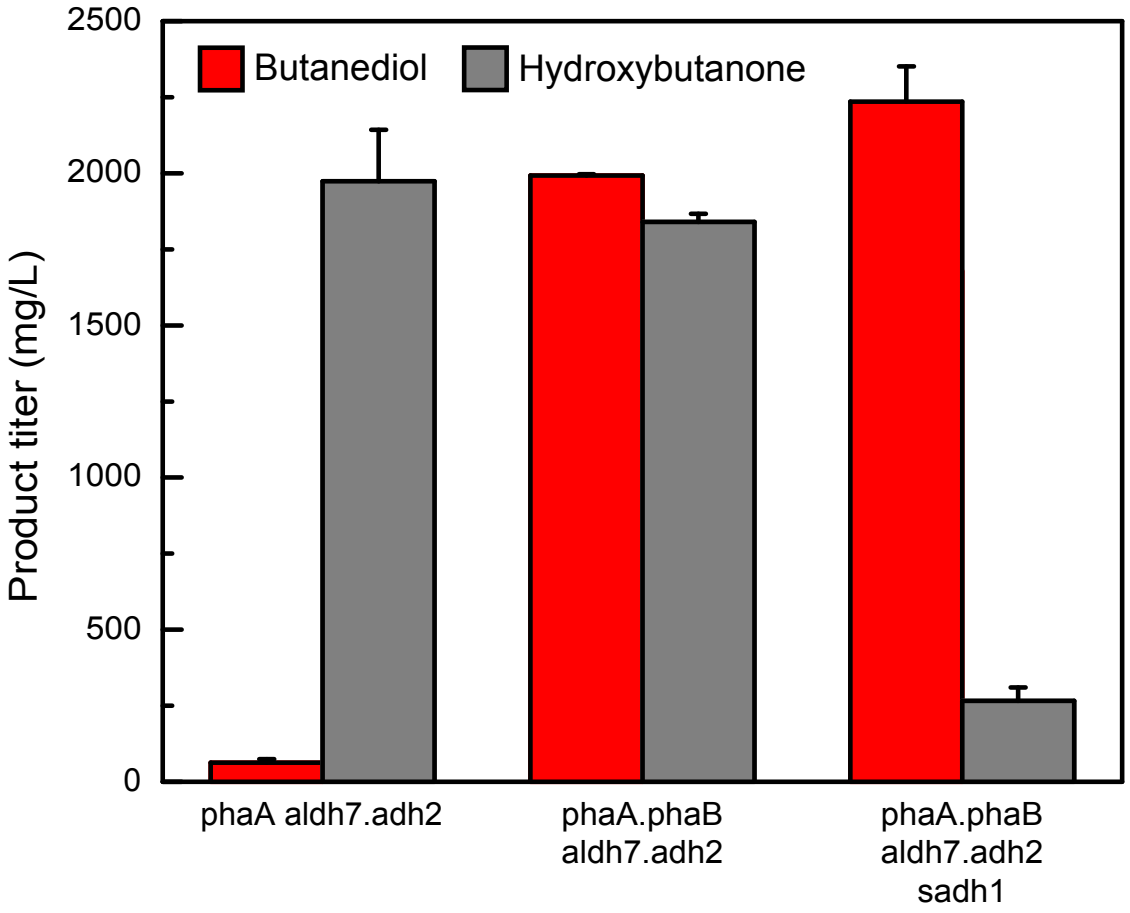


Figure 4.13 Control of butanediol:hydroxybutanone ratios through pathway design
Extensive screening of candidate ALDHs and (S)ADHs and pathway optimization enables tight control of the product profile. Maximum hydroxybutanone production is achieved with a pathway that only supplies acetoacetyl-CoA to aldh7.adh2. An even mixture of products can be achieved when an acetoacetyl-CoA reductase is added, thus allowing aldh7.adh2 to reduce both acetoacetyl-CoA and 3-hydroxybutyrl-CoA. Finally maximum butanediol titer can be achieved when the pathway is equipped with sadh1, which yields a two-tier pathway where half of the flux proceeds through 3-hydroxybutyryl-CoA to butanediol, and half of the flux proceeds through hydroxybutanone. The ability to deliver an arbitrary product profile through balancing expression level affords a great deal of control, and opens the door to using these compounds as polymer precursors. Data are mean ± s.d. (n = 3).

Directed evolution of aldehyde dehydrogenases

Concurrent to optimization of butanediol production through pathway design, we also employed directed evolution of alter the substrate specificity of ALDHs. We explored both DNA shuffling and saturation mutagenesis with mixed success. DNA shuffling was initially attractive as we already possessed the needed library of moderately diverse ALDHs. We also employed saturation mutagenesis, which is easier to implement, and proved more effective in this case.

DNA shuffling[37,38] is an effective means of producing diverse and highly-active libraries of enzymes for directed evolution (*Figure 4.14 A)*. DNA shuffling begins by

subjecting a library of diverse but related genes to partial digestion by DNaseI to generate fragments of a desired size. Fragments are reassembled by PCR without primers by relying on regions of homology within the related genes to serve as primer annealing and extension sites. Annealing and extension of fragments from different sources yields a longer chimeric fragment, and by the nature of PCR this process cascades to produce full-length chimeric genes with tunable fragment size and crossover frequency. This library is then subjected to selection to enrich improved variants, which are typically recycled for additional shuffling and selection.

In implementing DNA shuffling for directed evolution of ALDH substrate specificity[39] we used our existing library of ALDHs, which vary in percent identity from 52-97%. ALDH PCR products were treated with 2 or 4 units of DNaseI for increasing time intervals from 30 seconds to 7 minutes (*Figure 4.14 B)*. By varying the amount of DNaseI and the length of digestion a desired fragment size of ~400 bp is achieved. After purification of digested fragments in the desired size range, chimeric reassembly is achieved by PCR without primers (*Figure 4.14 C)*. Reassembly of small fragments or fragments with low homology can be quite challenging, and optimal reassembly conditions were determined by varying the number of PCR cycles, annealing temperature, amount of template, polymerase type, and more. Here full length reassembled products were achieved using 1 μL of template for 30 cycles and 0.1 μL of template for 25 or 30 cycles.

After generation of a diverse DNA shuffled ALDH library estimated to contain greater than 1 x 10^7 total mutants, anaerobic growth selection was employed in a similar manner to EMS mutagenesis selections in *Figure 3.13*. Clones from the shuffled library were sequenced prior to selection and found to be highly diverse and contained the desired average fragment size of ~400 bp.

Following selection multiple clones from independent cultures were sequenced (*Figure 4.15*) and encouragingly several cultures had become monoclonal during the selection. Each bar in the figure represents the chimeric makeup of twelve individual clones, and the colored segments denote the parent sequence of that fragment. One mutant was observed in five clones, two mutants were observed in two clones each, and the remaining three clones were unique but contained significant similarity to other clones.

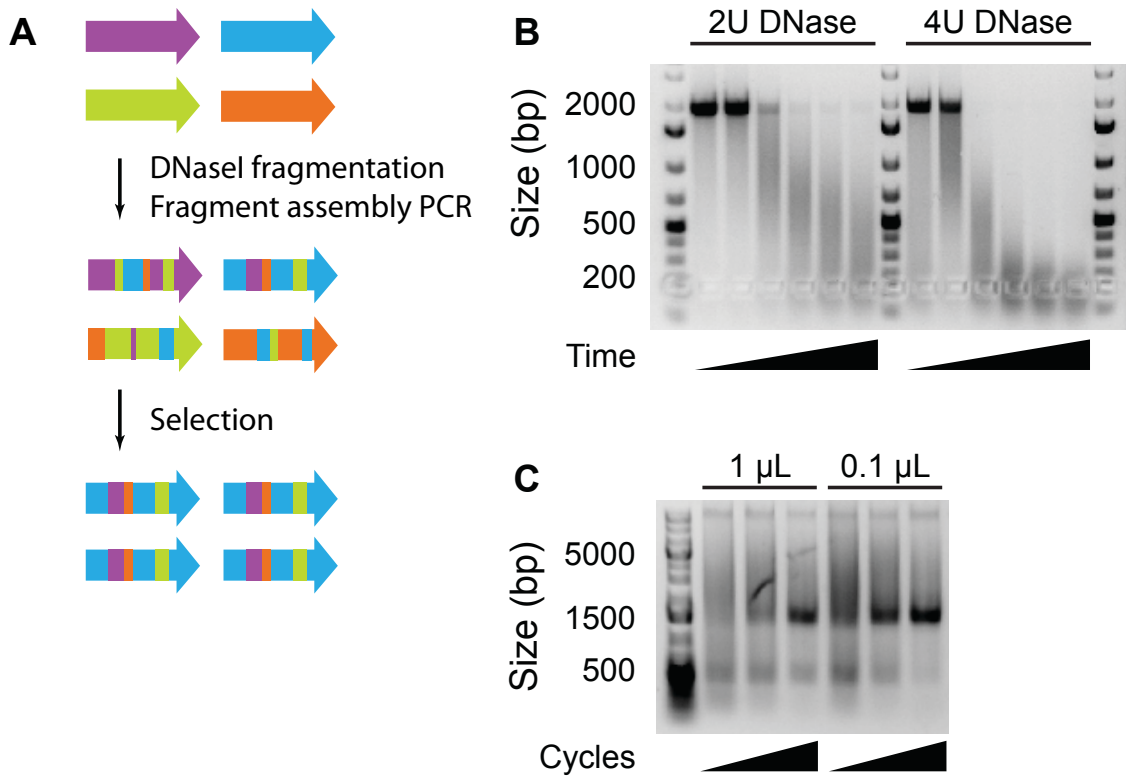


Figure 4.14 Generation of DNA shuffled ALDH libraries to improve substrate specificity and activity

DNA shuffling is an effective means of producing diverse and highly active libraries of enzymes for directed evolution. (A) DNA shuffling begins by subjecting a library of diverse but related genes to partial digestion by DNaseI to generate fragments of a desired size. Fragments are reassembled by PCR without primers by relying on regions of homology within the related genes to serve as primer annealing and extension sites. Annealing and extension of fragments from different sources yields a longer chimeric fragment, and by the nature of PCR this process cascades to produce full-length chimeric genes with tunable fragment size and crossover frequency. This library is then subjected to selection to enrich improved variants, which are typically recycled for additional shuffling and selection. (B) ALDH PCR products were treated with 2 or 4 units of DNaseI for increasing time intervals from 30 seconds to 7 minutes. By varying the amount of DNaseI and the length of digestion a desired fragment size of ~400 bp is achieved. (C) After purification of digested fragments in the desired size range, chimeric reassembly is achieved by PCR without primers. Reassembly of small fragments or fragments with low homology can be quite challenging, and optimal reassembly conditions must be determined by varying the number of PCR cycles, annealing temperature, amount of template, polymerase type, and more. Here full length reassembled products (1500 bp) were achieved using 1 µL of template for 30 cycles and 0.1 µL of template for 25 or 30 cycles.

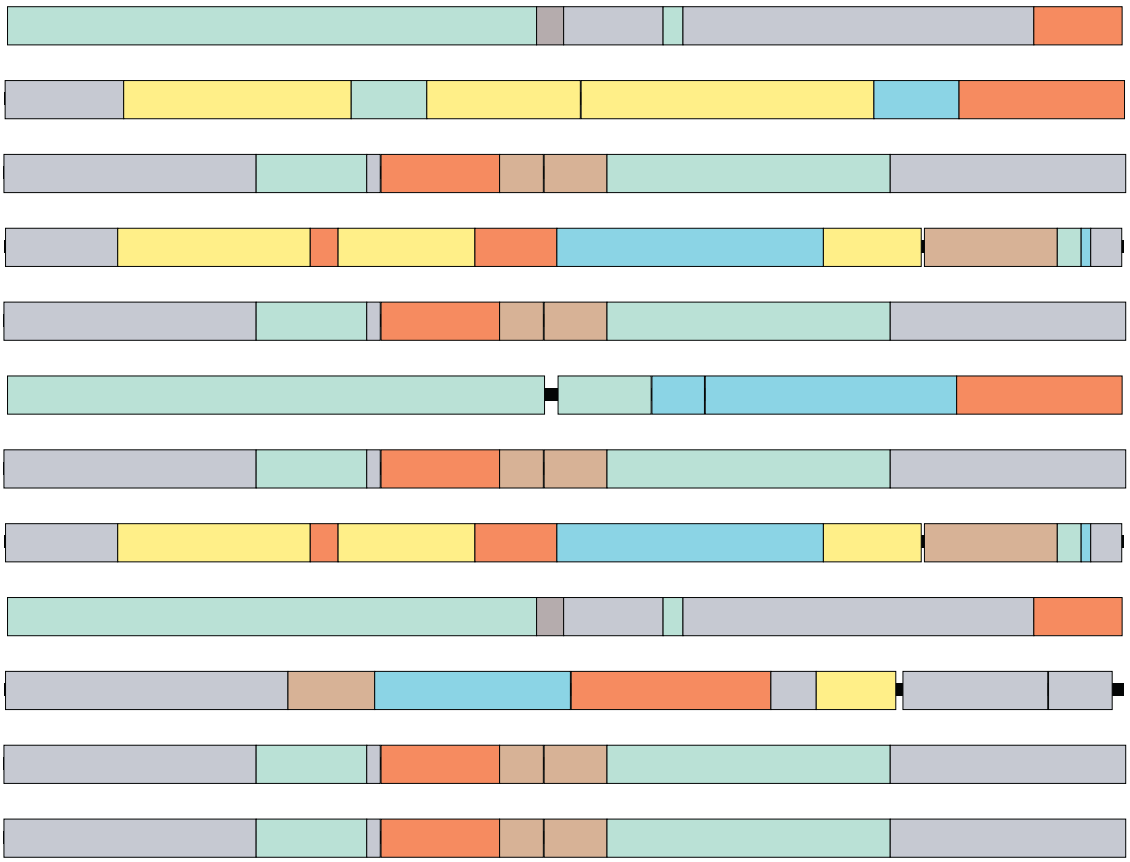


Figure 4.15 Chimeric structure of DNA shuffled ALDH clones following anaerobic growth selection

After generation of a diverse DNA shuffled ALDH library estimated to contain greater than 1 x 10^7 total mutants, anaerobic growth selection was employed in a similar manner to EMS mutagenesis selections in Figure 3.13. Clones from the shuffled library were sequenced prior to selection and found to be highly diverse and contained the desired average fragment size of ~400 bp. Following selection multiple clones from independent cultures were sequenced and encouragingly several cultures had become monoclonal during the selection. Each bar above represents the chimeric makeup of twelve individual clones, and the colored segments denote the parent sequence of that fragment. One mutant was observed in five clones, two mutants were observed in two clones each, and the remaining three clones are unique but contain significant similarity to other clones.

Unfortunately, despite evidence of efficient enrichment of DNA shuffled libraries such that monoclonal cultures arose, followup screening of these clones showed little improvement in total titer or substrate specificity. It is possible that DNA shuffling targeting different fragment sizes, or including a different subset of parental ALDH sequences would be more effective, but at this stage we turned to alternative methods.

As an alternative and complementary approach to DNA shuffling, saturation mutagenesis was employed[40,41]. Improvements to substrate specificity are more efficiently achieve by mutations close to the active site[42], so the X-ray crystal structure of aldh46 was determined to assist in selection of target residues. The preliminary structure was of moderate resolution (3.0 Å) and further crystallography attempts are ongoing, but

it was of sufficient quality to aide in selection of regions to target for saturation mutagenesis (*Figure 4.16*). Six regions of three residues each were chosen for independent NNK saturation mutagenesis. NNK mutagenesis includes 32 codons comprising every amino acid and one stop codon; this degeneracy normalizes the abundance of different amino acids and decreases the number of stop codons, which typically result in nonfunctional library members. With these parameters a moderately sized library of 1×10^5 must be produced to ensure 95% coverage of all possible variants. Following saturation mutagenesis of each region independently, additional mutagenesis of the remaining regions can be applied to the top hit for each region, allowing rapid exploration of the structure-function landscape[41].

Aldh7 mutant libraries were produced with saturation mutagenesis and selected through anaerobic growth (*Figure 4.17*). Titers at the endpoint of selection were highly variable and some cultures had lost all productivity. However, culture C4-2, containing a library targeting residues 445-447 (Thr-Phe-Thr, see *Figure 4.16*, blue region), produced 5 g/L butanediol and 1.5 g/L hydroxybutanone. This compares favorably to the wild type, which produced 4 g/L butanediol and 2 g/L hydroxybutanone, and the combined titer of 6.5 g/L is our highest achieved to date. Upon sequencing, clones in this culture were found to contain Thr445 mutated to serine or glycine. In the wild type structure Thr445 is 7.7 Å from the active site Cys279 and forms the opposite wall of the active site. Thr445 also makes close contacts with two other loops that further define the active site cavity opposite the catalytic cysteine.

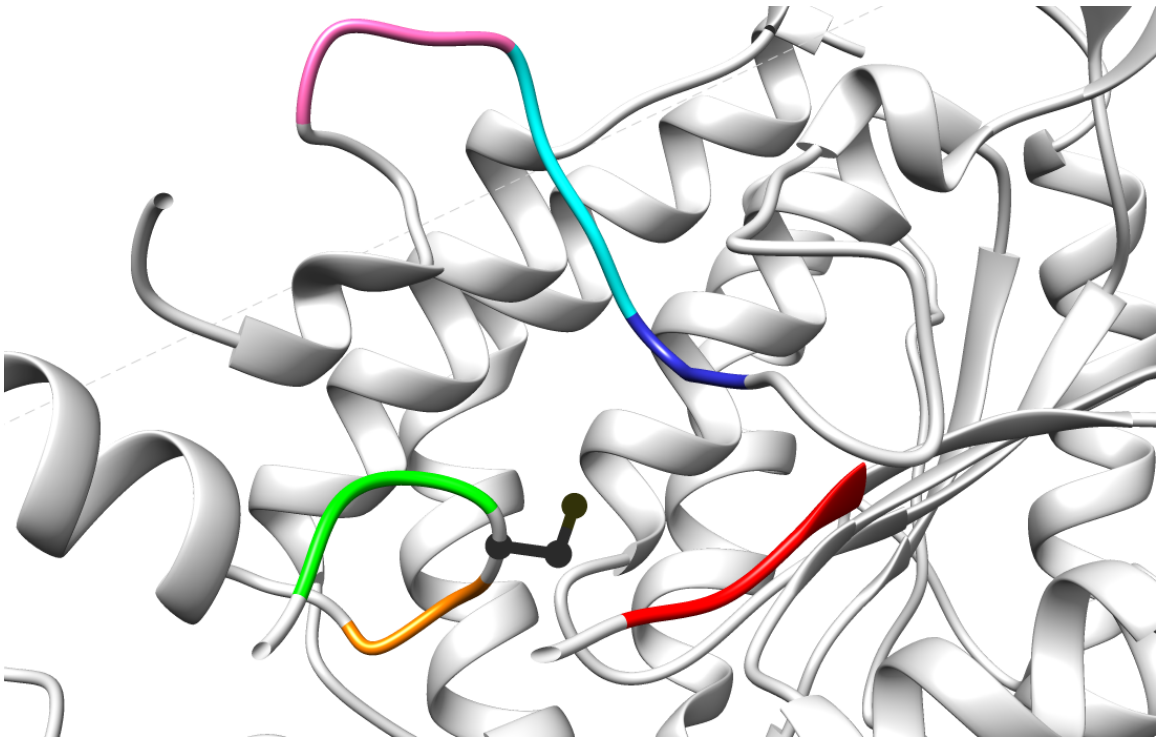


Figure 4.16 Active site of aldh46 and regions targeted for saturation mutagenesis
As an alternative and complementary approach to directed evolution of ALDHs for greater substrate specificity and activity, saturation mutagenesis was employed. Improvements to substrate specificity are more efficiently achieve by mutations close to the active site, so the X-ray crystal structure of aldh46 was determined to assist in selection of target residues. The catalytic cysteine is shown in sticks, and the targeted residues are depicted in colors. Six regions of three residues each were chosen for NNK saturation mutagenesis. NNK mutagenesis includes 32 codons comprising every amino acid and one stop codon; this degeneracy normalizes the abundance of different amino acids and decreases the number of stop codons, which typically result in nonfunctional library members. With these parameters a moderately sized library of 1 x 10^5 must be produced to ensure 95% coverage of all possible variants. Following saturation mutagenesis of each region independently, additional mutagenesis can be applied to the top hit for each region, allowing rapid exploration of the structure-function landscape.

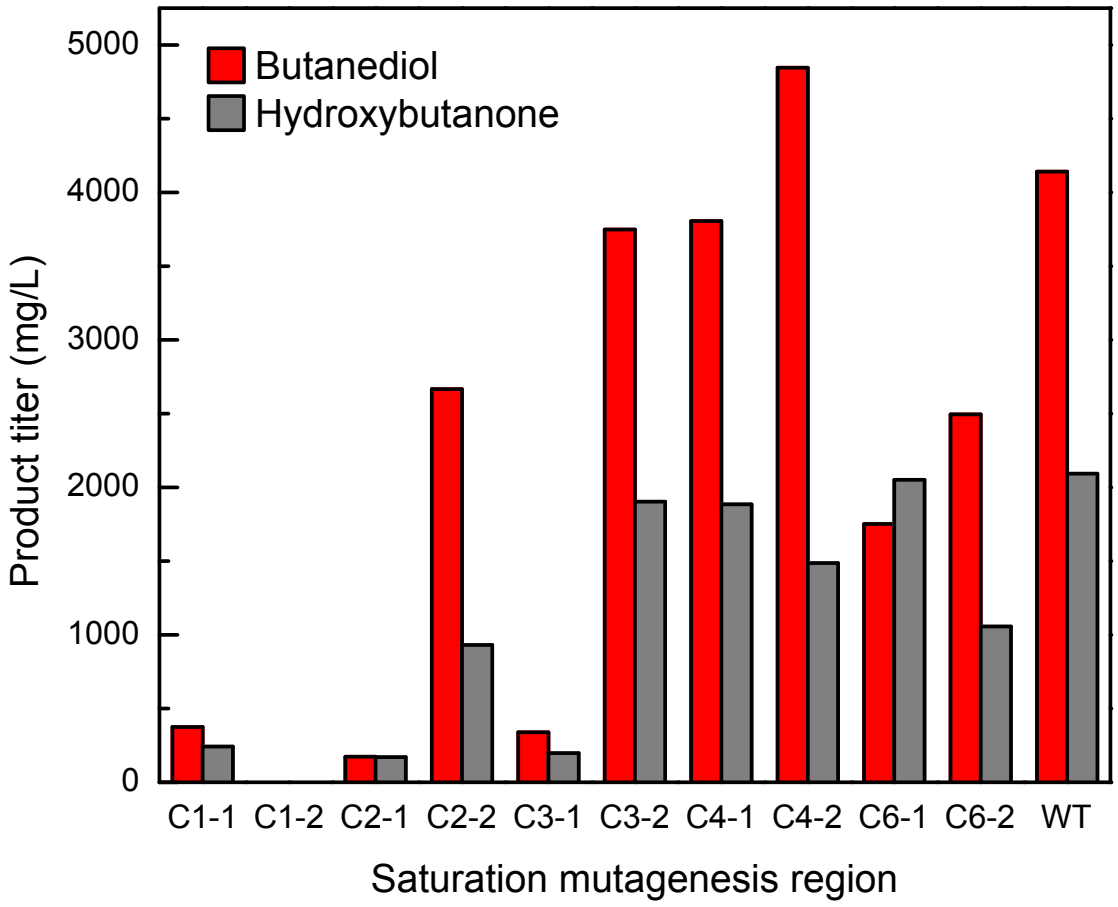


Figure 4.17 ALDH saturation mutagenesis improves butanediol titer and product mixture *Aldh7 mutant libraries were produced with saturation mutagenesis and selected through anaerobic growth. Titers at the endpoint of selection were highly variable and some cultures had lost all productivity. However, culture C4-2 ,containing a library targeting residues 445-447 (Thr-Phe-Thr, see Figure 4.16, blue region), produced 5 g/L butanediol and 1.5 g/L hydroxybutanone. This compares favorable to the wild type, which produced 4 g/L butanediol and 2 g/L hydroxybutanone, and the combined titer of 6.5 g/L is our highest achieved to date. Upon sequencing, clones in this culture were found to contain Thr445 mutated to serine or glycine. In the wild type structure Thr445 is 7.7 Å from the active site Cys279 and forms the opposite wall of the active site. Thr445 makes close contacts with two other loops that further define the active site cavity opposite the catalytic cysteine.*

4.4 Conclusions

Looking to make use of the diverse set of ALDHs characterized in our previous work, we designed modifications of our well-established butanol pathway for the production of commodity chemicals. The diversity of our ALDHs enabled us to develop functional pathways producing 1,3-butanediol. This pathway was optimized in the same manner as our previous work through identification of ADHs to catalyze the final reduction in the pathway. When a side-product of the pathway was identified it was effectively controlled through identification of secondary alcohol dehydrogenases that can redirect off-pathway

carbon back to butanediol production. We further made use of DNA shuffling and saturation mutagenesis to modify the substrate specificity and activity of ALDHs, and this work is ongoing.

Looking forward, this pathway can continue to be optimized in much the same way as the butanol pathway. Although the butanol pathway provided a very strong base for this project, the overall architecture and expression levels needed to optimize butanediol production may be distinct from those for butanol production. Continued directed evolution will be a large area of emphasis, as improvements in substrate specificity and total activity are still desired. This also highlights the generalizability of the anaerobic growth selection we have developed; both protein and genomic directed evolution are effectively pursued. Whole genome mutagenesis of butanediol production strains should also be explored. Many mutations will be similar to those uncovered for butanol production, but mutations related to product tolerance are likely to be different. Additionally, the accessible sequence space of genome mutagenesis is extremely large, and novel mutations are likely to be observed simply through repetition.

4.5 References

1. C. E. Nakamura and G. M. Whited, Metabolic engineering for the microbial production of 1,3-propanediol, *Curr. Opin. Biotechnol.* **2003**, *14*, 454–459.
2. H. Yim, R. Haselbeck, W. Niu, C. Pujol-Baxley, A. Burgard, J. Boldt, J. Khandurina, J. D. Trawick, R. E. Osterhout, R. Stephen, *et al.*, Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol, *Nat. Chem. Biol.* **2011**, *7*, 445–452.
3. J. A. Dietrich, J. L. Fortman and E. J. Steen, *Recombinant host cells for the production of malonate*, (US Patent Office, **2013**).
4. H. Lin, G. Bennett and K.-Y. San, Metabolic engineering of aerobic succinate production systems in Escherichia coli to improve process productivity and achieve the maximum theoretical succinate yield, *Metab. Eng.* **2005**, *7*, 116–127.
5. A. Matsuyama and Y. Kobayashi, Microbial production of optically active 1, 3-butanediol from 4-hydroxy-2-butanone, *Biosci. Biotechnol. Biochem.* **1993**,
6. B. B. Bond-Watts, R. J. Bellerose and M. C. Y. Chang, Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways, *Nat. Chem. Biol.* **2011**, *7*, 222–227.
7. H. J. Atkinson, J. H. Morris, T. E. Ferrin and P. C. Babbitt, Using Sequence Similarity Networks for Visualization of Relationships Across Diverse Protein Superfamilies, *PLoS ONE* **2009**, *4*, e4345.
8. S. Zhao, A. Sakai, X. Zhang, M. W. Vetting, R. Kumar, B. Hillerich, B. San Francisco, J. Solbiati, A. Steves, S. Brown, *et al.*, Prediction and characterization of enzymatic activities guided by sequence similarity and genome neighborhood networks, *Elife* **2014**, *3*, e03275.

9. S. Zhao, R. Kumar, A. Sakai, M. W. Vetting, B. M. Wood, S. Brown, J. B. Bonanno, B. S. Hillerich, R. D. Seidel, P. C. Babbitt, *et al.*, Discovery of new enzymes and metabolic pathways by using structure and genome context, *Nature* **2013**, *502*, 698–702.

10. R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, *et al.*, Pfam: the protein families database, *Nucleic Acids Res.* **2014**, *42*, D222–30.

11. W. Li, L. Jaroszewski and A. Godzik, Clustering of highly homologous sequences to reduce the size of large protein databases, *Bioinformatics* **2001**, *17*, 282–283.

12. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.* **2003**, *13*, 2498–2504.

13. Microbial conversion of glycerol to 1,3-propanediol by an engineered strain of Escherichia coli, *Appl. Env. Microbiol.* **2009**, *75*, 1628–1634.

14. C. Dellomonaco, J. M. Clomburg, E. N. Miller and R. Gonzalez, Engineered reversal of the β-oxidation cycle for the synthesis of fuels and chemicals, *Nature* **2011**, *476*, 355–359.

15. A. A. Ismaiel, C. X. Zhu, G. D. Colby and J. S. Chen, Purification and characterization of a primary-secondary alcohol dehydrogenase from two strains of Clostridium beijerinckii, *J. Bacteriol.* **1993**, *175*, 5097–5105.

16. L. H. Luo, J.-W. Seo, J.-O. Baek, B.-R. Oh, S.-Y. Heo, W.-K. Hong, D.-H. Kim and C. H. Kim, Identification and characterization of the propanediol utilization protein PduP of Lactobacillus reuteri for 3-hydroxypropionic acid production from glycerol, *Appl. Microbiol. Biotechnol.* **2010**, *89*, 697–703.

17. R.-C. Zheng, Z. Ge, Z.-K. Qiu, Y.-S. Wang and Y.-G. Zheng, Asymmetric synthesis of (R)-1,3-butanediol from 4-hydroxy-2-butanone by a newly isolated strain Candida krusei ZJB-09162, *Appl. Microbiol. Biotechnol.* **2012**, *94*, 969–976.

18. R. Machielsen, N. G. H. Leferink, A. Hendriks, S. J. J. Brouns, H.-G. Hennemann, T. Daussmann and J. van der Oost, Laboratory evolution of Pyrococcus furiosus alcohol dehydrogenase to improve the production of (2S,5S)-hexanediol at moderate temperatures, *Extremophiles* **2008**, *12*, 587–594.

19. J. van der Oost, W. G. Voorhorst, S. W. Kengen, A. C. Geerling, V. Wittenhorst, Y. Gueguen and W. M. de Vos, Genetic and biochemical characterization of a short-chain alcohol dehydrogenase from the hyperthermophilic archaeon Pyrococcus furiosus, *Eur. J. Biochem.* **2001**, *268*, 3062–3068.

20. A. Steinbüchel and H. G. Schlegel, NAD-linked L(+)-lactate dehydrogenase from the strict aerobe alcaligenes eutrophus. 2. Kinetic properties and inhibition by oxaloacetate, *Eur. J. Biochem.* **1983**, *130*, 329–334.

21. R. J. Lamed and J. G. Zeikus, Novel NADP-linked alcohol--aldehyde/ketone oxidoreductase in thermophilic ethanologenic bacteria, *Biochem J.* **1981**, *195*, 183–190.

22. T. Oda, K. Oda, H. Yamamoto, A. Matsuyama, M. Ishii, Y. Igarashi and H. Nishihara, Hydrogen-driven asymmetric reduction of hydroxyacetone to (R)-1,2-propanediol by Ralstonia eutropha transformant expressing alcohol dehydrogenase from Kluyveromyces lactis, *Microb. Cell Fact.* **2013**, *12*, 2.

23. S. M. Molinas, S. G. Altabe, F. R. Opperdoes, M. H. Rider, P. A. M. Michels and A. D. Uttaro, The multifunctional isopropyl alcohol dehydrogenase of Phytomonas sp. could be the result of a horizontal gene transfer from a bacterium to the trypanosomatid lineage, *J. Biol. Chem.* **2003**, *278*, 36169–36175.

24. A. Steinbüchel and H. G. Schlegel, A multifunctional fermentative alcohol dehydrogenase from the strict aerobe Alcaligenes eutrophus: purification and properties, *Eur. J. Biochem.* **1984**, *141*, 555–564.

25. R. Sutak, I. Hrdy, P. Dolezal, R. Cabala, M. Sedinová, J. Lewin, K. Harant, M. Müller and J. Tachezy, Secondary alcohol dehydrogenase catalyzes the reduction of exogenous acetone to 2-propanol in Trichomonas vaginalis, *FEBS J.* **2012**, *279*, 2768–2780.

26. D. Leitsch, C. F. Williams, D. Lloyd and M. Duchêne, Unexpected properties of NADP-dependent secondary alcohol dehydrogenase (ADH-1) in Trichomonas vaginalis and other microaerophilic parasites, *Exp. Parasitol.* **2013**, *134*, 374–380.

27. C. T. Hou, R. N. Patel, A. I. Laskin, I. Barist and N. Barnabe, Thermostable NAD-linked secondary alcohol dehydrogenase from propane-grown Pseudomonas fluorescens NRRL B-1244, *Appl. Env. Microbiol.* **1983**, *46*, 98–105.

28. N. García-Quintáns, G. Repizo, M. Martín, C. Magni and P. López, Activation of the diacetyl/acetoin pathway in Lactococcus lactis subsp. lactis bv. diacetylactis CRL264 by acidic growth, *Appl. Env. Microbiol.* **2008**, *74*, 1988–1996.

29. E. González, M. R. Fernández, C. Larroy, L. Solà, M. A. Pericàs, X. Parés and J. A. Biosca, Characterization of a (2R,3R)-2,3-butanediol dehydrogenase as the Saccharomyces cerevisiae YAL060W gene product. Disruption and induction of the gene, *J. Biol. Chem.* **2000**, *275*, 35876–35885.

30. H. Zhang, G. T. Lountos, C. B. Ching and R. Jiang, Engineering of glycerol dehydrogenase for improved activity towards 1, 3-butanediol, *Appl. Microbiol. Biotechnol.* **2010**, *88*, 117–124.

31. K. Yamada Onodera, H. Nariai, Y. Tani and H. Yamamoto, Gene Cloning of Dihydroxyacetone Reductase from a Methylotrophic Yeast, Hansenula ofunaensis, and its Expression in Escherichia coli HB101 for Production of Optically Active 2-Pentanol, *Eng. Life Sci.* **2004**, *4*, 418–425.

32. H. Yamamoto, N. Kawada and A. Matsuyama, Cloning and expression in Escherichia coli of a gene coding for a secondary alcohol dehydrogenase from Candida parapsilosis, *Biosci. Biotechnol. Biochem.* **1999**, *6*, 1051–1055.

33. H. Man, C. Loderer, M. B. Ansorge Schumacher and G. Grogan, Structure of NADH-Dependent Carbonyl Reductase (CPCR2) from Candida parapsilosis Provides Insight into Mutations that Improve Catalytic Properties, *ChemCatChem* **2014**, *6*, 1103–1111.

34. T. Yang, Z. Man, Z. Rao, M. Xu, X. Zhang and Z. Xu, Asymmetric reduction of 4-hydroxy-2-butanone to (R)-1,3-butanediol with absolute stereochemical selectivity by a newly isolated strain of Pichia jadinii, *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 1743–1752.

35. B. Kosjek, W. Stampfer, M. Pogorevc, W. Goessler, K. Faber and W. Kroutil, Purification and characterization of a chemotolerant alcohol dehydrogenase applicable to coupled redox reactions, *Biotechnol. Bioeng.* **2004**, *86*, 55–62.

36. P. Anbarasan, Z. C. Baer, S. Sreekumar, E. Gross, J. B. Binder, H. W. Blanch, D. S. Clark and F. D. Toste, Integration of chemical catalysis with extractive fermentation to produce fuels, *Nature* **2012**, *491*, 235–239.

37. W. Stemmer, Rapid evolution of a protein in vitro by DNA shuffling, *Nature* **1994**,

38. W. P. Stemmer, DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10747–10751.

39. H. Zhao and F. H. Arnold, Optimization of DNA shuffling for high fidelity recombination, *Nucleic Acids Res.* **1997**, *25*, 1307–1308.

40. R. M. Myers, L. S. Lerman and T. Maniatis, A general method for saturation mutagenesis of cloned DNA fragments, *Science* **1985**, *229*, 242–247.

41. M. T. Reetz and J. D. Carballeira, Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes, *Nat. Protoc.* **2007**, *2*, 891–903.

42. K. L. Morley and R. J. Kazlauskas, Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* **2005**, *23*, 231–237.

Appendix 1: *Complete list of plasmids and strains generated*

Table A1.1 Plasmids generated

#	Name	Marker	Origin	Promoter	Assembly	Description
543	pBBR2-mlc	Km	pBBR1	lacUV5	Restriction Enzyme	from pBBR1-MCS2
544	pBBR2-narL	Km	pBBR1	lacUV5	Restriction Enzyme	from pBBR1-MCS2
545	pBBR2-lrp	Km	pBBR1	lacUV5	Restriction Enzyme	from pBBR1-MCS2
603	pBBR2-P(empty).aceE.F.lpd (2)	Km	pBBR1		Restriction Enzyme	pdh from E. coli
604	pBBR2-Ptrc.aceE.F.lpd (2)	Km	pBBR1	trc	Restriction Enzyme	pdh from E. coli
605	pBBR2-Ptac.aceE.F.lpd (2)	Km	pBBR1	double tac	Restriction Enzyme	pdh from E. coli
606	pBBR2-Ppro.aceE.F.lpd (2)	Km	pBBR1	pro	Restriction Enzyme	pdh from E. coli
607	pBBR2-ParcA.aceE.F.lpd (2)	Km	pBBR1	arcA promoter from cydA (E. coli genomic)	Restriction Enzyme	pdh from E. coli
608	pBBR2-Pfnr.aceE.F.lpd (2)	Km	pBBR1	FNR promoter from fdhF (E.coli genomic)	Restriction Enzyme	pdh from E. coli
645	pCOLA-Plac.bmoR-PbmoR.cat	Km, Cm	COLA	lac, bmoR promoter from bmo operon (Thauera butanivorans)	Restriction Enzyme	butanol reporter construct, Cm selection
646	pCOLA-Ptrc.bmoR-PbmoR.cat	Km, Cm	COLA	trc, bmoR promoter from bmo operon (Thauera butanivorans)	Restriction Enzyme	butanol reporter construct, Cm selection
647	pCOLA-Plac.bmoR-PbmoR.cheZ	Km	COLA	lac, bmoR promoter from bmo operon (Thauera butanivorans)	Restriction Enzyme	butanol reporter construct, motility selection

#	Name	Marker	Origin	Promoter	Assembly	Description
648	pET16x-His-PhaA.ZR L88A	Cb	ColE1	T7	Restriction Enzyme	
649	pET16x-His-PhaA.ZR M288A	Cb	ColE1	T7	Restriction Enzyme	
650	pET16x-His-PhaA.ZR L88G	Cb	ColE1	T7	Restriction Enzyme	
651	pET16x-His-PhaA.ZR M288G	Cb	ColE1	T7	Restriction Enzyme	
653	pBMOE1:V2	Cb	ColE1	bmoR (sigma 70), PbmoR	Restriction Enzyme	GFPuv butanol reporter construct from Jeff Dietrich, Keasling Lab
654	pBMOA1:V2	Cb	p15a	bmoR (sigma 70), PbmoR	Restriction Enzyme	GFPuv butanol reporter construct from Jeff Dietrich, Keasling Lab
655	pBMOS1:V2	Cb	pSC101 (possibly not WT)	bmoR (sigma 70), PbmoR	Restriction Enzyme	GFPuv butanol reporter construct from Jeff Dietrich, Keasling Lab
656	pBMOE1:V3	Cb, Tc	ColE1	bmoR (sigma 70), PbmoR	Restriction Enzyme	GFPuv-tetA fusion selection construct from Jeff Dietrich, Keasling Lab
692	pRSF-TdTer.adhE2	Sp	RSF1030	double tac	Restriction Enzyme	
693	pCDF-TdTer.adhE2	Sp	CloDF13	double tac	Restriction Enzyme	
694	pCDF-ccr.adhE2	Sp	CloDF13	double tac	Restriction Enzyme	from 10.1038/290264a0
695	pCDF2	Sp	CloDF13cop2 (200 copy number)		Restriction Enzyme	

#	Name	Marker	Origin	Promoter	Assembly	Description
696	pCDF3	Sp	CloDF13cop3 (70 copy number)		Restriction Enzyme	from 10.1038/290264a0
697	pCWOri-*20eGFP	Cb	ColE1	double tac	Restriction Enzyme	GT(G3S)3RT linker to eGFP (F64L for improved folding at 37C, S65T for 5X amplitude, 488ex 507em), linker may be prone to deletions by
698	pCWOri-TdTer.TbELO1	Cb	ColE1	double tac	Restriction Enzyme	synthetic ELO1 from Trypanosoma brucei
699	pBBR2-cobB	Km	pBBR1	lacUV5	Restriction Enzyme	from pBBR1-MCS2
701	pCWOri-TdTer.TbELO1.eGFP	Cb	ColE1	double tac	Restriction Enzyme	synthetic ELO1 from Trypanosoma brucei, C-terminal eGFP w/ GSAGSAAGSGES linker
704	pCDF2-TdTer.adhE2	Sp	CloDF13cop2 (200 copy number)	double tac	Restriction Enzyme	from 10.1038/290264a0
705	pCDF3-TdTer.adhE2	Sp	CloDF13cop3 (70 copy number)	double tac	Restriction Enzyme	from 10.1038/290264a0
707	pET23a-sbmoR	Cb	ColE1	T7	Restriction Enzyme	synthetic bmoR
710	pCOLA-cat.sacB	Km, Cm, sacB	COLA		Restriction Enzyme	sacB from B. subtilis inserted into FLP-CmR cassette from pKD3
721	pCWOri-TdTer.yTbELO1	Cb	ColE1	double tac	Restriction Enzyme	N-terminal 25AA from E. coli yidC fused to TbELO1
722	pPro18-TdTer.TbELO1	Cb	ColE1	pro	Restriction Enzyme	
723	pTet-TdTer.TbELO1	Cb	ColE1	tet	Restriction Enzyme	

#	Name	Marker	Origin	Promoter	Assembly	Description
727	pPro18-TdTer.yTbELO1	Cb	ColE1	pro	Restriction Enzyme	N-terminal 25AA from E. coli yidC fused to TbELO1
728	pPro18-TdTer.yTbELO1.eGFP	Cb	ColE1	pro	Restriction Enzyme	N-terminal 25AA from E. coli yidC fused to TbELO1, C-terminal eGFP
729	pTet-TdTer.yTbELO1	Cb	ColE1	tet	Restriction Enzyme	N-terminal 25AA from E. coli yidC fused to TbELO1
730	pTet-TdTer.yTbELO1.eGFP	Cb	ColE1	tet	Restriction Enzyme	N-terminal 25AA from E. coli yidC fused to TbELO1, C-terminal eGFP
740	pBca94 Bsb1004	Cm, Km	ColE1	Pc+	Restriction Enzyme	Pc+ promoter from Pseudomonas aeruginosa, Gabe Lopez from the Anderson lab
762	pCWori-TdTer.CbALD.bdhB	Cb	ColE1	double tac	Restriction Enzyme	strep-tagged C. acetobutylicum adhE2 w/ inactivated ALDH domain
882	pCWori-strep.adhE2 (C244A)	Cb	ColE1	double tac	Restriction Enzyme	strep-tagged C. acetobutylicum adhE2 w/ inactivated ADH domain
883	pCWori-strep.adhE2 (H721A, H735A)	Cb	ColE1	double tac	Restriction Enzyme	synthetic aldehyde dehydrogenase from C. beijerinckii
884	pBAD33-CbALD	Cb	ColE1	Pbad	Restriction Enzyme	Cm selection biosensor with exsAD circuit
885	pMD001	Cb, Cm	ColE1	Ppro,Pc,PbmoR,Plac	Gibson	strep-tagged with TEV cleavage site codon optimized adhE2 from C. acetobutylicum
1062	pCWori-strep_TEV_adhE2	Cb	ColE1	double tac	Gibson	strep tagged codon optimized adhE2 from C. acetobutylicum
1063	pCWori-strep_adhE2	Cb	ColE1	double tac	Gibson	

#	Name	Marker	Origin	Promoter	Assembly	Description
1064	pCWori-strep_adhE2 (846)	Cb	ColE1	double tac	Gibson	strep tagged first 846 amino acids of codon optimized adhE2 from C. acetobutylicum
1065	pBu2	Cm	p15A	Pbad, Ptrc, double tac	Gibson	single plasmid Bu2 pathway
1152	pCDF3-TdTer.aldh46	Sp	CloDF13cop3	double tac	Gibson	TdTer, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052, contains unannotated C-terminal adhE2 fragment after aldh46
1221	pCWori-ter.aldh46.ADH	Cb	ColE1	double tac	Gibson	TdTer, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052, codon optimized ADH domain from adhE2 (449-859)
1274	pCWori-strep_aldh46	Cb	ColE1	double tac	Gibson	strep tagged codon optimized aldh from Clostridium beijerinckii NCIMB 8052
1275	pMD140	Cb	ColE1	PbmoR	Gibson	eGFP bmoR repoter in ColE1 plasmid, based on plasmid from Jeff Dietrich
1276	pCWO.trc-tdTer-aldh46.adh	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 and codon optimized ADH domain from adhE2 (449-859) in operon under trc promoter
1277	pCWO.trc-tdTer-aldh4.adh	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Vibrio shilonii AK1 (ZP_01868679) and codon optimized ADH domain from adhE2 (449-859) in operon under trc promoter
1278	pCWO.trc-tdTer-aldh10.adh	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium carboxidivorans P7 (ZP_05393779) and codon optimized ADH domain from adhE2 (449-859) in operon under trc promoter
1318	pT533-phaA.HBD	Cm	p15A	T5	Gibson	phaA.HBD operon under T5 promoter, front end plasmid for 1,3-butanediol production

#	Name	Marker	Origin	Promoter	Assembly	Description
1319	pT533-phaA.phaB	Cm	p15A	T5	Gibson	phaA.phaB operon under T5 promoter, front end plasmid for 1,3-butanediol production
1436	pCWori-aldh3.adh	Cb	ColE1	double tac	Gibson	codon optimized butyraldehyde dehydrogenase from Clostridium saccharoperbutylacetonicum N1-4(HMT) and codon optimized ADH domain from adhE2 (449-859) in operon under double tac promoter, replaced pCWori stutter RBS w/ original Calysta RBS
1437	pCWori-aldh6.adh	Cb	ColE1	double tac	Gibson	codon optimized hypothetical protein CLOBOL_07248 from Clostridium bolteae ATCC BAA-613 and codon optimized ADH domain from adhE2 (449-859) in operon under double tac promoter, replaced pCWori stutter RBS w/ original Calysta RBS
1438	pCWori-aldh7.adh	Cb	ColE1	double tac	Gibson	codon optimized ethanolamine utilization protein EutE from Clostridium botulinum B str. Eklund 17B and codon optimized ADH domain from adhE2 (449-859) in operon under double tac promoter, replaced pCWori stutter RBS w/ original Calysta RBS
1439	pCWori-aldh46.adh	Cb	ColE1	double tac	Gibson	codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 and codon optimized ADH domain from adhE2 (449-859) in operon under double tac promoter, replaced pCWori stutter RBS w/ original Calysta RBS
1522	pRecA	Tc	RepA101ts	Pbad	Gibson	arabinose inducible recA (BW25113) expression for complementing recA1 strains, tetracycline marker, temperature sensitive origin (grow at 30C, cure at 42C)
1585	pLambdaRed	Tc	oriR101ts	Lambda Pr	Gibson	oligo recombineering plasmid, beta expressed from lambda Pr promoter
1647	pCDF3-aldh1	Sp	CloDF13cop3 (70 copy number)	double tac	Gibson	codon optimized coenzyme A acylating aldehyde dehydrogenase Clostridium beijerinckii, contains unannotated C-terminal adhE2 fragment after aldh1, GI:4884855

#	Name	Marker	Origin	Promoter	Assembly	Description
1648	pCDF3-aldh2	Sp	CloDF13cop3 (70 copy number)	double tac	Gibson	codon optimized ethanolamine utilization protein EutE Escherichia coli CFT073, contains unannotated C-terminal adhE2 fragment after aldh2, GI:26250354
1649	pCDF3-aldh3	Sp	CloDF13cop3 (70 copy number)	double tac	Gibson	codon optimized butyraldehyde dehydrogenase from Clostridium saccharoperbutylacetonicum N1-4(HMT), contains unannotated C-terminal adhE2 fragment after aldh3, GI:31075383
1650	pCDF3-aldh4	Sp	CloDF13cop3 (70 copy number)	double tac	Gibson	codon optimized Ethanolamine utilization protein eutE Vibrio shilonii AK1, contains unannotated C-terminal adhE2 fragment after aldh4, GI:149190407
1651	pCDF3-aldh5	Sp	CloDF13cop3 (70 copy number)	double tac	Gibson	codon optimized hypothetical protein RUMGNA_01022 from Ruminococcus gnavus ATCC 29149, contains unannotated C-terminal adhE2 fragment after aldh5, GI:154503198
1652	pCDF3-aldh6	Sp	CloDF13cop3 (70 copy number)	double tac	Gibson	codon optimized hypothetical protein CLOBOL_07248 from Clostridium bolteae ATCC BAA-613, contains unannotated C-terminal adhE2 fragment after aldh6, GI:160942363
1653	pCDF3-aldh7	Sp	CloDF13cop3 (70 copy number)	double tac	Gibson	codon optimized ethanolamine utilization protein EutE from Clostridium botulinum B str. Eklund 17B, contains unannotated C-terminal adhE2 fragment after aldh7, GI:187934965
1654	pCDF3-aldh8	Sp	CloDF13cop3 (70 copy number)	double tac	Gibson	codon optimized coenzyme A acylating aldehyde dehydrogenase from Clostridium saccharobutylicum, contains unannotated C-terminal adhE2 fragment after aldh8, GI:189310620
1655	pCDF3-aldh9	Sp	CloDF13cop3 (70 copy number)	double tac	Gibson	codon optimized ethanolamine utilization protein EutE from Clostridium botulinum E1 str. 'BoNT E Beluga', contains unannotated C-terminal adhE2 fragment after aldh9, GI:251780016

#	Name	Marker	Origin	Promoter	Assembly	Description
1656	pCDF3 -aldh10	Sp	CloDF13cop 3 (70 copy number)	double tac	Gibson	codon optimized Aldehyde Dehydrogenase from Clostridium carboxidivorans P7, contains unannotated C-terminal adhE2 fragment after aldh10, GI:255526882
1657	pCDF3 -aldh11	Sp	CloDF13cop 3 (70 copy number)	double tac	Gibson	codon optimized aldehyde dehydrogenase from Clostridium saccharolyticum WM1, contains unannotated C-terminal adhE2 fragment after aldh11, GI:302386203
1658	pCDF3 -aldh12	Sp	CloDF13cop 3 (70 copy number)	double tac	Gibson	codon optimized aldehyde dehydrogenase from Geobacillus sp. Y4.1MC1, contains unannotated C-terminal adhE2 fragment after aldh12, GI:312110932
1659	pCDF3 -aldh13	Sp	CloDF13cop 3 (70 copy number)	double tac	Gibson	codon optimized Acetaldehyde dehydrogenase (acetylating) from Clostridium sp. DL-VIII, contains unannotated C-terminal adhE2 fragment after aldh13, GI:359413662
1660	pCDF3 -aldh14	Sp	CloDF13cop 3 (70 copy number)	double tac	Gibson	codon optimized hypothetical protein HMPREF9942_01197 from Fusobacterium nucleatum subsp. animalis F0419, contains unannotated C-terminal adhE2 fragment after aldh14, GI:371960349
1661	pCDF3 -aldh15	Sp	CloDF13cop 3 (70 copy number)	double tac	Gibson	codon optimized hypothetical protein HMPREF0402_00608 from Fusobacterium sp. 12_1B, contains unannotated C-terminal adhE2 fragment after aldh15, GI:373496187
1662	pCDF3 -aldh46	Sp	CloDF13cop 3 (70 copy number)	double tac	Gibson	codon optimized aldehyde dehydrogenase from Clostridium beijerinckii NCIMB 8052, contains unannotated C-terminal adhE2 fragment after aldh46, GI:150018649
1866	pCWO. trc-ter-aldh46. adh2	Cb	ColE1	Trc, double tac	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with ADH from Campylobacter fetus subsp. fetus (strain 82-40) under Trc promoter

#	Name	Marker	Origin	Promoter	Assembly	Description
1867	pCWO.trc-ter-aldh46.adh8	Cb	ColE1	Trc, double tac	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with ADH from Olsenella uli (strain ATCC 49627) under Trc promoter
1868	pCWO.trc-ter-aldh46.adh12	Cb	ColE1	Trc, double tac	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with ADH from Clostridium spiroforme DSM 1552 under Trc promoter
1906	pCWO.trc-TdTer-aldh46.adh14	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized ADH from Succinatimonas hippei YIT 12066 under Trc promoter
1907	pCWO.trc-TdTer-aldh46.dhaT2	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized dhaT from Halanaerobium hydrogeniformans (Halanaerobium sp. (strain sapolanicus)) under Trc promoter
1908	pCWO.trc-TdTer-aldh46.dhaT3	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized dhaT from Citrobacter freundii under Trc promoter
1909	pCWO.trc-TdTer-aldh46.dhaT4	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized dhaT from Clostridium novyi (strain NT)i under Trc promoter
1910	pCWO.trc-TdTer-aldh46.dhaT5	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized dhaT from Pelobacter carbinolicus (strain DSM 2380 / Gra Bd 1) under Trc promoter
1911	pCWO.trc-TdTer-aldh46.adh3	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized ADH from Olsenella sp. oral taxon 809 str. F0356 under Trc promoter

#	Name	Marker	Origin	Promoter	Assembly	Description
1912	pCWO.trc-TdTer-aldh46.adh4	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized ADH from Anaerofustis stercorihominis DSM 17244 unter Trc promoter
1913	pCWO.trc-TdTer-aldh46.adh8	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized ADH from Olsenella uli (strain ATCC 49627) under Trc promoter
1914	pCWO.trc-TdTer-aldh46.adh9	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized ADH from Streptococcus gallolyticus (strain ATCC 43143 / F-1867) under Trc promoter
1915	pCWO.trc-TdTer-aldh46.adh13	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized ADH from Tetragenococcus halophilus (strain DSM 20338) under Trc promoter
1916	pCWO.trc-TdTer-aldh46.adh2	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized ADH from Campylobacter fetus subsp. fetus (strain 82-40) under Trc promoter
1917	pCWO.trc-TdTer-aldh46.adh5	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized ADH from Bacillus subtilis (strain 168) under Trc promoter
1918	pCWO.trc-TdTer-aldh46.adh6	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized ADH from Faecalibacterium prausnitzii M21/2 under Trc promoter
1919	pCWO.trc-TdTer-aldh46.adh7	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized ADH from Erysipelotrichaceae bacterium 3_1_53 under Trc promoter

#	Name	Marker	Origin	Promoter	Assembly	Description
1920	pCWO.trc-TdTer-aldh46.adh10	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized ADH from Desulfurispirillum indicum (strain ATCC BAA-1389 / S5) under Trc promoter
1921	pCWO.trc-TdTer-aldh46.adh12	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized ADH from Clostridium spiroforme DSM 1552 under Trc promoter
1922	pCWO.trc-TdTer-aldh46.dhaT6	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized dhaT from Pelotomaculum thermopropionicum (strain DSM 13744 / JCM 10971 / SI) under Trc promoter
1923	pCWO.trc-TdTer-aldh46.dhaT7	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized dhaT from Clostridium perfringens D str. JGS1721 under Trc promoter
1924	pCWO.trc-TdTer-aldh46.dhaT8	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with codon optimized dhaT from Ilyobacter polytropus (strain DSM 2926 / CuHBu1) under Trc promoter
1934	pET23a-HisTEV_aldh3	Cb	ColE1	T7	Gibson	his tagged with TEV linker codon optimized butyraldehyde dehydrogenase from Clostridium saccharoperbutylacetonicum N1-4(HMT)
1935	pET23a-HisTEV_aldh6	Cb	ColE1	T7	Gibson	his tagged with TEV linker codon optimized hypothetical protein CLOBOL_07248 from Clostridium bolteae ATCC BAA-613
1936	pET23a-HisTEV_aldh7	Cb	ColE1	T7	Gibson	his tagged with TEV linker codon optimized ethanolamine utilization protein EutE from Clostridium botulinum B str. Eklund 17B

#	Name	Marker	Origin	Promoter	Assembly	Description
1937	pET23a-HisTEV_aldh46	Cb	ColE1	T7	Gibson	his tagged with TEV linker codon optimized ALDH from Clostridium beijerinckii NCIMB 8052
1938	pET23a-HisTEV_aldh4	Cb	ColE1	T7	Gibson	his tagged with TEV linker codon optimized Ethanolamine utilization protein eutE Vibrio shilonii AK1
1939	pET23a-HisTEV_aldh8	Cb	ColE1	T7	Gibson	his tagged with TEV linker codon optimized ALDH from Clostridium saccharobutylicum
1940	pET23a-HisTEV_aldh10	Cb	ColE1	T7	Gibson	his tagged with TEV linker codon optimized ALDH from Clostridium carboxidivorans
1941	pET23a-HisTEV_aldh11	Cb	ColE1	T7	Gibson	his tagged with TEV linker codon optimized ALDH from Clostridium saccharolyticum
2076	pCWO.trc-TdTer-aldh7.adh2	Cb	ColE1	double tac, trc	Golden Gate	TdTer under double tac promoter, codon optimized ethanolamine utilization protein EutE from Clostridium botulinum B str. Eklund 17B in operon with ADH from Campylobacter fetus subsp. fetus (strain 82-40) under Trc promoter
2078	pCWO.trc-TdTer-RFP.adh2	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, BsaI entry vector (ATTC/TAGA) for ALDH DNA shuffling libraries into operon with ADH from Campylobacter fetus subsp. fetus (strain 82-40) under Trc promoter
2079	pCWO.trc-TdTer-RFP	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, BsaI entry vector (ATTC/TAGT) for aldh.adh pair cloning under trc promoter

#	Name	Marker	Origin	Promoter	Assembly	Description
2080	pT533-phaA	Cm	p15a	T5	Gibson	phaA under T5 promoter, front end plasmid for 4-hydroxy-2-butanone production
2083	pKD46-cas9	Cb	repA101ts	Pbad, S. pyogenes native	Gibson	pKD46 with cas9, from JGI via Gabe Lopez in the Anderson lab, grow at 30C and cure at 42C, derived from strain 2005 with BsaI site removed
2096	pCWO.trc-TdTer-aldh3.adh2	Cb	ColE1	double tac, trc	Golden Gate	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with ADH from Campylobacter fetus subsp. fetus (strain 82-40) under Trc promoter
2097	pCWO.trc-TdTer-aldh3.adh8	Cb	ColE1	double tac, trc	Golden Gate	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with ADH from Olsenella uli (strain ATCC 49627) under Trc promoter
2098	pCWO.trc-TdTer-aldh3.adh12	Cb	ColE1	double tac, trc	Golden Gate	TdTer under double tac promoter, codon optimized ALDH from Clostridium beijerinckii NCIMB 8052 in operon with ADH from Clostridium spiroforme DSM 1552 under Trc promoter
2099	pCWO.trc-TdTer-aldh6.adh2	Cb	ColE1	double tac, trc	Golden Gate	TdTer under double tac promoter, codon optimized hypothetical protein CLOBOL_07248 from Clostridium bolteae ATCC BAA-613 in operon with ADH from Campylobacter fetus subsp. fetus (strain 82-40) under Trc promoter
2100	pCWO.trc-TdTer-aldh6.adh8	Cb	ColE1	double tac, trc	Golden Gate	TdTer under double tac promoter, codon optimized hypothetical protein CLOBOL_07248 from Clostridium bolteae ATCC BAA-613 in operon with ADH from Olsenella uli (strain ATCC 49627 under Trc promoter
2101	pCWO.trc-TdTer-aldh7.adh8	Cb	ColE1	double tac, trc	Golden Gate	TdTer under double tac promoter, codon optimized ethanolamine utilization protein EutE from Clostridium botulinum B str. Eklund 17B in operon with ADH from Olsenella uli (strain ATCC 49627 under Trc promoter

#	Name	Marker	Origin	Promoter	Assembly	Description
2102	pCWO.trc-TdTer-aldh7.adh12	Cb	ColE1	double tac, trc	Golden Gate	TdTer under double tac promoter, codon optimized ethanolamine utilization protein EutE from Clostridium botulinum B str. Eklund 17B in operon with ADH from Clostridium spiroforme DSM 1552 under Trc promoter
2103	pCWO.trc-TdTer-aldh9.adh2	Cb	ColE1	double tac, trc	Golden Gate	TdTer under double tac promoter, codon optimized ethanolamine utilization protein EutE from Clostridium botulinum E1 str. 'BoNT E Beluga' in operon with ADH from Campylobacter fetus subsp. fetus (strain 82-40) under Trc promoter
2104	pCWO.trc-TdTer-aldh9.adh8	Cb	ColE1	double tac, trc	Golden Gate	TdTer under double tac promoter, codon optimized ethanolamine utilization protein EutE from Clostridium botulinum E1 str. 'BoNT E Beluga' in operon with ADH from Olsenella uli (strain ATCC 49627 under Trc promoter
2105	pCWO.trc-TdTer-aldh9.adh12	Cb	ColE1	double tac, trc	Golden Gate	TdTer under double tac promoter, codon optimized ethanolamine utilization protein EutE from Clostridium botulinum E1 str. 'BoNT E Beluga' in operon with ADH from Clostridium spiroforme DSM 1552 under Trc promoter
2106	pCWO.trc-TdTer-aldh14.adh2	Cb	ColE1	double tac, trc	Golden Gate	TdTer under double tac promoter, codon optimized hypothetical protein HMPREF9942_01197 from Fusobacterium nucleatum subsp. animalis F0419 in operon with ADH from Campylobacter fetus subsp. fetus (strain 82-40) under Trc promoter
2107	pCWO.trc-TdTer-aldh14.adh8	Cb	ColE1	double tac, trc	Golden Gate	TdTer under double tac promoter, codon optimized hypothetical protein HMPREF9942_01197 from Fusobacterium nucleatum subsp. animalis F0419 in operon with ADH from Olsenella uli (strain ATCC 49627 under Trc promoter
2110	pMOD-Sp	Cb, Sp	ColE1		Gibson	Tn5 transposon encoding spectinomycin resistance
2111	pMOD-Sp.trc	Cb, Sp	ColE1	trc	Gibson	Tn5 transposon encoding spectinomycin resistance and outward facing Ptrc promoter

#	Name	Marker	Origin	Promoter	Assembly	Description
2112	pMODts-Sp	Tc, Sp	repA101ts		Gibson	Tn5 transposon encoding spectinomycin resistance, temperature sensitive origin (grow at 30C, cure above 37C)
2181	pCWO.trc-ter-RFP.adh8	Cb	ColE1	double tac, trc	Gibson	TdTer under double tac promoter, golden gate RFP cassette (ATTC/TAGA) for cloning ALDHs in operon with ADH from Olsenella uli (strain ATCC 49627) under Trc promoter
2182	pET23a-StrepTEV_aldh7	Cb	ColE1	T7	Gibson	Strep tagged with TEV linker codon optimized ethanolamine utilization protein EutE from Clostridium botulinum B str. Eklund 17B

Table A1.2 Strains generated

#	Organism	Name	Description
881	*E. coli*	MC1.27	DH1 ΔcobB::
1553	*E. coli*	MC1.49	DH1 ΔackA-pta ΔadhE ΔldhA ΔpoxB ΔfrdBC ΔmutS via MC1.24, use for lambda red oligo recombineering, this strain grows very slowly and may be slightly mutagenic
1637	*E. coli*	BW25113-T1R	BW25113 ΔfhuA, base strain for Keio collection
1677	*E. coli*	BL21 (DE3) Star T1R	RNaseE mutation to increase mRNA stability, ΔfhuA, from A. Martin lab
1687	*E. coli*	MC2.16-T1R	BW25113 ΔadhE ΔfhuA, P1 transduced fhuA:Km from 1637 parent to 1320 then recycled Km marker
1688	*E. coli*	MC2.20-T1R	BW25113 ΔackA-pta ΔadhE ΔldhA ΔpoxB ΔfhuA, P1 transduced fhuA:Km from 1637 parent to 1434 then recylced Km marker
1689	*E. coli*	MC2.21-T1R	BW25113 ΔackA-pta ΔadhE ΔldhA ΔfhuA, P1 transduced fhuA:Km from 1637 parent to 1433 then recycled Km marker
1690	*E. coli*	MC2.22-T1R	BW25113 ΔadhE ΔldhA ΔfhuA, P1 transduced fhuA:Km from 1637 parent to 1432 then recycled Km marker
1691	*E. coli*	MC2.24-T1R	BW25113 ΔackA-pta ΔadhE ΔldhA ΔpoxB ΔfrdBC ΔfhuA, P1 transduced fhuA:Km from 1637 parent to 1435 then recycled Km marker
1692	*E. coli*	MC2.48-T1R	BW25113 ΔackA-pta ΔadhE ΔldhA ΔpoxB ΔfrdBC ΔatoB ΔyqeF ΔfhuA, P1 transduced fhuA:Km from 1637 parent to 1560 then recycled Km marker
1707	*E. coli*	MC2.25-T1R	BW25113 ΔackA-pta ΔadhE ΔldhA ΔpoxB ΔfrdBC ΔfhuA ΔtolC

Appendix 2: *Oligonucleotides used for plasmid and strain construction*

Table A2.1 Oligonucleotides used for plasmid and strain construction

Name	Sequence
pBBR2 KmR F1	gcatgcggccgctagcttgcagtgggcttacatgg
pBBR2 KmR R1	gcatgcggccgctagcttgcagtgggcttacatgg
pTrc99a Trc F1	gatcactagtgtttgacagcttatcatcgactgcacg
pTrc99a Trc R1	gatcctcgagctgtttcctgtgtgaaattgttatccgctc
pCWOri Ptac F1	gatcactagtagcttactccccatccccctg
pCWOri Ptac R1	gatcctcgaggatcctgtttcctgtgtgaaattgttatccgc
cydA promoter (arcA) F1	gatcactagttttgcgtttatcttcactctcaagccacg
cydA promoter (arcA) R1	gatcctcgagcatgactccttgctcatcgcatgaag
fdhF promoter (fnr) F1	gatcactagtaatgtctgccgcgtgatgg
fdhF promoter (fnr) R1	gatcctcgagtcatcggtctcgctccagtt
fdhF promoter (fnr) R2	gatcctcgagcggtctcgctccagttaatcaaatcac
pPro18 Ppro F1	gatcgtcgactcagcttttcagccgccgc
pPro18 Ppro R1	gatcactagtttgttatcaacttgttatttgcgttgataaagacaaagc
pKD3 SOE F1	ggatccgagctcgtgtaggctggagctgcttcgaag
pKD3 SOE R1	ggtccatatgaatatcctccttagttcctattccgaag
sacB SOE F1	gatcgcatgccccatcacatatacctgccgttcacta
sacB SOE R1	gcctacacgagctcggatccttatttgttaactgttaattgtccttgttcaaggatgct
pKD3 SOE F2	cagttaacaaataaggatccgagctcgtgtaggctggagctgcttcgaag
bmoR F1 SOE	gagctcaggagaggatccatgtccaagatgcaagagttcgcg
bmoR R1	gatctctagattatgtgccgatccgcgactg
Plac F1	ggcgcgccgcgcaacgcaattaatgtgagttagctc
Plac R1 SOE	gcatcttggacatggatcctctcctgagctcagctgtttcctgtgtgaaattgttatccg
PbmoR F1	gcatgcggccgccccccccaacgacgtccgtc
PbmoR R1 SOE	catatgcctcctactagtttgtgtgttctgctgtcggtagc
cat F1 SOE	cagaacacacaaactagtaggaggcatatgatggagaaaaaaatcactggatataccacc
cat R1	gcatggtaccttacgccccgccctgcc
pKD3 SOE R2	ctgacatgggaattagccatggtccatatg
sacB F2	gatcgcatgcgtgtaggctggagctgcttcgaagttcctatactttctagagaataggaacttccccatcacatatacctgccgttcact
sacB R2 SOE	gagctcggatccttatttgttaactgttaattgtccttgttcaaggatgct
pKD3 F3 SOE	caaggacaattaacagttaacaaataaggatccgagctcttacgccccgccctgccac
cat F2 SOE	cagaacacacaaactagtaggaggcatatggagaaaaaaatcactggatataccacc
cheZ F1	gatccatatgatgcaaccatcaatcaaacctgctg
cheZ R1	gatcggtaccttaaaatccaagactatccaacaaatcgtccacc
bmoR-FL F1	gccatgtccaagatgcaagagttcgcg
bmoR-FL R1	gatcctcgagttatgtgccgatccgcgactg
bmoR-N R1	gatcctcgagtgctgcgtcgcccgtgtc
phaA L88G QCF1	cgcctggggcatgaaccagggttgcggctcgggcctgc
phaA L88G QCR1	gcaggcccgagccgcaaccctggttcatgccccaggcg
phaA L88A QCF1	cgcctggggcatgaaccaggcttgcggctcgggcctgc
phaA L88A QCR1	gcaggcccgagccgcaagcctggttcatgccccaggcg
phaA M288G QCF1	gtcggcgtcgatcccaaggtcggcggcaccggcccgatccc
phaA M288G QCR1	gggatcgggccggtgccgccgaccttgggatcgacgccgac
phaA M288A QCF1	gtcggcgtcgatcccaaggtcgcgggcaccggcccgatccc
phaA M288A QCR1	gggatcgggccggtgcccgcgaccttgggatcgacgccgac
bmoR-FL R2	gcatgcggccgctgtgccgatccgcgactggc
bmoR-N R2	gcatgcggccgctgctgcgtcgcccgtgt
Ptrc F1	gatcgagctcctgtttcctgtgtgaaattgttatccgctc
Ptrc R1	gcatggcgcgcccgactgcacggtgcacc
Ptac F1	gatcgagctcgatcctgtttcctgtgtgaaattgttatccg
Ptac R1	gcatggcgcgccagcttactccccatcccccctg
pKD3 F4 SOE	ggacaattaacagttaacaaataaggatccgagctcttacgccccgccctgcc
pKD3 R4	gatccatatgatggagaaaaaaatcactggatataccaccgttg
sacB F4	gcatgcatgcatacctgccgttcactattatttagtgaaatgag
sacB R4 SOE	gagctcggatccttatttgttaactgttaattgtccttgttcaaggatgc
pKD3 R5	gatccatatgggcgcgcctacctgtgac
aadA F1	gatcaagcttttatttgccgactaccttggtgatctcg
aadA R1 SOE	ctgtgggtaactttgtatgtgtccgc
ter-adhE2 CDF F1	gatctctagatcactgcccgctttccagtc
ter-adhE2 CDF R1	gcatcctgaggttaaaaagatttgatataaatgtctttcagctcagagatcagc
Clo DF13-cop2 QCF1	gccagttaccacggttaagcagtttcccaactgacttaaccttcgatcaaacc

Name	Sequence
Clo DF13-cop2 QCR1	ggtttgatcgaaggttaagtcagttgggaaactgcttaaccgtggtaactggc
Clo DF13-cop3 QCF1	gatttggttgctgtgctctgcgaaaaccagttaccacggttaagcagttcc
Clo DF13-cop3 QCR1	ggaactgcttaaccgtggtaactggttttcgcagagcacagcaaccaaatc
sbmoR R100	ctcgagctacgtaccgatgc
sbmoR F100	tatgggcgccatgagcaaaatgc
CloDF13 F1	aatagctagctcactcggtcgctacg
CloDF13cop2 R1 SOEQ	aaactgcttaaccgtggtaactggc
CloDF13cop2 F1 SOEQ	tcccaactgacttaaccttcgatcaaaccac
CloDF13 R1	gaaatctagagcggttcagtagaaaagatcaaagg
CloDF13cop3 R1 SOEQ	ttttcgcagagcacagcaaccaaatc
CloDF13cop3 F1 SOEQ	accagttaccacggttaagcagttcc
ter-adhE2 CDF F2	gatctctagatcactgcccgctttccagtcgg
ter-adhE2 CDF R2	gcatcctgaggttaaaaagatttgatataaatgtctttcagctcagagatcagcggg
sbmoR F101	gccatgagcaaaatgcaggagttcg
aadA CloDF F1	gatacaaagcttttatttgccgactaccttggtgatctcg
aadA CloDF R1	gatacacctgagggatcaaaggatcttcttgagatcctttttttctgcg
PvuII CmR F1	ctgaacggtctggttataggtacattgagc
PvuII CmR R1 SOE	ggcaggtatatgtgatgggggatccggcgcgcctacctgtgac
sacB F5 SOE	ggatcccccatcacatatacctgccgttcactattatttagtg
sacB R5 SOE	cctattctctagaaagtataggaacttcgagctcttatttgttaactgttaattgtccttgttcaaggatgctg
pKD3 F5 SOE	gttaacaaataagagctcgaagttcctatactttctagagaataggaacttcggaatagg
pKD3 R6	atttttgcggccgcaagatccgc
eGFP F1	tgtgggtaccggcggtgg
eGFP R1 SOEQ	cattgaacaccataggtcagagtagtgacaagtgttggccatgg
eGFP F1 SOEQ	cttgtcactactctgacctatggtgttcaatgcttttcccgttatccg
eGFP R1	cgactctagattatttgtagagctcatccatgcc
CloDF13cop2 R2 SOEQ	gttaagtcagttgggaaactgcttaaccgtggtaactggc
CloDF13cop2 F2 SOEQ	cacggttaagcagtttcccaactgacttaaccttcgatcaaacc
CloDF13cop3 R2 SOEQ	gtggtaactggttttcgcagagcacagcaaccaaatc
CloDF13cop3 F2 SOEQ	gtgctctgcgaaaaccagttaccacggttaagcagttcc
sTbELO1 F100	gatagaattcaagaaggagatataccatgttcttcacgccgccgc
sTbELO1 R100	gataggtaccctacttaactttcttctccacggaaccgttg
sTbELO1-GFP R101 SOE	agattcgccagaaccagcagcggagccagcggatcccttaactttcttctccacggaaccgttgc
eGFP-link F2 SOE	ggatccgctggctccgctgctggttctggcgaatctagtaaaggagaagaacttttcactggagttgtcc
eGFP R2	gatcggtaccttatttgtagagctcatccatgcc
cobB KF1	atgctgtcgcgtcgggggattgtgtaggctggagctgcttc
cobB KR1	tcaggcaatgcttcccgcttttaatcggaattagccatggtccatatgaatatcctccttag
cobB F1	gataggtaccatgctgtcgcgtcgggg
cobB R1	gatcgaattctcaggcaatgcttcccgcttttaatc
CloDF F2	gactcctgagggatcaaaggatcttcttgagatccttttttctgcg
sbmoR F102	gtacggtaccggcgccatgagcaaaatgcaggagttcg
cobB KVF1	tgagcgttatctggatttattagcggtttgtc
cobB KVR1	atatgggttatcaggctttatgtttgtatatcgataatagctttg
sacB F6	gatcctgcagcccatcacatatacctgccgttcact
pKD3 F6	gatcgaattcgtgtaggctggagctgcttcg
pKD3 R7	gatcctgcaggcgcgcctacctgtgacg
yidC ELO signal F2	gtcatcgctttgctgttcgtgtctttcatgatctggcaagcctgggagcagatgttcttcacgccgccgc
yidC ELO signal F1	gatagaattcaagaaggagatataccatggattcgcaacgcaatcttttagtcatcgctttgctgttcgtgtc
Ter.ELO 10K RBS F1	gatagctagcaagtcgaggacctacgtcgaacgggagaggagcgacatgatcgtcaagccaatggtgcg
Ter.ELO R1	gataaagcttggtaccctacttaactttcttctccacggaaccgttg
galK cat.sacB KF1	gtttgcgcgcagtcagcgatatccattttcgcgaatccggagtgtaagaacgtgtaggctggagctgcttc
galK cat.sacB KR1	ttcatattgttcagcgacagcttgctgtacggcaggcaccagctcttccgccaattgagatctgccatatgaatatcctccttag
Ter.ELO.eGFP R1	cgataagcttggtaccttatttgtagagctcatcc
galK KVF1	gcctgcaaaaagaatattttgccgaacagaaatc
galK KVR1	cgggcaaggtttatcaactgtggc
ampR F1	aggaattcgatatcaagcttatcgataccgtcgacctcgaggggggggcccgcggccgcaggtggcacttttcggggaaatg
ampR SOE R1	ggattttggtcatgactatggcttggattctcaccaataaaaaacgccc
ColE1 SOE F1	gagaatccaagccatagtcatgaccaaaatcccttaacgtgagttttcg
ColE1 R1	gcggcggctgaaaagctgaggcgcgcccctaggcgttcggctgc

Name	Sequence
prpR.Ppro F1	ctcgccgcagccgaacgcctaggggcgcgcctcagcttttcagccgccgc
prpR.Ppro SOE R1	taccttacctgtcgaaccgggatccttgttatcaacttgttatttgcgttgataaagacaaagc
exsD SOE F1	ggatcccggttcgacaggtaaggtaatggagcaggaagacgataagcagtac
exsD R1	ccaatcccgcatcggtacgtagcctcaggtcagctctgccagtagaagtgatcc
Plac R1	ccgcagcacctactatctgtggggtaccgcgcaacgcaattaatgtgagttagctc
Plac SOE F1	gctcatagatctctgcttatataagtgggatagctgtttcctgtgtgaaattgttatccg
sbmoR SOE R1	cagctatcccacttatataagcagagatctatgagcaaaatgcaggagttcgcg
sbmoR F1	cacttctactggcagagctgacctgaggctacgtaccgatgcgggattgg
PbmoR F1	actcacattaattgcgttgcgcggtaccccacagatagtaggtgctgcgg
PbmoR SOE R1	ctccttgcatactagttctcgccctccttatctattctgcg
exsA SOE F1	gagggcgagaactagtatgcaaggagccaaatctcttggc
exsA R1	tgagtggcagggcggggcgtaatctagatcagttatttttagcccggcattcgtc
Pc F1	atgattacgccaagctcggaattaaccctcactaaagggaacaaaagctggcggccgcgatcttccaggtgcgccagg
Pc SOE R1	gaccgaatttagtatgggtctccgaccctgcgaactcggcaag
cat SOE F1	ggtcggagacccatactaaattcggtcatatggagaaaaaaatcactggatataccaccgttg
cat R1	atgccgggctaaaaataactgatctagattacgccccgccctgcc
cobB KF2	catctaaccgattaaacaacagaggttgcttgcactacgcttatcaggcctgcaagggga
cobB KR2	tccccttgcaggcctgataagcgtagtgcaagcaacctctgttgtttaatcggttagatg
galK KF2	atgagtctgaaagaaaaaacacaatctctgtttgcgtgtaggctggagctgcttc
galK KR2	tcagcactgtcctgctccttgtgccaattgagatctgccatatgaatatcctccttag
galK KF3	atccattttcgcgaatccggagtgtaagaaacgaaactcccgcactggcacccgatggtc
galK KR3	gaccatcgggtgccagtgcgggagtttcgtttcttacactccggattcgcgaaaatggat
lac.PDH KF1	gatactcgagccattttcgcgaatccggagtgtaagaagcgcaacgcaattaatgtgagttagc
lac.PDH R1	gatagctagcagctgtttcctgtgtgaaattgttatccg
lac.PDH F2	gtaatctagaggtttaaggagatatcctatgagtactgaaatc
lac.PDH KR2	gataccgcggcgggtgccagtgcgggagtttcgtttacttcttcttcgctttcgggttcgg
lac.sbmoR NF3	aggaaacagctatcccacttatataagcagagatctatgagcaaaatgcaggagttcgcg
lac.sbmoR NF2	actttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatcccacttatataagcagag
lac.sbmoR NF1	gataggtaccgcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttg
pBBR2 KmR F2	gataagtacttagcttgcagtgggcttacatgg
pBBR2 KmR R2	gataagtacttcagaagaactcgtcaagaaggcgatag
exsD R2	ccgcagcacctactatctgtggggtaccccctcaggtcagctctgccagtagaagtgatcc
cobB KF3	gccttcctacatctaaccgattaaacaacagaggttgctcgtgtaggctggagctgcttc
cobB KR3	attaattgcgtccccttgcaggcctgataagcgtagtgcaccaattgagatctgccatatgaatatcctccttag
pBAD/BT KmR SLIC F1	attacgccccgccctgccactcatcgcagtactagttagcttgcagtgggcttacatgg
pBAD/BT KmR SLIC R1	gccggacgcatcgtggccggcatcaccggcgcatgctcagaagaactcgtcaagaaggcg
CbALD F1	gatagaattcctaccacatcccggagcaccatatgaacaaagataccctgatcccgac
CbALD SOE R1	gcttcctcctagtaggccggtacctcagccggccagtacgc
bdhB SOE F1	ggtaccggcctactaggaggaagcatggtggacttcgagtattctatcccaac
bdhB R1	gataaagcttttaaacggatttcttgaaaatttgcagtacctcg
Pc R2	gatacatatgaccgaatttagtatgggtctccgaccctgcgaactcggcaag
sbmoR R2	gataggtaccagatctatgagcaaaatgcaggagttcgcg
cobB KF4	tgcgtggtgcggccttcctacatctaaccgattaaacaacagaggttgctcgtgtaggctggagctgcttc
cobB KR4	cgcaaattcaattaattgcgtcccctgcaggcctgataagcgtagtgcaccaattgagatctgccatatgaatatcctccttag
galK KF4	gtttgcgcgcagtcagcgatatccattttcgcgaatccggagtgtaagaacgtgtaggctggagctgcttc
galK KR4	agttaacagtcggtacggctgaccatcgggtgccagtgcgggagtttcgtccaattgagatctgccatatgaatatcctccttag
lac.PDH KF3	gatactcgaggtttgcgcgcagtcagcgatatccattttcgcgaatccggagtgtaagaagcgcaacgcaattaatgtgagttagc
lac.PDH KR3	gataccgcggagttaacagtcggtacggctgaccatcgggtgccagtgcgggagtttcgtttacttcttcttcgctttcgggttcgg
Pc F2 SOE	ccccatggatttaacgagcaaggtcaacc
Pc R3 SOE	ggttgaccttgctcgttaaatccatgggg
Pc F3	gatagcggccgcgatcttccagg
sbmoR SLIC R3	agccgcagcacctactatctgtggggtaccagatctatgagcaaaatgcaggagttcgcg
sbmoR SLIC F2	tcacttctactggcagagctgacctgagggctacgtaccgatgcgggattgg
pBAD/BT KmR SLIC R2	atagtaagccagtatacactccgctagcgcggccgctcagaagaactcgtcaagaaggcg
pBAD/BT KmR GF1	gccccgccctgccactcatcgcagtactggcaccaataactgccttaaaaaaattacgcc
pBAD/BT KmR GR1	gtcagtgacggcgcgccgcggccgctcagaagaactcgtcaagaaggcgatag
pBAD/BT p15A GF1	ttcttctgagcggccgcggcgccgtcactgactcgctacgctcg
pBAD/BT p15A GR1	acgcatcgtggccggcatcaccggcgccaatattttatctgattaataagatgatcttcttgagatcgttttggtc
sbmoR GF1	atcacttctactggcagagctgacctgaggctacgtaccgatgcgggattgg
sbmoR GR1	atcccacttatataagcagagatctatgagcaaaatgcaggagttcgcg
Plac GF1	agatctctgcttatataagtgggatagctgtttcctgtgtgaaattgttatccg
Plac GR1	cagcacctactatctgtggggtaccgcgcaacgcaattaatgtgagttagc
pMD000 cassette GF1	ggtaccccacagatagtaggtgctgcgg
pMD000 cassette	gactgcttgccgagttcgcag

Name	Sequence
GR1	
Pc GF1	gaccctgcgaactcggcaagcagtcggttgaccttgctcgttaaatccatgggg
Pc GR1	cccgaaaagtgccacctgcggccgcgatcttccaggtgcgccagg
pBAD/BT KmR GF2	gccccgccctgccactcatcgcagtacttagcttgcagtgggcttacatgg
FbFP GF100	ggagacccatactaaattcggtcatatggcgtcgttccagtcgttc
FbFP GR100	ccgggctaaaaataactgatctagactcgagcagcttttcatattccttctgcttg
strep adhE2 (ALDH*) GF1	ccatcgatgcttaggaggtcatatgatggcaagctggagccacccgcagttcgaaaagggtgcaggtatgaaagtcacgaaccag aaggaactgaag
strep adhE2 (ALDH*) GR1	tttgacagcttatcatcgataagcttggtaccttaaaaagatttgatataaatgtctttcagctcagagatcagc
sbmoR GF2	caggcacaggatcacttctactggcagagctgacctgaggctacgtaccgatgcgggattgg
sbmoR GR2	atttcacacaggaaacagctatcccacttatataagcagagatctatgagcaaaatgcaggagttcgcg
Plac GF2	ttttgctcatagatctctgcttatataagtgggatagctgtttcctgtgtgaaattgttatccg
Plac GR2	gacaggagcatgagcagccgcagcacctactatctgtggggtaccgcgcaacgcaattaatgtgagttagc
ALDH GF1	gttgaacgtttcgatcgtatttaagaattctcataactaatttagaggaaggtcctactatgaaagtcacgaaccagaag
ALDH GR1	ggcatttatatatcctccttaatgggtaccttagccgtatttgaagtaaattttc
ADH GF1	tttacttcaaatacggctaaggtacccattaaggaggatatataaatgcctgcgtttcgcgctgaaag
ADH GR1	catgtttgacagcttatcatcgataagcttttaaaaagatttgatataaatgtctttcag
yfiQ KF1	atgagtcagcgaggactggaagcactactgcgaccaaaatcgatagcggtcgtgtaggctggagctgcttc
yfiQ KR1	tcatgattcctcgcgctgggcaagatttagcgtaagcccaacgatcccctccaattgagatctgccatatgaatatcctccttag
yfiQ KVF1	aggccaatattgtactgccgag
yfiQ KVR1	gaaatcgacatcatcaactgattg
sbmoR GR3	ggagcatgagcagccgcagcacctactatctgtggggtaccatgagcaaaatgcaggagttc
ampR pBBR GF1	aagcgcagatgacccgcgaccagaccacgtttgcggccgcctatggcttggattctcacc
ampR pBBR GR1	ccaacgcgcggggagaggcggtttgcgtattgggcgcatgcaggtggcacttttcggg
ampR pBBR GR2	ggcatcagcaccttgtcgccttgcgtataatatttgcccatggaggtggcacttttcggg
Plac GF3	tttccagacgcgcgaactcctgcattttgctcatggtaccgtatcccgtagaagtgatgaagctgtttcctgtgtgaaattg
Pc GO1	ccatactaaattcggtcatatggagaaaaaaatcactggatataccaccgttgatatatc
Pc GO2	tatgaccgaatttagtatgggtctccgaccctgcgaactcggcaagcagtcggttgacct
Pc GO3	cgctgagctttaggaggcgcccccatggatttaacgagcaaggtcaaccgactgcttgcc
Pc GO4	gcgcctcctaaagctcagcgcatgctagcaccgggctatgcatcgcttttgtcactgaga
Pc GO5	aggtgcgccagggcgaatcgcaggcgagcgcccgggaaggagaggtcaacgccaccatcg
Pc GO6	cgattcgccctggcgcacctggaagatcgcggccgcaggtggcacttttcggggaaatgt
ColE1 GF1	tgagtgagctaactcacattaattgcgttgcgcggtacctcatgaccaaaatcccttaac
ColE1 GR1	acagtactgcgatgagtggcagggcggggcgtaacatatgcgttcggctgcggcg
CmR GF1	tgagtgagctgataccgctcgccgcagccgaacgcatatgttacgccccgccctgc
CmR GR1	gaaaggcgagatcaccaaggtagtcggcaaataaactagttacctgtgacggaagatcac
aadA GF1	atttattctgcgaagtgatcttccgtcacaggtaactagtttatttgccgactaccttgg
aadA GR1	ggtttttgcttttggaggggcagaaagatgaatgactgtcccgccgcagtctcacg
rnpB terminator GO1	cgtaaaaacccgcttcggcgggtttttgcttttggaggggcagaaagatgaatgactgtc
rnpB terminator GO2	cccctccaaaagcaaaaacccgccgaagcgggtttttacgtaaatcaggtgaaactgacc
Plac.sbmoR GF1	aagcgggtttttacgtaaatcaggtgaaactgaccgactcgagctacgtaccgatgcggg
Plac.sbmoR GR1	cagtggaacgaaaactcacgttaagggattttggtcatgaggtaccgcgcaacgc
rnpB terminator GO3	ggggcagaaagatgaatgactgtcggcgcgcccctaggcgttcggctgcggcgagcggta
rnpB terminator GO4	cgcgccgacagtcattcatctttctgcccctccaaaagcaaaaacccgccgaagcgggtt
rnpB terminator GO5	gatcggtcagtttcacctgatttacgtaaaaacccgcttcggcgggtttttgcttttgga
rnpB terminator GO6	tttacgtaaatcaggtgaaactgaccgatcagcttttcagccgccgccagaacgtcgtcc
ilv.rrnD terminator GO1	gggtcttagttcgttaaggcttgatctctatcagctctgccagtagaagtgatcctgtgc
ilv.rrnD terminator GO2	gccttaacgaactaagacccccgcaccgaaaggtccgggggttttttttgaccttaaaaa
ilv.rrnD terminator GO3	tgaacactctcccgcctcaggtgtctgctcctcggttatgtttttaaggtcaaaaaaaac
ilv.rrnD terminator GO4	ctgaggcgggagagtgttcaccgacaaacaacagataaaacaaaaggcccagtcttccga
ilv.rrnD terminator GO5	taggggaactgccagacatcaaataaaacaaaaggctcagtcggaagactgggccttttg
ilv.rrnD terminator GO6	gatgtctggcagttcccctacgtaccgatgcgggattggctccaatg
B1006.rrnB terminator GO1	aaaaaaaccccgccctgtcaggggcgggggttttttttttcagttatttttagcccggcat
B1006.rrnB terminator GO2	tgacagggcgggggttttttttttctagacaggagagcgttcaccgacaaacaacagataaa
B1006.rrnB terminator GO3	aataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggt
B1006.rrnB terminator GO4	actgagcctttcgttttatttgatgcctggcagttcccttacgccccgccctgccactca
pBAD/BT KmR GF3	ccttaaaaaaattacgccccgccctgccactcatcgcagtacttagcttgcagtgggctt

Name	Sequence
pBAD/BT KmR GR3	aaaaaaaaccccgccctgtcaggggcggggttttttttttcagaagaactcgtcaagaag
pBAD/BT p15A GF3	aaaaccccgcccctgacagggcggggttttttttgcggccgcgtcactgactcgctacgc
pBAD/BT p15A GR3	ccggacgcatcgtggccggcatcaccggcaatattttatctgattaataagatgatcttc
strep.tev.adhE2 GO1	tcgaactgcgggtggctccagcttgccatatgacctcctaagcatcgatggatcctgttt
strep.tev.adhE2 GO2	tatggcaagctggagccacccgcagttcgaaaagggtgcaggtgagaatctctacttcca
strep.tev.adhE2 GF1	ggtgcaggtgagaatctctacttccagggtaccggcgccatgaaagtcacgaaccagaag
strep.tev.adhE2 GR1	gggctcagatctgctcatg
ilv.rrnD terminator GO7	cgaggagcagacacctgaggctacgtaccgatgcgggattggctccaatgtgccggacgg
100.5 backbone GF1	tgagtgagctaactcacattaattgcgttgcgcggtacctcatgaccaaaatcccttaac
100.5 backbone GR1	ccgtccggcacattggagccaatcccgcatcggtacgtagctcgagtcggtcagtttcac
ilv.rrnD terminator GO8	tgtttgtcggtgaacactctcccgcctcaggtgtctgctcctcggttatgtttttaaggt
ilv.rrnD terminator GO9	gctcagtcggaagactgggccttttgttttatctgttgtttgtcggtgaacactctcccgcctcaggtgtctgctcctcggttatgtttt
ilv.rrnD terminator GO10	aaaacaaaaggcccagtcttccgactgagccttttgttttatttgatgtctggcagttcccctacgtaccgatgcgggattggctccaat
pBu2 GF1	ttcaaaaatcgttaacccggggatcctctagagtcgacctgcagtcactgcccgctttcc
pBu2 GR1	tctcatccgccaaaacagccaagcttgcatgcttaaaaagatttgatataaatgtctttc
pBu2 GR2	cgatcgaaacgttcaacttc
pBu2 GF2	gtctaatggtttcgacgttg
rnpB terminator GO7	aacccgcttcggcgggtttttgcttttggaggggcagaaagatgaatgactgtcggcgcgcccctaggcgttcggctgcggcgagcg gta
rnpB terminator GO8	cctccaaaagcaaaaacccgccgaagcgggtttttacgtaaatcaggtgaaactgaccgatcagcttttcagccgccgccagaac gtcgt
B1006.rrnB terminator GO5	gtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctcctgtctagatcagttatttttagcccggcattcgtcct
B1006.rrnB terminator GO6	aaacgaaaggcccagtctttcgactgagcctttcgttttatttgatgcctggcagttcccttacgccccgccctgccactcatcgcagta
PbmoR GF1	catatttgaatgtatttagaaaaataaacaaatagctagcccacagatagtaggtgctgc
PbmoR GR1	aaagttcttctcctttactcataattctagcccgcctctagattgtgtgttctgctgtcg
eGFP GF1	cagaacacacaatctagaggcgggctagaattatgagtaaaggagaagaacttttcactg
eGFP GR1	ccccgccctgtcaggggcgggggttttttttgcatgcttatttgtagagctcatccatgc
B1006 terminator GO1	aaaaccccgcccctgacagggcgggggttttttttcactcggtcgctacgctccgggcgt
sbmoR.1 GF1	gacgcggggccgggcgaggcggacgtgcatgtgctgcaggacgcggcgggacgcgcag
sbmoR.1 GR1	cgtcccgccgcgtcctgcagcacatgcacgtccgcctcgcccggccccgcgtccgctg
strep.adhE2 GF1	aggatccatcgatgcttaggaggtcatatggcaagctggagccacccgcagttcgaaaagggtgcaggtatgaaagtcacgaacc agaag
ilv.rrnD terminator GF1	tcgaaggcttcaaaaatcgttaacccggggatcctagagatcaagccttaacgaactaag
ilv.rrnD terminator GR1	acaggtttcccgactggaaagcgggcagtgactgcaggtcgacgggaactgccagacatc
B1006.rrnB terminator GO7	tcctgtctagaaaaaaaaaccccgccctgtcaggggcgggggttttttttttcagttatttttagcccggcattcgtccttcccctgccgc
B1006.rrnB terminator GO8	cccgcccctgacagggcgggggttttttttctagacaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtctttc gac
B1006.rrnB terminator GO9	agggcggggcgtaagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgt cg
B1006.rrnB terminator GO10	gttttatttgatgcctggcagttcccttacgccccgccctgccactcatcgcagtactgttgtaat
M13.rrnB terminator GO1	gagattttcaacatgaaaaaattattatttctagacaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtctttc gac
M13.rrnB terminator GO2	gtgaacgctctcctgtctagaaataataatttttttcatgttgaaaatctccaaaaaaaaaggctccaaaaggagcctttaattgtatcgg
M13.rrnB terminator GO3	aaagcaagctgataaaccgatacaattaaaggctccttttggagccttttttttg
M13.rrnB terminator GO4	aaaaggagcctttaattgtatcggtttatcagcttgcttttcagttatttttagcccggcattcgtccttcccctgccgc
adhE2 846 GR1	tgctcatgtttgacagcttatcatcgataagcttggtaccttacgggtaacgcgggtttg
adhE2 GF1	aactacgaggcagaagttgaac
adhE2_bdhB GR1	gattttatctttaccgaaaaagatacgggttgggatagaataccacagcatgttttcacg
adhE2_bdhB GF1	tattctatcccaacccgtatc
term 3 GF100	ggctgcacgccgagcc
term 3 GR100	gtcggcagaatgcttaatg
ALDH 400 GR1	tcatgtttgacagcttatcatcgataagcttggtaccttactgggagcttggcatgttga
ALDH 430 GR1	tcatgtttgacagcttatcatcgataagcttggtaccttagttttgggaaacgctgttgc
ALDH 463 GR1	tcatgtttgacagcttatcatcgataagcttggtaccttagccgtatttgaagtaaattt

Name	Sequence
pMD100.6 backbone GF1	ttttttttcactcggtcgctacgctccgggcgtaagcttgacagtcattcatctttctgc
pMD100.6 backbone GR1	gctagctatttgtttatttttctaaatacattc
adhE2_bdhB GR2	tctgctcatgtttgacagcttatcatcgataagcttggtaccccacagcatgttttcacg
exsA GF1	cgctaccgacagcagaac
exsA GR1	catgttgaaaatctccaaaaaaaaaggctccaaaaggagcctttaattgtatcggtttatcagcttgcttttcagttatttttagcccgg
cat GF1	acaaacaacagataaaacgaaaggcccagtctttcgactgagcctttcgttttatttgatgcctggcagttcccttacgccccgccctgc
cat GR1	caaggtcaaccgactgc
term 3 GBF1	ctttttttttggagattttcaacatgaaaaaattattatttctagacaggagagcgttcaccgacaaacaacagataaaacgaaaggccc
term 3 GBR1	gggcctttcgttttatctgttgtttgtcggtgaacgctctcctgtctagaaataataatttttttcatgttgaaaatctccaaaaaaaaag
adhE2_bdhB GF2	atcaaatctgttgcagaacgccgtgaaaacatgctgtggtattctatcccaacccgtatc
B1006 terminator GBR1	acgcccggagcgtagcgaccgagtgaaaaaaaaaccccgccctgtcaggggcggggtttt
cat GF2	aggacgaatgccgggctaaaaataactgatctagagactgagcctttcgttttatttgatgcctggcagttcccttacgccccgccctgc
cat GF3	aggacgaatgccgggctaaaaataactgatctagaaacaacagataaaacgaaaggcccagtctttcgactgagcctttcgttttatttg
cat GF4	aggacgaatgccgggctaaaaataactgatctagacaggagagcgttcaccgacaaacaacagataaaacgaaaggc
PbmoR GR2	ccggagcgtagcgaccgagtgaaattctagcccgcctctagattgtgttctgctgtcg
eGFP GF2	tggagcattcgaagccgctaccgacagcagaacacacaatctagaggcgggctagaattatgagtaaaggagaagaacttttcactg
eGFP GR2	cgagtgaaattctagcccgcctaaaaaaaaaccccgccctgtcaggggcgggtttttttttgcatgcttatttgtagagctcatccatgc
900+20K GO1	gcatctttttgtacctataatagattcatcacctactaaagcaagggagggtaccatgagcaaaatgcaggagttcgcgc
900+20K GO2	gatgaatctattataggtacaaaaagatgcgaagtcaatactctttttggtacctcatgaccaaaatcccttaacgtga
700+20K GO1	atcccttgcggcgatataatgtgtggatcacaatccacagccacggaggggtaccatgagcaaaatgcaggagttcgcgc
700+20K GO2	tgatccacacattatatcgccgcaagggataaaagtcaatactctttttggtacctcatgaccaaaatcccttaacgtga
500+20K GO1	atttttctgatacttacagccatgcatcaaacaacgaccaggagggtaccatgagcaaaatgcaggagttcgcgc
500+20K GO2	ttgatgcatggctgtaagtatcagaaaaattttcctgatgtcaaggtacctcatgaccaaaatcccttaacgtga
300+20K GO1	atccggctcgtagtgtttgtggaaactcaaagacagagaaggagggtaccatgagcaaaatgcaggagttcgcgc
300+20K GO2	tttgagtttccacaaacactacgagccggatgattaattgtcaaggtacctcatgaccaaaatcccttaacgtga
100+20K GO1	atcccttgcggcgatataatagattcatagttacacgaaatataaggcgggtaccatgagcaaaatgcaggagttcgcgc
100+20K GO2	ctatgaatctattatatcgccgcaagggataaaagtcaatactctttttggtacctcatgaccaaaatcccttaacgtga
50+20K GO1	atccggctcgtataatgtgtggagaggcttcgcgggaggaggagggtaccatgagcaaaatgcaggagttcgcgc
50+20K GO2	gaagcctctccacacattatacgagccggatgattaataaaaaaggtacctcatgaccaaaatcccttaacgtga
10+20K GO1	ttttgtacctagatttaacgtatagccgacattagaggaggaagggtaccatgagcaaaatgcaggagttcgcgc
10+20K GO2	gtcggctatacgttaaatctaggtacaaaaagatgcgatgtcgaggtacctcatgaccaaaatcccttaacgtga
1+20K GO1	tccggctcgtataatgtgtggatataatctacagtataagagagggtaccatgagcaaaatgcaggagttcgcgc
1+20K GO2	agattatatccacacattatacgagccggatgattaattgtcaaggtacctcatgaccaaaatcccttaacgtga
bmoR GF1	aagcgggtttttacgtaaatcaggtgaaactgaccgactcgagctatgtgccgatccgcg
bmoR GR1	acgaaaactcacgttaagggattttggtcatgaggtaccatgtccaagatgcaagagttc
EcAdhE GF1	aaatctgttgcagaacgccgtgaaaacatgctgtggcacaaacttccgaaatctatctac
EcAdhE GR1	tctgctcatgtttgacagcttatcatcgataagcttggtaccttaagcggatttttcgc
900+20K GF1	tgatgaatctattataggtacaaaaagatgcgaagtcaatactctttttggtacctcatgaccaaaatcc
900+20K GR1	tcgcatctttttgtacctataatagattcatcacctactaaagcaagggagggtaccatgagcaaaatgc
pMD105 GF1	ctgtctttgagtttccacaaacactacgagccggatgattaattgtcaaggtacctcatgaccaaaatcc
pMD105 GR1	attaatcatccggctcgtagtgtttgtggaaactcaaagacagagaaggagggtaccatgagcaaaatgc
pMD106 GF1	ctatgaatctattatatcgccgcaagggataaaagtcaatactctttttggtacctcatgaccaaaatcc
pMD106 GR1	ttatcccttgcggcgatataatagattcatagttacacgaaatataaggcgggtaccatgagcaaaatgc
pMD109 GF1	actgtagattatatccacacattatacgagccggatgattaattgtcaaggtacctcatgaccaaaatcc
pMD109 GR1	ttaatcatccggctcgtataatgtgtggatataatctacagtataagagagggtaccatgagcaaaatgc
pMD120 GF1	ggtttggatccacacattatacgagccggatgattaattgtcaagctagcccacagatagtaggtgctgc
pMD120 GR1	tccggctcgtataatgtgtggatccaaaccgatcccgaggggaggtaccatgtccaagatgcaagagttc
pMD121 GF1	aacggcgatacgttaaatctaggtacaaaaagatgcgatgtcgagctagcccacagatagtaggtgctgc
pMD121 GR1	ttttgtacctagatttaacgtatcgccgttacacaaaaaggagggtaccatgtccaagatgcaagagttc
pMD122 GF1	tggacgtttccacacattatacgagccggatgattaataaaaaagctagcccacagatagtaggtgctgc
pMD122 GR1	atccggctcgtataatgtgtggaaacgtccaacgcccacgggaggtaccatgtccaagatgcaagagttc
pMD123 GF1	tcatgaatctattatatcgccgcaagggataaaagtcaatactctttttgctagcccacagatagtaggtgctgc
pMD123 GR1	atcccttgcggcgatataatagattcatgacacaaagctttactaggggtaccatgtccaagatgcaagagttc
pMD124 GF1	tgtaatgttccacaaacactacgagccggatgattaattgtcaagctagcccacagatagtaggtgctgc
pMD124 GR1	atccggctcgtagtgtttgtggaacattacaacgagccaaggaggtaccatgtccaagatgcaagagttc
pMD125 GF1	ttgatgcatggctgtaagtatcagaaaaattttcctgatgtcaagctagcccacagatagtaggtgctgc
pMD125 GR1	tttttctgatacttacagccatgcatcaaacaacgaccaggagggtaccatgtccaagatgcaagagttc
pMD126 GF1	gtggactcatccacacattatatcgccgcaagggataaaagtcaatactctttttccacagatagtaggtgctgc
pMD126 GR1	atcccttgcggcgatataatgtgtggatgagtccacaagaaacccggaggtaccatgtccaagatgcaagagttc
adhE2 850 GR1	gctcatgtttgacagcttatcatcgataagcttggtaccttacagctcagagatcagcgg
adhE2 854 GR1	tttgacagcttatcatcgataagcttggtaccttaataaatgtctttcagctcagagatc

Name	Sequence
ALDH46 GF1	taactacgaggcagaagttgaac
ALDH46 fucO GR1	atcattctgttagccattgtctcccccccctgcgccggctcgagttagccggccagaacac
ALDH46 yqhD GR1	agattaaagttgttcatctttacctcctgatagaagtctcgagttagccggccagaacac
fucO ALDH46 GF1	ggccggctaactcgagccggcgcagggggggagacaatggctaacagaatgattctgaac
yqhD ALDH46 GF1	ggccggctaactcgagacttctatcaggaggtaaagatgaacaactttaatctgcacacc
fucO GR1	gtatcacgaggccctttcgtcttcaagcagatcgagctcttaccaggcggtatggtaaag
yqhD GR1	atcacgaggccctttcgtcttcaagcagatcgagctcttagcgggcggcttcgtatatac
pMD13x GF1	ttcatctttctgcccctccaaaagcaaaaacccgccgaagcgggtttttacgtaaatcaggtgaaactgaccgactatgtgccgatcc gc
pMD130 GR1	tatatcgccgcaagggataaaagtcaatactctttttggtaccaggtggcacttttcggg
pMD131 GR1	atggctgtaagtatcagaaaaattttcctgatgtcaaggtaccaggtggcacttttcggg
mRFP GF1 100	agtattgacttttatcccttgcggcgatataatagattcatgacacaaagctttactaggggtaccatggcgagtagcgaagac
mRFP GF1 500	gacatcaggaaaatttttctgatacttacagccatgcatcaaacaacgaccaggagggtaccatggcgagtagcgaagac
mRFP GR1	ggaggggcagaaagatgaatgactgtccatatgaaaaaaaaccccgccctgtcaggggcggggttttttttttttaagcaccggtgga gtg
aldh46 bdhB GR1	tactcgaagtccaccatttttacctctctctgtttagctcgagttagccggccagaacac
bdhB aldh46 GF1	ttctggccggctaactcgagctaaacagagagaggtaaaaatggtggacttcgagtattc
bdhB GR1	cacgaggccctttcgtcttcaagcagatcgagctcttaaacggatttcttgaaaatttgc
ilvE KF1	cacgttgccatctgccagagcacaaccacatcacaacaaatccgcgcctgagcgcaaaaggaatataaaagtgtaggctggagc tgct
ilvE KR1	ttatttactcccagtgtctgtctcgtaaatgggacggtgcgtgccgtcccatttttgtatcatatgaatatcctccttagttcctattc
ilvE KVF1	agtcagttaaataaactggtggacg
ilvE KVR1	accatgagtggtggtggc
eGFP GF3	attttattacgcagaatagataaggagggcgagaatgagtaaaggagaagaacttttcac
eGFP GR3	tttatttgatgcctggtctagactcgagttaggatccttatttgtagagctcatccatgc
pMD140 GF1	tgggattacacatggcatggatgagctctacaaataaggatcctaactcgagtctagacc
pMD140 GR1	tgggacaactccagtgaaaagttcttctcctttactcattctcgccctccttatctattc
ilvE GF1	acacaatattttattacgcagaatagataaggagggcgagaatgaccacgaagaaagctg
ilvE GR1	atttgatgcctggtctagactcgagttaggatccttattgattaacttgatctaaccagc
pMD150 GF1	agataaatggggctggttagatcaagttaatcaataaggatcctaactcgagtctagacc
pMD150 GR1	cccattgaaccaaatgtaatcagctttcttcgtggtcattctcgccctccttatctattc
.03 HBD QF1	cgactacgggtaattcggaac
.03 HBD QR1-1	atcaacaaattcatccttaatgtcg
.03 HBD QR1-2	gtttttgttaatgaagtccaggc
.03 HBD QR1	cgactgtcgttccgaattacc
.03 HBD QF1-1	ggttgccgtatcctggtgacg
.03 HBD QF1-2	gccacccaatcggtgcttctg
pTT QF1	aatgagctgttgacaattaatcatc
pTT QR1	ccaccgaacttaccaacagc
pTT QF2	cgtataatgtgtggaattgtgagc
pTT QR2	actactgcgcccagttccgg
pTT QF3	agctggcacgacaggtttcc
pTT QR3	cacagcttccgatggctgcc
.3 crt QF1	ggctgcacacaagacagacc
.3 crt QR1-1	gatttcgccaataacgtagtcc
.3 crt QR1-2	gcacttcgctgtcgttctcg
.3 crt QR1	gtcttgtgtgcagcctctcg
.3 crt QF1-1	tcgctcaaggcgcactcccg
.3 crt QF1-2	gtaaatcactgcataattcgtgtcg
CmR QF1	ggggcgaagaagttgtcc
CmR QR1	atattcgcaagatgtggcg
CmR QF2	ttgccatacggaattccg
CmR QR2	gaggcatttcagtcagttgc
adhE2 delete GO1	cccgccctgtcaggggcggggttttttttgaattcttaaatacgatcgaaacgttcaac
adhE2 delete GO2	aaaaccccgcccctgacagggcggggttttttttggtaccaagcttatcgatgataagct
aldh46 max GF1	agcggataacaatttcacacaggaaacagctcgagctatcaataaggaggaagtaatgaataaagacaccctgattcc
aldh46-bdhB max GR1	tactcgaagtccaccatttatacctcctcttattagctctagattagccggccagaacac
aldh46-bdhB max GF1	ggccggctaatctagagctaataagaggaggtataaatggtggacttcgagtattctatc
bdhB max GR1	gcgtaatacgactcactatagggcgaattggagctcttaaacggatttcttgaaaatttg
aldh46-yqhD max GR1	agattaaagttgttcatgtgttacctccttattagtttctagattagccggccagaacac
aldh46-yqhD max GF1	tggccggctaatctagaaactaataaggaggtaacacatgaacaactttaatctgcacac
yqhD max GR1	gtgagcgcgcgtaatacgactcactatagggcgaattggagctcttagcgggcggcttcg
aldh46-fucO max GR1	atcattctgttagccatttttacctcctctggtgccgtctagattagccggccagaacac
aldh46-fucO max GF1	ggccggctaatctagacggcaccagaggaggtaaaaatggctaacagaatgattctgaac

Name	Sequence
fucO max GR1	cgcgcgtaatacgactcactatagggcgaattggagctcttaccaggcggtatggtaaag
phaA GF1	ttgggctagcgaattcgagctcaaggagatatacatatgactgatgttgtgattgtaagc
phaA-rrnB GR1	caggctgaaaatcttctctcatccgccaaaacagccctcgagttatttgcgttcgacagc
phaA-rrnB GF1	tggtatgggtgtagcactggctgtcgaacgcaaataactcgagggctgttttggcggatg
rrnB GR1	gacgccagaagcattggtgcaccgtgcagtcggccgagagtttgtagaaacgcaaaaagg
phaB GF1	aaagggaacaaaagctgggtaccccagcaaataggaggaagccatgacccagcgcatcgc
phaB GR1	tgagcgcgcgtaatacgactcactatagggcgaattggagctcttagcccatgtgcaggc
ter epPCR F1	cgatgcttaggaggtcatatg
ter epPCR R1	ttttacctccttttttgaattctta
ter delete GO1	gatccatcgatgcttaggaggtcatatgtaaaaaaaaaccccgcccctgacagggcgggg
ter delete GO2	attcatttttttacctccttttttgaattcaaaaaaaaccccgccctgtcaggggcggggt
ter epPCR GF1	ggataacaatttcacacaggaaacaggatccatcgatgcttaggaggtca
ter epPCR GR1	ttcgtggtcggaatcagg
300ter epPCR GF1	ggataacaatttcacacaggaaacaggatccatcgatgcttaggaggtcatatgtatactcacttcagaaacgtatcatg
300ter epPCR GR1	ctgcttcagttccttctggttcgtgactttcatggtatatctccttcttgaattcttaaatacgatcgaaac
pCDF3-300Ter GF1	gttgagggtattaactacgaggc
pCDF3-300Ter GR1	ttggcttgacgatcatgatacgtttctgaagtgagtatacatatgacctcctaagcatcg
phaB MF1	caagaccgcattcgacaaagttaaatctgaagtaggcgaggtagatgtcctgatcaacnnsnnsnnsnnsnnscgtgatgttgtgt tccg
phaB R1	gttgatcaggacatctacctcg
pBBR2 short GF1	tatctggacaagggaaaacgcaagcgcaaagagaaagcaggtagcttgcagtgggcttac
pBBR2 short GR1	ctttctctttgcgcttgcgttttcccttgtccagataaaaatcaacaacttaaaaaaggg
ter delete GO3	tgaattcaaaaaaaaccccgccctgtcaggggcggggttttttttttacatatgacctcctaagcatcgatggatcctgtttcctgtgtga
ter delete GO4	catatgtaaaaaaaaaccccgccctgacagggcgggttttttttgaattcaaaaaaggaggtaaaaaatgaataaagacaccct gatt
phaA GF2	ctctctactgtttctccataccc
phaA rbs HBD GR1	tttcattttggggcgactgtcgttccgaattacccgtagtcgttatttgcgttcgacagc
phaA rbs HBD GF1	aacgactacgggtaattcggaacgacagtcgccccaaaatgaaaaaggtttgcgttattg
HBD GR1	atcaggctgaaaatcttctctcatccgccaaaacagccctcgagttacttggagtaatcg
Ptrc GF1	tggatggaggcggataaag
Ptrc rbs crt GR1	catgactcggtctgtcttgtgtgcagcctctcgactcgctcacaattccacacattatac
Ptrc rbs crt GF1	tgagcgagtcgagaggctgcacacaagacagaccgagtcatggaactgaacaacgtgatc
crt GR1	ttcggtgcgggggtcttag
phaB GGF1	gatcgaggtctcccagcaaataggaggaagccatg
phaB GGR1	ctagctggtctctgggcgaattggagctcttag
pBBR2 phaB GGF1	ctagctggtctctgctggggtacccagcttttg
pBBR2 phaB GGR1	gatcgaggtctcagccctatagtgagtcgtattacg
phaB GF2	cagctatgaccatgattacgccaagcgcgcaattaaccctcactaaagggaacaaaagctgggtaccccagcaaataggagga agccatg
phaB GR2	gggttttcccagtcacg
pCDF3GG GO1	ttgagaccctaagcatcgaggtctcggatcctgtttcctgtgtgaaattgttatccgctc
pCDF3GG GO2	cgagacctcgatgcttagggtctcactagtaaaaaaggaggtaaaaaatgaataaagaca
ter epPCR GGF1	gatcgaggtctcggatccatcgatgcttaggaggtcatatg
ter epPCR GGR1	gatcgaggtctctctagtcttaaatacgatcgaaacgttc
lac tac ter GF1	gagccttttgttttatttgatg
lac tac ter GR1	tgtctttattcatttttttacctccttttttgaattcttaaatacgatcgaaacgttcaac
aldh46 GF2	cgaggcagaagttgaacgtttcgatcgtatttaagaattcaaaaaaggaggtaaaaaatg
aldh46 GR2	tgaaaatcttctctcatccgccaaaacagccaagcttgcatgcttagccggccagaacac
aldh46 GF3	tgagggtattaactacgaggc
aldh46.ADH GR1	actttgaaccacagcattaggacctcctctggtaagctctagattagccggccagaacac
aldh46.ADH GF1	gttctggccggctaatctagagcttaccagaggaggtcctaatgctgtggttcaaagtcc
ADH GR1	catgtttgacagcttatcatcgataagcttttaaaaagatttgatataaatgtctttcag
pBu2-TS GF1	ctctgagctgaaagacatttatatcaaatctttttaagcatgcctgttttggcggatgag
pBu2-TS GR1	agagaacatactggctaaatacggaaggatctgaactagttacctgtgacggaagatcac
pBu2-TS GF2	tttattctgcgaagtgatcttccgtcacaggtaactagttcagatccttccgtatttagc
pBu2-TS GR2	gcggttgctggcgcctatatcgccgacatcaccggcggccgcgttgatgataccgctgcc
pBu2-TS GF3	aatgcacccagtaaggcagcggtatcatcaacgcggccgccggtgatgtcggcgatatag
pBu2-TS GR3	tcacaacatcagtcatatgtatatctccttgagctcgaattcgctagcccaaaaaaacgg
aldh4/10 GF1	ttgtgagcggataacaatttcacacaggaaacaggaattcaaaaaaggaggtaaaaaatg
aldh4.ADH GR1	ctttgaaccacagcattataacctccttagtattgagagctcttaaacaatgcgcagacc
aldh4.ADH GF1	ctgcgcattgtttaagagctctcaatactaaggaggttataatgctgtggttcaaagtcc
aldh10.ADH GR1	tttgaaccacagcatttttaacctccttatggctcgagctcttacacgatgcgaaaagag
aldh.ADH GF1	tttcgcatcgtgtaagagctcgagccataaggaggttaaaaatgctgtggttcaaagtcc
ADH GR2	catgtttgacagcttatcatcgataagcttttaaaaagatttgatataaatgtctttcag
trc.crt delete GO1	caagcttgcatgcctgcaggtcgactctagattagcccatgtgcaggccaccgttcaggg
trc.crt delete GO2	gaacggtggcctgcacatgggctaatctagagtcgacctgcaggcatgcaagcttggctg
lac-tac.ter GF2	cccagtcttccgactgag
lac-tac.ter GR2	gttgaacgtttcgatcgtatttaaaagcttaaaaaaaaaccccgcccctgacagggcggg

Name	Sequence
rrnb-trc.46.adh GF1	ttaaaaaaaaaccccgcccctgacagggcggggttttttttatggaggcggataaagttg
rrnb-trc.46.adh GR1	gctgaaaatcttctctcatccgccaaaacaggcatgcttaaaaagatttgatataaatgtctttcag
dnaQ GF1	cccgttttttgggctagcgaattcgagctcaaggagatatacatatgagcactgcaattacacgccagatcgttctcgcgaccgcgac c
dnaQ.umuDC GR1	ggtgaaggaaagcccatttgtcctccgtttctagagatctagattatgctcgccagaggc
dnaQ.umuDC GF1	tctggcgagcataatctagatctctagaaacggaggacaaatgggctttccttcaccggc
umuDC.recA1 GR1	atagccatagaccctcctgtctctgtgtcatgcttatttgaccctcagtaaatcagaac
umuDC.recA1 GF1	agggtcaaataagcatgcacacagagacaggagggtctatggctatcgacgaaaacaaac
recA1.recA2 GR1	ccacctgccccaagcgcgatatccagtgaaagcgaaccggtagagatggttttcacatcc
recA1.recA2 GF1	ccatcatgcgcctgggtgaagaccgttccatggatgtgaaaaccatctctaccggttcgc
recA2 GR1	ctgaaaatcttctctcatccgccaaaacagcggccgttaaaaatcttcgttagtttctgc
strep_aldh46 GF1	ggatccatcgatgcttaggaggtcatatggcaagctggagccacccgcagttcgaaaagggtgcaggtatgaataaagacaccct gattc
aldh46 GR1	gctcagatctgctcatgtttgacagcttatcatcgataagcttttagccggccagaacac
HisTev_aldh46 GF1	tcatcatgagaatctctacttccagggtaccggcgccatgaataaagacaccctgattcc
HisTev_aldh46 GR1	gttagcagccggatctcagtggtggtggtggtggtgctcgagtttagccggccagaacac
HisTev_aldh3 GF1	catcatgagaatctctacttccagggtaccggcgccatgattaaggacactctcgtaagc
HisTev_aldh3 GR1	agcagccggatctcagtggtggtggtggtggtgctcgagtttaacccgccagaacacaac
HisTev_aldh6 GF1	catcatcatgagaatctctacttccagggtaccggcgccatgaaagagggtgtaattcgc
HisTev_aldh6 GR1	tagcagccggatctcagtggtggtggtggtggtgctcgagtttaacgaatgctaaaggcg
HisTev_aldh7 GF1	catcatcatcatgagaatctctacttccagggtaccggcgccatggaacgcaacttgtcg
HisTev_aldh7 GR1	agcagccggatctcagtggtggtggtggtggtgctcgagtttaaccggccagaacgcaac
ter delete GO5	agcggataacaatttcacacaggaaacaggatccgaattcaaaaaaggaggtaaaaaatg
ter delete GO6	catttttacctccttttttgaattcggatcctgtttcctgtgtgaaattgttatccgct
aldh3 GR1	actttgaaccacagcattaggacctcctctggtaagctctagattaacccgccagaacac
aldh6 GR1	tttgaaccacagcattaggacctcctctggtaagctctagattaacgaatgctaaaggcg
aldh7 GR1	actttgaaccacagcattaggacctcctctggtaagctctagattaaccggccagaacgc
dnaQ GR1	tatcaggctgaaaatcttctctcatccgccaaaacagaagcttttatgctcgccagaggc
pMut2 GF1	gctaatgcacccagtaaggcagcggtatcatcaacctgcagggtgatgccggccacgatg
pMut2 GR1	cgagatcaccaaggtagtcggcaaataactcgagaagagtttgtagaaacgcaaaaagac
aadA GF1	gacggatggtctttttgcgtttctacaaactcttctcgagttatttgccgactaccttgg
aadA GR1	actagagaacatactggctaaatacggaaggatctgatctagaccgccgcagtctcacgc
repA101ts GF1	ggtcgctacgctccgggcgtgagactgcggcggtctagatcagatccttccgtatttagc
repA101ts GR1	atcctctacgccggacgcatcgtggccggcatcaccctgcaggttgatgataccgctgcc
aldh46 GF4	gaagataagattctgaaacatgagc
aldh46.yqhD GR1	agattaaagttgttcatctttacctcctgatagaagtctcgagttagccggccagaacac
aldh46.yqhD GF1	tggccggctaactcgagacttctatcaggaggtaaagatgaacaactttaatctgcacac
yqhD GR2	tcatgtttgacagcttatcatcgataagcttgagctcttagcgggcggcttcgtatatac
aldh46.fucO GR1	atcattctgttagccattgtctccccccctgcgccggctcgagttagccggccagaacac
aldh46.fucO GF1	ggccggctaactcgagccggcgcagggggggagacaatggctaacagaatgattctgaac
fucO GR2	gctcatgtttgacagcttatcatcgataagcttgagctcttaccaggcggtatggtaaag
aldh46.dhaT GR1	ttagccggccagaacac
aadA GF2	tcgatgaattttctcgggtgttctcgcatattggctcgaattcccgccgcagtctcacgc
aadA GR2	tgaagctcttgttggctagtgcgtagtcgttggcaagcttttatttgccgactaccttgg
aadA GR3	ctgacgccagaagcattggtgcaccgtgcagtcgaagcttttatttgccgactaccttgg
Ptrc GF2	cgtgaaaggcgagatcaccaaggtagtcggcaaataaaagcttcgactgcacggtgcacc
Ptrc GR1	gttggccgattcattaagacagctgtctcttatacacatctcgccgagttacctcctacctcgagctgtttcctgtgtgaaattgttatc
yqhD GF1	gtaacttcacgcgccaacgtcgttgtgttctggccggctaactcgagacttctatcaggaggtaaagatgaacaactttaatctgcaca c
ME F1	ctgtctcttatacacatctcaaccatc
ME R1	ctgtctcttatacacatctcaaccc
recA GF1	ttttgggctagcgaattcgagctcaaggagatatacatatggctatcgacgaaaacaaac
recA GR1	taatggtttcttagacgtcaggtggcactcgagttaaaaatcttcgttagtttctgctac
TcR GF1	agcgaaggcgtagcagaaactaacgaagatttttaactcgagtgccacctgacgtctaag
TcR GR1	actagagaacatactggctaaatacggaaggatctgatctagacgtattgggcgcatttg
oriR101ts GF1	gagaactgtgaatgcgcaaatgcgcccaatacgtctagatcagatccttccgtatttagc
oriR101ts GR1	gcagatcctctacgccggacgcatcgtggccggcatggtaccgttgatgataccgctgcc
Pbad GF1	tggctaatgcacccagtaaggcagcggtatcatcaacggtaccatgccggccacgatgcg
Pbad GR1	ccaacgctttctgtttgttttcgtcgatagccatatgtatatctccttgagctcgaattc
FRT1 GO1	cgaagttcctattctctagaaagtataggaacttcgaagcagctccagcctacactgatggttgagatgtgtataagagacagctgtc gt
FRT1 GO2	tcgaagttcctatactttctagagaataggaacttcggaataggaacttcgaattcccgccgcagtctcacgcccggagcgtagcgac cg
FRT2 GO1	agttcctattccgaagttcctattctctagaaagtataggaacttcctcgagttatttgccgactaccttggtgatctcgcctttcacgt
FRT2 GO2	ctatactttctagagaataggaacttcggaataggaactaaggaggatattcatatgtcagggttgagatgtgtataagagacagctgt c
cI857 FLP GF1	ttctagagaataggaacttcggaataggaacttcgaattctcagccaaacgtctcttcag
cI857 FLP GR1	tgaatgtatttagaaaaataaacaaatagctagcttatatgcgtctatttatgtaggatg

Name	Sequence
mutS KF1	atgagtgcaatagaaaatttcgacgcccatacgcccatgatgcagcagtagtgtaggctggagctgcttc
mutS KR1	ttacaccaggctcttcaagcgataaatccactccagcgcctgacgcggggcatatgaatatcctccttagttcctattc
cra stop LR1	a*g*t*g*gtccgcgacactcccgccagccgagcgatttcatccagtcattaaattgccccttgcgtaaaatgtaaaaaccatagccattgtac
cra stop LR2	g*t*a*c*aatggctatggtttttacattttacgcaaggggcaatttaatgactggatgaaatcgctcggctggcgggagtgtcgcggaccact
fadR stop LR1	a*a*t*g*tactcttccgcgaaacccgccgggctttgcgccttaattcattaagtgagatttccataacacagcaaaacaaagttggactcatc
fadR stop LR2	g*a*t*g*agtccaactttgttttgctgtgttatggaaatctcacttaatgaattaaggcgcaaagcccggcgggtttcgcggaagagtacatt
arcA stop LR1	g*t*a*c*ttcctgtttcgatttagttggcaatttaggtagcaaactaatgaaccccgcacattcttatcgttgaagacgagttggtaacacgc
arcA stop LR2	g*c*g*t*gttaccaactcgtcttcaacgataagaatgtgcggggttcattagtttgctacctaaattgccaactaaatcgaaacaggaagtac
crp stop LR1	t*a*t*c*tggctctggagaaagcttataacagaggataaccgcgctaatgacttggcaaaccgcaaacagacccgactctcgaatggttcttg
crp stop LR2	c*a*a*g*aaccattcgagagtcgggtctgtttgcggtttgccaagtcattagcgcggttatcctctgttataagctttctccagagccagata
cya stop LR1	a*t*t*t*atggcatccagtctctgtttcagagtctcaatatagagtcattagacgtatcgcctgatttgctacccgtcatgactgtgattccg
cya stop LR2	c*g*g*a*atcacagtcatgacgggtagcaaatcaggcgatacgtctaatgactctatattgagactctgaaacagagactggatgccataaat
iclR stop LR1	a*a*a*a*tgaaaatgatttccacgatacagaaaaaggagactgtctaatgagcacccattcccgcgaaacgcggcagaaaaccccgccgttgcc
iclR stop LR2	g*g*c*a*acggcgggttttctgccgcgtttcgcgggaatgggtgctcattagacagtctccttttttctgtatcgtggaaatcattttcatttt
lrp stop LR1	a*t*c*g*atacggtcgagatctttgccagggcgcttcttgctatctcattatattattgtctctctgttattccttccctactcctgtctgacc
lrp stop LR2	g*g*t*c*agacaggagtagggaaggaatacagagagacaataatataatgagatagcaagaagcgccctggcaaagatctcgaccgtatcgat
fnr stop LR1	a*t*g*t*taaaattgacaaatatcaattacggcttgagcagaccttaatgaccggaaaagcgaattatacggcgcattcagtctggcggttgt
fnr stop LR2	a*c*a*a*ccgccagactgaatgcgccgtataattcgcttttccggtcattaaggtctgctcaagccgtaattgatatttgtcaattttaacat
narL stop LR1	t*t*a*t*tcccgaaaaaactttcacagacgtccaaggagataccctaatagaatcaggaaccggctactatcctgctgattgacgatcacccg
narL stop LR2	c*g*g*g*tgatcgtcaatcagcaggatagtagccggttcctgattctattagggtatctccttggacgtctgtgaaagtttttcgggaataa
mlc stop LR1	g*g*t*c*tgctttatttgatcaatgtgcccaggctggttttcagctcattacgcatactccctatattttcgcgctccgaaataatctgtagg
mlc stop LR2	c*c*t*a*cagattatttcggagcgcgaaaatatagggagtatgcgtaatgagctgaaaaccagcctgggcacattgatcaaataaagcagacc
mutS off	a*t*c*a*caccccatttaatatcagggaaccggacataaccccatcagtgcaatagaaaatttcgacgcccatacgcccatgatgcagcagta
mutS stop LR1	c*c*a*t*cacaccccatttaatatcagggaaccggacataaccccctaataggcaatagaaaatttcgacgcccatacgcccatgatgcagcag
mutS off wt F1	gaaccggacataaccccatg
mutS off mut F1	gaaccggacataaccccatc
mutS stop wt F1	ccggacataaccccatgagt
mutS stop mut F1	ggaaccggacataaccccctaatag
mutS R1	tcagccggttcgctcag
cra stop wt F1	cgcaaggggcaattgtgaa
cra stop mut F1	attttacgcaaggggcaatttaatg
cra stop R1	cggtattgcttcgctttgc
arcA stop wt F1	gcaatttaggtagcaaacatgcag
arcA stop mut F1	tggcaatttaggtagcaaactaatga
arcA R1	gagaggatctgatgcatttccg
crp stop wt F1	ggataaccgcgcatggtg
crp stop mut F1	agaggataaccgcgctaatga
crp R1	gaggatcatttctttaccctcttcg
mlc stop wt F1	aaaatatagggagtatgcggtggtt
mlc stop mut F1	cgaaaatatagggagtatgcgtaatga
mlc R1	cccaccgccggacggcc
Tn5:Sp GF1	ggctaatgcacccagtaaggcagcggtatcatcaacggtaccattcaggctgcgcaactg
Tn5:Sp GR1	tcatgataataatggtttcttagacgtcaggtggcactcgaggtcagtgagcgaggaagc
cI857 GF1	atgcacccagtaaggcagcggtatcatcaacggtacctcagccaaacgtctcttcaggcc
cI857 GR1	gccagcgttgcgagtgcagtactcattcgacctccttagtacatgcaaccattatcaccg
bet GF2	ggcggtgataatggttgcatgtactaaggaggtcgaatgagtactgcactcgcaacgctg
bet GR2	ataataatggtttcttagacgtcaggtggcactcgagtcatgctgccaccttctgctctg
mutS GF3	ctgaaacagaaagccgcagagcagaaggtggcagcatgatgtcacaggataaggaggtccatatgagtgcaatagaaaatttcgacgccc
mutS1 mutS2 GR1	gcaatcagtgcggtctggcgcatataggtacttttatcgcccatgttcggaccggtgatg

Name	Sequence
mutS1 mutS2 GF1	catgttgatcatcaccggtccgaacatgggcgataaaagtacctatatgcgccagaccgc
mutS2 GR1	atggtttcttagacgtcaggtggcactcgagttacaccaggctcttcaagcgataaatcc
mutS GF1	tttttgggctagcgaattcgagctcaaggagatatacatatttaataaggaggtccacccatatgagtgcaatagaaaatttcgacgcc c
mutS GF2	tttttgggctagcgaattcgagctcaaggagatatacatgaaaccaacaacaggagaatcatatgagtgcaatagaaaatttcgacg ccc
mutS2 GR2	atggtttcttagacgtcaggtggcactcgagttacaccaggctcttcaagcgataaatcc
ilv.rrnD terminator GF2	tggatttatcgcttgaagagcctggtgtaatagagatcaagccttaacgaactaagaccc
ilv.rrnD terminator GR2	gtcagtggcctgaagagacgtttggctgagggaactgccagacatcaaataaaacaaaag
term cI857 GF1	tgagccttttgttttatttgatgtctggcagttccctcagccaaacgtctcttcaggcca
dmlA stop LR1	t*g*g*t*gatgttgtcaattttgatggtcaggaagtgagaacccataataaatgaaaacgatgcgtattgctgcgatcccgggagacg ggatt
dmlA stop LR2	a*a*t*c*ccgtctcccgggatcgcagcaatacgcatcgttttcatttattatgggttctcacttcctgaccatcaaaattgacaacatcac ca
dmlA W34* LR1	c*c*t*t*cctgaagggattcgcgtgttacaggctgccgctgagcgctgaggcttcgccttgagttttgagcaaatggagtgggcgagc tgcga
dmlA W34* LR2	t*c*g*c*agctcgcccactccatttgctcaaaactcaaggcgaagcctcagcgctcagcggcagcctgtaacacgcgaatcccttc aggaagg
aadA F1	ccgtcatacttgaagctagacagg
aadA F2	cgagatcaccaaggtagtcggc
aadA R2	cttcggcgatcaccgcttccctc
aadA R1	aggccgccatccactgcggagcc
CmR F2	gcgatgagtggcagggcggggcg
CmR F1	caaggcgacaaggtgctgatgcc
CmR R2	cggtggtatatccagtgatttttttctcc
CmR R1	aaggccgtaatatccagctgaac
tonA KF1	cttatctttataataatcattctcgtttacgttatcattcactttacatcagagatataccagtgtaggctggagctgcttcgaagttcc
tonA KR1	ttcctgcataacagccaacttgtgaaatgggcacggaaatccgtgccccaaaagagaaacatatgaatatcctccttagttcctattcc g
tolC KO	taaatgtgaatttcagcgacgtttgactgccgtttgagcagtcatatgacgacgacggggcttcggccccgtctgaacgtaaggcaac gt
tolC KVF1	catcgtttttgccaaatgtaacgggc
tolC KVR1	cagcataaaaacagtggcaaccgc
pac1 tolC KF1	aacggcctgcgggccgttttgttttgtctggattttgcgctttttgcccattgaggcacattaacgccctatggc
pac1 tolC KR1	gcgctgtcgcgtatttttttcacaattcccgggcaattttcgtctgaatgctctagggcggcggatttgtcc
pac1 pacWT KF1	aacggcctgcgggccgttttgttttgtctggattttgcgctttttgcccacatgatcattgatcactctaatgatcaacatgcag
pac1 pacWT KR1	gcgctgtcgcgtatttttttcacaattcccgggcaattttcgtctgaatgcgttccaggcgatcacgtaaggtc
pac1 KVF1	atgatttacagtactttagcggaggaactctc
pac1 KVR1	ctgaaacttgcatcttgttacctcaaaaaatc
tolC GF1	tgcacccagtaaggcagcggtatcatcaacggtaccttgaggcacattaacgccctatgg
tolC GR1	tcagttacggaaagggttatgaccgttac
rrnB-1 GF1	ggtattaactacgaggcagaagttg
rrnB-1 GR1	gttccctactctcgcatgggctgaccccacactaccatcg
rrnB-2 GF1	cgatggtagtgtggggtcagcccatgcgagagtagggaac
rrnB-2 GR1	cagcttccgatggctgcc
BsaI delete GO1	gctgataaatctggagccggtgagcgtggctcccgcggtatcattgcagcactggggcca
BsaI delte GO2	tggccccagtgctgcaatgataccgcgggagccacgctcaccggctccagatttatcagc
aldh3 C275A GR1	caaaaacaaaaacctctttctccgcgatcgccggcagattgttgtcgaagctgcagcc
aldh3 C275A GF1	ggctgcagcttcgacaacaatctgccggcgatcgcggagaaagaggtttttgtttttg
aldh46 C275A GR1	caaacacgaaaacctctttctccgcaatcgccggcaggttgttatcgaagc
aldh46 C275A GF1	gcttcgataacaacctgccggcgattgcggagaaagaggttttcgtgtttg
specR GF1	aaggcgtagcagaaactaacgaagatttttaactcgagcagccaggacagaaatgcctcg
specR GR1	accagcagtcctgttattagctcagtaaagctttatttgccgactaccttggtgatctcg
specR GO1	ctagatgcagaaataaaaaggcctgcgattaccagcagtcctgttattagctcagtaaag
specR GO2	aatcgcaggccttttttatttctgcatctagatcagatccttccgtatttagccagtatgt
pac1-Sp45 KF1	tcccgcctgctgatattgaaactggctgcgtctcgcgcgctcccgtcagacagccaggacagaaatgcctcg
pac1-Sp45 KR1	ttcatggttgtttttacgggtggttaactgagcggcgaatgttaacacaatgcagaaataaaaaggcctgcgattac
pac1-Sp45 KVF1	gcgctcaagcgtgtaatactgtcg
pac1-Sp45 KVR1	caatcaccagtgaatgcccggc
pRecA GF1	taaagctttactgagctaataacaggactgctggtaatcgcaggccttttttatttctgcatctagatcagatccttccgtatttagccag
pRecA GR1	agcgaagtcgaggcatttctgtcctggctgctcgagttaaaaatcttcgttagtttctgc
SpecR GF2	aaggcgtagcagaaactaacgaagatttttaactcgagcagccaggacagaaatgcctcg
SpecR GR2	tctagatgcagaaataaaaaggcctgcgattaccagcagtcctgttattagctcagtaaagctttatttgccgactaccttggtgatctc
pac1-Sp225 KF1	tcggcaaaatcgcgatgttgtttagctgacgatgtgtttttttgcgccgatcagccaggacagaaatgcctcg
pac1-Sp225 KR1	gagatcccataaacacggctgccggatgcggcgtaaacgctgcatccggctgcagaaataaaaaggcctgcgattac
pac1-Sp225 KVF1	ggaaggcggagatgctcatcc

Name	Sequence
pac1-Sp225 KVR1	ccgcattgaaaatttatccgcgcg
pac1-Sp405 KF1	atataagttataaataagttaatgcaagtttaatgatttgatgtgatgtacagccaggacagaaatgcctcg
pac1-Sp405 KR1	cctctctaccggagccgtgcgctccggttttcccttcttttttacatcctgtgcagaaataaaaaggcctgcgattac
pac1-Sp405 KVF1	atttcagccactcgcgtggc
pac1-Sp405 KVR1	caccgcaaaatgtggtagcgattg
pac1-Sp585 KF1	tgcataaaaaagagcattcagttacctgaatgctctgaggctgatgacaacagccaggacagaaatgcctcg
pac1-Sp585 KR1	ggatgcggcgtgaacgccttttccggtctacgcattagacagttcttcgttgcagaaataaaaaggcctgcgattac
pac1-Sp585 KVF1	gccagtggattcaggtttcgctg
pac1-Sp585 KVR1	accagcctgctggttgcgc
pac1-Sp855 KF1	tttgaaagggcgaagatctgcaacggaaagatgatgtctttgttaaggcccagccaggacagaaatgcctcg
pac1-Sp855 KR1	aggatcagtgaaaatgtcatttttggcgtaattacccttactttatgcattgcagaaataaaaaggcctgcgattac
pac1-Sp855 KVF1	cagcccctggatcgttactgtctac
pac1-Sp855 KVR1	aggtagttcaataacctgatgtaccgagg
pac1-Sp1305 KF1	gatgggcgctttttttatttaatcgataaccagaagcaataaaaaatcaacagccaggacagaaatgcctcg
pac1-Sp1305 KR1	ttccatcggattcatcttagataaagtgagattatatagtgaaatccgattgcagaaataaaaaggcctgcgattac
pac1-Sp1305 KVF1	agcttgttaaagagatcacttctactgatgatttc
pac1-Sp1305 KVR1	gcgcccagtaccagtttagtagc
pac1-Sp45R KF1	acaacgtccatggagcaagttacagcgtaaaacgcataatattgctgcgccagccaggacagaaatgcctcg
pac1-Sp45R KR1	gaagccacaccaggcatataattattcgctacggcgagcaataatttttatgcagaaataaaaaggcctgcgattac
pac1-Sp45R KVF1	tgagcctcaggatcggtcgc
pac1-Sp45R KVR1	gacgagctgcgcgatcaaaac
aldh3 GF2	cggataacaatttcacacaggaaacaggaattcaaaaaaggaggtaaaaaatgattaagg
aldh3.adh2 GR1	ttcgatacgggttggattgcagtaggaaaagttgaccatctgttcctcctctggatttgttttgtctagattaacccgccagaacacaac
aldh3.adh2 GO1	ttttcctactgcaatccaacccgtatcgaattcggcaaaggtaaagaaaactccatcggt
aldh3.adh2 GO2	accgatggagttttctttacctttgccgaattcgatacgggttggattgcagtaggaaaa
aldh3.adh8 GR1	ggtcggtacgtggaacatgaagtcgtacatttattgccccccttgttgctatgagtctagattaacccgccagaacacaac
aldh3.adh12 GR1	cgggttggagtaggtaaagtcgcccagcatttctttcccccttaggaactagggttctagattaacccgccagaacacaac
aldh6 GF2	gcggataacaatttcacacaggaaacaggaattcaaaaaaggaggtaaaaaatgaaagag
aldh6.adh2 GR1	cgatacgggttggattgcagtaggaaaagttgaccatattggttacccccttagattaggatctagattaacgaatgctaaaggcgtcc
aldh6.adh2 GO1	acttttcctactgcaatccaacccgtatcgaattcggcaaaggtaaagaaaactccatcg
aldh6.adh2 GO2	cgatggagttttctttacctttgccgaattcgatacgggttggattgcagtaggaaaagt
aldh6.adh8 GR1	ggtcggtacgtggaacatgaagtcgtacatttggtccccctttgcgctaggattctctagattaacgaatgctaaaggcgtcc
aldh6.adh12 GR1	cgggttggagtaggtaaagtcgcccagcatggtgtcgtgtcctcctttatgtcggtctagattaacgaatgctaaaggcgtcc
aldh7 GF2	gagcggataacaatttcacacaggaaacaggaattcaaaaaaggaggtaaaaaatggaac
aldh7.adh2 GR1	ttcgatacgggttggattgcagtaggaaaagttgaccatcttatacctctttgtgtgtgaaagctctagattaaccggccagaacgcaac
aldh7.adh2 GO1	ttttcctactgcaatccaacccgtatcgaattcggcaaaggtaaagaaaactccatcggt
aldh7.adh2 GO2	accgatggagttttctttacctttgccgaattcgatacgggttggattgcagtaggaaaa
aldh7.adh8 GR1	ggtcggtacgtggaacatgaagtcgtacattaaaagtcctccccctgtgggttttctagattaaccggccagaacgcaac
aldh7.adh12 GR1	cgggttggagtaggtaaagtcgcccagcataagtatatctccttaagttccttggtctagattaaccggccagaacgcaac
aldh9 GF2	gcggataacaatttcacacaggaaacaggaattcaaaaaaggaggtaaaaaatgaatgac
aldh9.adh2 GR1	ttggattgcagtaggaaaagttgaccattggcggaagcctccctttgtatgatctagattacttaatagaaaagttgtcaaccataacg
aldh9.adh2 GO1	ttcctactgcaatccaacccgtatcgaattcggcaaaggtaaagaaaactccatcggtga
aldh9.adh2 GO2	tgccgaattcgatacgggttggattgcagtaggaaaagttgaccattggcggaagcctcc
aldh9.adh8 GR1	tcggtacgtggaacatgaagtcgtacatttagtgcccctccttaagattagcgtctagattacttaatagaaaagttgtcaaccataacg
aldh9.adh12 GR1	ggttggagtaggtaaagtcgcccagcatagtttagtacctccttcgttaggtatctagattacttaatagaaaagttgtcaaccataacg
aldh14 GF2	cggataacaatttcacacaggaaacaggaattcaaaaaaggaggtaaaaaatggaatttg
aldh14.adh2 GR1	atacgggttggattgcagtaggaaaagttgaccatgtgattcctcctagattcgggttcttctagattatttaatgctaaaaccgcccac
aldh14.adh2 GO1	caacttttcctactgcaatccaacccgtatcgaattcggcaaaggtaaagaaaactccat
aldh14.adh2 GO2	atggagttttctttacctttgccgaattcgatacgggttggattgcagtaggaaaagttg
aldh14.adh8 GR1	ggtcggtacgtggaacatgaagtcgtacataggattatcctccttctcgcgctagtctagattatttaatgctaaaaccgcccac
aldh14.adh12 GR1	cgggttggagtaggtaaagtcgcccagcatgcttgcctcctggtgtcttagagtatctagattatttaatgctaaaaccgcccac
CmR GF1	gacggacgcctttagcattcgttaataaggtaccaagcttttagccctgccactcatcgc
CmR GR1	gtctcacgcccggagcgtagcgaccgagtgagctagctacctgtgacggaagatcacttc
CmR GR2	tagctatcgccatgtaagcccactgcaagctgctagctacctgtgacggaagatcacttc
KmR GF1	tttattctgcgaagtgatcttccgtcacaggtagctagcagcttgcagtgggcttacatg
KmR GR1	gttaaatttttgttaaatcagctcatttttttaaccaatcagaagaactcgtcaagaaggc
CbR GF1	gccttcttgacgagttcttctgattggttaaaaaatgagctgatttaacaaaaatttaac
CbR GR1	tctcacgcccggagcgtagcgaccgagtgagctagcttaccaatgcttaatcagtgaggc
RFP BsaI GF1	gagcggataacaatttcacacaggaaacaggaattcgagacctccctatcagtgatagag
RFP BsaI GR1	gaccatgtattcctcgcttagagtttaatcttctagagagacctataaacgcagaaaggc
aldh1 shuffle GGR1	ttgatcggtctctctagattaaccggccagaacgcag
aldh2 shuffle GGR1	ttgatcggtctctctagattagcgaatgttcagcgcc
aldh3 shuffle GGR1	ttgatcggtctctctagattaacccgccagaacacaac
aldh4 shuffle GGR1	ttgatcggtctctctagattaaacaatgcgcagaccgc
aldh5 shuffle GGR1	ttgatcggtctctctagattaacgaatgcacagagaatcgg
aldh6 shuffle GGR1	ttgatcggtctctctagattaacgaatgctaaaggcgtcc

Name	Sequence
aldh7 shuffle GGR1	ttgatcggtctctctagattaaccggccagaacgcaac
aldh8 shuffle GGR1	ttgatcggtctctctagattaaccaacaaacacacagcgac
aldh9 shuffle GGR1	ttgatcggtctctctagattacttaatagaaaagttgtcaaccataacg
aldh10 shuffle GGR1	ttgatcggtctctctagattacacgatgcgaaaagagtccac
aldh11 shuffle GGR1	ttgatcggtctctctagattaaatgatacgcaaaccgccc
aldh12 shuffle GGR1	ttgatcggtctctctagattagcgaatagacaaggcgtcg
aldh13 shuffle GGR1	ttgatcggtctctctagattagccggccaggacgc
aldh14 shuffle GGR1	ttgatcggtctctctagattatttaatgctaaaaccgcccacc
aldh15 shuffle GGR1	ttgatcggtctctctagattatttaatgctcagaccaccgac
aldh46 shuffle GGR1	ttgatcggtctctctagattagccggccagaacacaac
aldh round 1 amp F1	ctggcaggttaccgtcacg
aldh round 1 amp R1	tttgttcatgtctttcagttctttcag
ampR GF1	atatctataccaaagacgttctggaag
colE1 GR1	ggtttccggcaccagaagc
aldh round 1 library GGF1	gctcagcggtctcgaattcaaaaaaggaggtaaaaaatg
aldh round 1 library GGR1	tgtatcggtctctctagaaaccacagcatgttctctgc
aldh round 2 amp F1	gcgccgacatcataacggttctggc
aldh round 2 amp R1	gttttttgcgccatattcgttcag
aldh round 2 library GGR1	tgtatcggtctcaccatgtattcctcgcttagagtttaatc
RFP v2 GO1	atggtcaacttttcctactgcaatccaacccgtatcgaattcggcaaaggtaaagaaaac
RFP v2 GO2	ggttggattgcagtaggaaaagttgaccatagagacctataaacgcagaaaggcccaccc
aldh GF1	ttgtgagcggataacaatttcacacaggaaacaggaattcaaaaaaggaggtaaaaaatg
(aldh3).adh2 GF1	ttaatctagacaaaacaaatccagaggaggaacagatggtcaacttttcctactgcaatc
adh GR1	gctcattaggcgggctcag
(aldh6).adh2 GF1	ttaatctagatcctaatctaaaggggtaaccaatatggtcaacttttcctactgcaatc
(aldh7).adh2 GF1	ttaatctagagctttcacacacaaagaggtataagatggtcaacttttcctactgcaatc
(aldh9).adh2 GF1	gtaatctagatcatacaaaagggaggcttccgccaatggtcaacttttcctactgcaatc
(aldh14).adh2 GF1	ataatctagaagaacccgaatctaggaggaatcacatggtcaacttttcctactgcaatc
(aldh3).adh8 GF1	ttaatctagactcatagcaacaaggggggcaataaatgtacgacttcatgttccacgtac
(aldh6).adh8 GF1	ttaatctagagaatcctagcgcaaaggggaccaaatgtacgacttcatgttccacgtac
(aldh7).adh8 GF1	ttaatctagaaaacccacaaggggggaggacttttaatgtacgacttcatgttccacgtac
(aldh9).adh8 GF1	gtaatctagacgctaatcttaaggaggggcactaaatgtacgacttcatgttccacgtac
(aldh14).adh8 GF1	ataatctagactagcgcgagaaggaggataatcctatgtacgacttcatgttccacgtac
(aldh3).adh12 GF1	ggttaatctagaaccctagttcctaaggggaaagaaatgctgggcgactttacctactc
(aldh6).adh12 GF1	cgttaatctagaccgacataaaggaggacacgacaccatgctgggcgactttacctactc
(aldh7).adh12 GF1	ggttaatctagaccaaggaacttaaggagatatacttatgctgggcgactttacctactc
(aldh9).adh12 GF1	aagtaatctagatacctaacgaaggaggtactaaactatgctgggcgactttacctactc
(aldh14).adh12 GF1	aaataatctagatactctaagacaccaggaggcaagcatgctgggcgactttacctactc
RFP BsaI GR2	gatctgctcatgtttgacagcttatcatcgatactaggagacctataaacgcagaaaggc
RFP BsaI GF2	gataacaatttcacacaggaaacaggaattcagagacctccctatcagtgatagagattg
RFP BsaI GR3	tgaccatgtattcctcgcttagagtttaatcttctaggagacctataaacgcagaaaggc
BsmBI GF1	gcttgtctgctcccggcatccgcttacagacaagctgtgacccactccgggagctgcatg
BsmBI GR1	gggccaggcggtgaagggcaatcagctgttgcccgtatcactggtgaaaagaaaaaccac
RFP BsmBI GF1	caatttcacacaggaaacaggaattcagagacgtccctatcagtgatagagattgacatc
RFP BsmBI GR1	catgtattcctcgcttagagtttaatcttctaggagacgtataaacgcagaaaggcccac
(aldh14).adh12 GGF1	gatacaggtctcaagacaccaggaggcaagcatgctgggcgactttacct
aldh14.(adh12) GGR1	gatacaggtctccgtcttagagtatctagattatttaatgctaaaaccgccca
(aldh14).adh8 GGF1	gatacaggtctcacgagaaggaggataatcctatgtacgacttcatgttccacg
aldh14.(adh8) GGR1	gatacaggtctctctcgcgctagtctagattatttaatgctaaaaccgccca
(aldh14).adh2 GGF1	gatacaggtctcaccgaatctaggaggaatcacatggtcaacttttcctactgcaa
aldh14.(adh2) GGR1	gatacaggtctcctcgggttcttctagattatttaatgctaaaaccgccca
aldh14 GGF1	gatacaggtctccattcaaaaaaggaggtaaaaaatggaatttg
(aldh9).adh2 GGF1	gatacaggtctccaagggaggcttccgccaatggtcaacttttcctactgcaa
aldh9.(adh2) GGR1	gatacaggtctcacctttttgtatgatctagattacttaatagaaaagttgtcaaccataacg
aldh9 GGF1	gatacaggtctcaattcaaaaaaggaggtaaaaaatgaatgactttaac
(aldh6).adh12 GGF1	gatacaggtctctaaggaggacacgacaccatgctgggcgactttacct
aldh6.(adh12) GGR1	gatacaggtctcccctttatgtcggtctagattaacgaatgctaaaggcgtcc
adh12 GGR1	gatacaggtctcgactagtttatttcatggattgtttcaggatg
adh8 GGR1	gatacaggtctcaactagtttacagagcagcacggtagatttc
(aldh6).adh2 GGF1	gatacaggtctcaaaaggggtaaccaatatggtcaacttttcctactgcaa
aldh6.(adh2) GGR1	gatacaggtctctctttagattaggatctagattaacgaatgctaaaggcgtcc
aldh6 GGF1	gatacaggtctctattcaaaaaaggaggtaaaaaatgaaagagg
adh2 GGR1	gatacaggtctccactagtttagtaagccaggttcagaatttcttc
aldh3 GGF1	gatacaggtctcgattcaaaaaaggaggtaaaaaatgattaaggac
(aldh3).adh8 GGF1	gatacaggtctcacaacaagggggcaataaatgtacgacttcatgttccacg

Name	Sequence
aldh3.(adh8) GGR1	gatacaggtctcagttgctatgagtctagattaacccgccagaacacaa
(aldh3).adh2 GGF1	gatacaggtctcctccagaggaggaacagatggtcaacttttcctactgcaa
aldh3.(adh2) GGR1	gatacaggtctcgtggatttgttttgtctagattaacccgccagaacacaa
aldh round 1 library GGF2	gatacaggtctcaattcaaaaaaggaggtaaaaaatg
aldh round 1 library GGR2	gatacaggtctcctctagaaaccacagcatgttctctgc
RFP adh2 GO1	ggtctcctagaagattaaactctaagcgaggaatacatggtcaacttttcctactgcaatccaacccgtatcgaattcggcaaaggtaaa
RFP adh2 GF1	gtgggcctttctgcgtttataggtctcctagaagattaaactctaagcgaggaatacatg
phaA GF1	taacaatttcacacagagctcaaggagatatacatatgactgatgttgtgattgtaagcg
phaA GR1	aaaacagccaagcttgcatgcctgcaggtcgactctagattatttgcgttcgacagccag
aldh7N C1 RFP GR1	agggatgtcaatctctatcactgatagggaggtctctgcaaccctcaatgatgttctttg
RFP aldh7N C1 GF1	tcgacaaagcggcaaagaacatcattgagggttgcagagacctccctatcagtgatagag
RFP aldh7C C1 GR1	ttcttgatcatgttctgaatcagatcattcgcgacatgagacctataaacgcagaaaggc
aldh7C C1 RFP GF1	gctcaccttcgggtgggcctttctgcgtttataggtctcatgtcgcgaatgatctgattc
aldh7N C2 RFP GR1	gatagggatgtcaatctctatcactgatagggaggtctcttgatgttctttgccgctttg
RFP aldh7N C2 GF1	ttccgccgacatcgacaaagcggcaaagaacatcaagagacctccctatcagtgatagag
RFP aldh7C C2 GR1	gttctgaatcagatcattcgcgacattctcgaacactgagacctataaacgcagaaaggc
aldh7C C2 RFP GF1	tggctcaccttcgggtgggcctttctgcgtttataggtctcagtgttcgagaatgtcgcg
aldh7N C3 RFP GR1	gatagggatgtcaatctctatcactgatagggaggtctctcgaagcttttggcattcttc
RFP aldh7N C3 GF1	tacgacgatcttcgtgaagaatgccaaaagcttcgagagacctccctatcagtgatagag
RFP aldh7C C3 GR1	acgggtgaagttacgtgcgctggtgatgccctcacctgagacctataaacgcagaaaggc
aldh7C C3 RFP GF1	ctggctcaccttcgggtgggcctttctgcgtttataggtctcaggtgagggcatcaccag
aldh7N C4 RFP GR1	tgatagggatgtcaatctctatcactgatagggaggtctctcacacccgcgaagcttttg
RFP aldh7N C4 GF1	tcttcgtgaagaatgccaaaagcttcgcgggtgtgagagacctccctatcagtgatagag
RFP aldh7C C4 GR1	cggcgctgacgggtgaagttacgtgcgctggtgatgtgagacctataaacgcagaaaggc
aldh7C C4 RFP GF1	ggctcaccttcgggtgggcctttctgcgtttataggtctcacatcaccagcgcacgtaac
aldh7N C5 RFP GR1	actgatagggatgtcaatctctatcactgatagggaggtctctcaccgaaacccacaccc
RFP aldh7N C5 GF1	gaatgccaaaagcttcgcgggtgtgggtttcggtgagagacctccctatcagtgatagag
RFP aldh7C C5 GR1	cagaacgcaacggcgctgacgggtgaagttacgtgctgagacctataaacgcagaaaggc
aldh7C C5 RFP GF1	ggctcaccttcgggtgggcctttctgcgtttataggtctcagcacgtaacttcacccgtc
aldh7N C6 RFP GR1	tagggatgtcaatctctatcactgatagggaggtctcttcagcagggtcttaatcaaacc
RFP aldh7N C6 GF1	caccggtggtccgggtttgattaagaccctgctgaagagacctccctatcagtgatagag
RFP aldh7C C6 GR1	gatgttctttgccgctttgtcgatgtcggcggaatctgagacctataaacgcagaaaggc
aldh7C C6 RFP GF1	tggctcaccttcgggtgggcctttctgcgtttataggtctcagattccgccgacatcgac
aldh7 C1 F	ttgcagcttcgacaacaatctgccgtgtnnknnknnkaaagaggtttttgtgttcgagaa
aldh7 C1 R	gacattctcgaacacaaaaacctctttmnnmnnmnnacacggcagattgttgtcgaagct
aldh7 C2 F	atcattgagggttgcagcttcgacaacnnknnknnktgtatcgcggaaaaagaggttttt
aldh7 C2 R	acacaaaaacctcttttccgcgatacamnnmnnmnngttgtcgaagctgcaaccctcaa
aldh7 C3 F	ttcgcgggtgtgggtttcggtgcagaannknnknnkaccttcaccatcgctggcccgacc
aldh7 C3 R	caccggtcgggccagcgatggtgaaggtmnnmnnmnnttctgcaccgaaacccacacccg
aldh7 C4 F	tgtgggtttcggtgcagaaggctttacgnnknnknnkatcgctggcccgaccggtgaggg
aldh7 C4 R	gatgccctcaccggtcgggccagcgatmnnmnnmnnncgtaaagccttctgcaccgaaacc
aldh7 C5 F	cagaaggctttacgaccttcaccnnknnknnkccgaccggtgagggcatcaccagc
aldh7 C5 R	gtgcgctggtgatgccctcaccggtcggmnnmnnmnnnggtgaaggtcgtaaagccttctg
aldh7 C6 F	ctgaactccggtaagaaagcgattggcnnknnknnkggtaatccgccggttattgttgat
aldh7 C6 R	aatcatcaacaataaccggcggattaccmnnmnnmnngccaatcgctttcttaccggagt
StrepTEV_aldh7 GF1	tcgaaaagggtgcaggtgagaatctctacttccagggtgccatggaacgcaacttgtcgg
StrepTEV_aldh7 GO1	actgcgggtggctccagcttgccatatgtatatctccttcttaaagttaaacaaaattat
StrepTEV_aldh7 GO2	tacatatggcaagctggagccacccgcagttcgaaaagggtgcaggtgagaatctctact
aldh3.(adhX) GGR1	gatacaggtctcttagattaacccgccagaacacaa
aldh6.(adhX) GGR1	gatacaggtctcttagattaacgaatgctaaaggcgtcc
aldh7.(adhX) GGR1	gatacaggtctcttagattaaccggccagaacgc
aldh46.(adhX) GGR1	gatacaggtctcttagattagccggccagaacacaac
pET28-RFP GF1	tgtttaactttaagaaggagatatacatatgagagacctccctatcagtgatagagattg
pET28-RFP GR1	tttcgggctttgttagcagccggatcctcgagtcagggagacctataaacgcagaaaggc
pac1-Cm45 KF1	tcccgcctgctgatattgaaactggctgcgtctcgcgcgctcccgtcagagtgtaggctggagctgcttc
pac1-Cm45 KR1	ttcatggttgtttttacgggtggttaactgagcggcgaatgttaacacaacatatgaatatcctccttagttcctattc
pac1-Sp45-2 KF1	tcccgcctgctgatattgaaactggctgcgtctcgcgcgctcccgtcagaccgccgcagtctcacgc
pac1-Sp45-2 KR1	ttcatggttgtttttacgggtggttaactgagcggcgaatgttaacacaattatttgccgactaccttggtg
pac3-Sp45 KF1	aaactttaccgtgccctaatacgacaaaagcccagactttgcagcctggacccgccgcagtctcacg
pac3-Sp45 KR1	tttacttcttcatgccaaaagggagctatctcccttgtttgaattgaaaattatttgccgactaccttggtg
pac3-Sp45 KVF1	gcagattgttgtgcaccac
pac3-Sp45 KVR1	ggaaatcagcatcgtggg
pac1 GF	aaacggcaattttcgtctgaatgcg
pac1 GR	aaaacgcattcagacgaaaattgcc
laclq Ptrc GF1	aaattcgaaaagcctgctcaacgagcaggcttttttgaattccagatcaattcgcgcgcg

Name	Sequence
laclq Ptrc GR1	gccagcgaggaggctgggaccatgccggccatcgtctgtttcctgtgtgaaattgttatc
KmR sgRNA GF1	tgtggaattgtgagcggataacaatttcacacaggaaacagacgatggccggcatggtcc
KmR sgRNA GR1	aaataggcgtatcacgaggccctttcgtcttcacctcgagaaaaaagcctgctcgttgag
laclq Ptrc GF2	aaagtatgtaaatagacctcaactgaggtctttttttgaattccagatcaattcgcgcgc
laclq Ptrc GF3	aagcgggggttttttgtctctgcttttgccgctttgaattccagatcaattcgcgcgc
pKD46 GF1	tggtttatgcagtgacggc
pKD46 GR1	ggacgacgtggtgttagctgtgcatggattcttcgtctgtttctac
cas9 GF1	aataccagtagaaacagacgaagaatccatgcacagctaacaccacgtcg
cas9 GR1	gtagataacaaatacgattcttccgac
cas9 GR2	atcaggatttaaatctccctcaatc
fucO gRNA R	aaaactgccgtttcgttcagaatca
fucO gRNA F	aaactgattctgaacgaaacggcag
ahr gRNA R	aaaaccttcttttgcggcatagctt
ahr gRNA F	aaacaagctatgccgcaaaagaagg
yiaY gRNA R	aaaactctgtgaatgtcatcggcgc
yiaY gRNA F	aaacgcgccgatgacattcacagag
eutG gRNA R	aaaaccgctctttcaggcgttcgat
eutG gRNA F	aaacatcgaacgcctgaaagagcgg
adhP gRNA R	aaaactgtttcagtgagcgcagtgt
adhP gRNA F	aaacacactgcgctcactgaaacag
yqhD gRNA 1 R	aaaacaacagaatgcgggttggggt
yqhD gRNA 1 F	aaacaccccaacccgcattctgttg
yqhD gRNA 3 R	aaaacaacccgcattctgtttggta
yqhD gRNA 3 F	aaactaccaaacagaatgcgggttg
yqhD gRNA 2 R	aaaacgcattctgtttggtaaaggc
yqhD gRNA 2 F	aaacgcctttaccaaacagaatgcg
fucO KVR1	aaaaagcgccacgcagctg
fucO KVF1	ggctctcaaaaatcgtaaggc
ahr KVR1	tggcgactatgaaactattctcg
ahr KVF1	gtatttctacggttgattttgaatcaac
yiaY KVR1	ataaagttgtcatagcgggcttttaac
yiaY KVF1	tacgcgtcgtcgtggaaatg
eutG KVR1	tcatgatgtacatgatgatttcgttaattc
eutG KVF1	ggtcacggtgaagagatttgg
adhP KVR1	gggagcaatttccattgctc
adhP KVF1	aaaatggcgcagcagcaag
yqhD cut repair 90 R	caagtaatgaacaactttaatctgcacaccccaacccgcattctggcttttacgcctcaaactttcgttttcgggcatttcgtccagac
yqhD cut repair 90 F	gtctggacgaaatgcccgaaaacgaaagtttgaggcgtaaaaagccagaatgcgggttggggtgtgcagattaaagttgttcatta cttg
yqhD KVR1	aactttatccgccagcaagc
yqhD KVF1	gcggcggtatcaatcgag
yqhD clean repair 90 R	agatcgttctctgccctcatattggcccagcaaagggagcaagtagctttttacgcctcaaactttcgttttcgggcatttcgtccagac
yqhD clean repair 90 F	gtctggacgaaatgcccgaaaacgaaagtttgaggcgtaaaaagctacttgctccctttgctgggccaatatgagggcagagaac gatct
fucO clean repair 120	ttcaaaacctatgggttacgaattgaagagtaatttcgtaaagcaacaaggagaaggatgatgcgctgatgtgataatgccggatac gacgtttgcggcgtatccggctcgccgttaccg
ahr clean repair 120	caggcaccccgccctgccatgctctacacttcccaaacaacaccagagaaggaccaaaaaaaatcattcgcagcgctgatctga ggcgctgccctctttcgcacatattctgttttgtcg
yiaY clean repair 120	ttttcgaatgaaaatcctcagtaagctgcccgccctttttttacactttcaggagtgtgtttcatcatttccacaacggctggcaaattgttag ccgctttttcaactatctctgtaaccc
eutG clean repair 120	accccgctggcgatcgccagtagtgggcgagagaaagcggaggggctatatgcaaaatgaatgccggatgcgacgcttgccgc atcttatccagcctacgggattgcacatgtagggcgg
adhP clean repair 120	ggcgaataaatctcatttgcctcacctgctatgcagaacatcatccgaaaaggaggaactgaggcctttgctgcgactgccatgttcg ggtcgcagcatcgcacactctccaacatgaaa
yqhD clean repair 120	gtgttgtgaacttaagtctggacgaaatgcccgaaaacgaaagtttgaggcgtaaaaagctacttgctccctttgctgggccaatatg agggcagagaacgatctgcctgatgtttttca
yqhD cut repair 120	gtgttgtgaacttaagtctggacgaaatgcccgaaaacgaaagtttgaggcgtaaaaagccagaatgcgggttggggtgtgcagat taaagttgttcattacttgctccctttgctgggc
yqhD cut repair 150	gcaggctccggtgaggtgttgtgaacttaagtctggacgaaatgcccgaaaacgaaagtttgaggcgtaaaaagccagaatgcgg gttggggtgtgcagattaaagttgttcattacttgctccctttgctgggccaatatgagggcaga
yqhD cut repair 200	tcctcctttatatgaactcaccggagcaggctccggtgaggtgttgtgaacttaagtctggacgaaatgcccgaaaacgaaagtttga ggcgtaaaaagccagaatgcgggttggggtgtgcagattaaagttgttcattacttgctccctttgctgggccaatatgagggcagag aacgatctgcctgatgtttttcat
yqhD clean repair 150	gcaggctccggtgaggtgttgtgaacttaagtctggacgaaatgcccgaaaacgaaagtttgaggcgtaaaaagctacttgctccct ttgctgggccaatatgagggcagagaacgatctgcctgatgtttttcattgtgatcgccagcg
yqhD clean repair 200	tcctcctttatatgaactcaccggagcaggctccggtgaggtgttgtgaacttaagtctggacgaaatgcccgaaaacgaaagtttga ggcgtaaaaagctacttgctccctttgctgggccaatatgagggcagagaacgatctgcctgatgtttttcattgtgatcgccagcgcc

Name	Sequence
	ctggctctcaatgctcatttctg
pacWT gRNA R	aaaactttgacatgtgactttcgtt
pacWT gRNA F	aaacaacgaaagtcacatgtcaaag
pac gRNA R	aaaacgggtgtcgcgttccaggcga
pac gRNA F	aaactcgcctggaacgcgacacccg
pacWT KVR1	ttttagtggcttcctgactgg
pacWT KVF1	acacgcccacgtagatttcag
pac KVR1	tcaggcggagtaccttcatcatc
pac KVF1	ttgacatgtgactttcgttaccc
pacWT-FRT R1	t*g*taaccaccatctcgcg
pacWT-FRT F1	c*a*gccttgggtaatggcg
pacWT R1	g*a*tggtgtaaccaccatctcg
pacWT F1	a*g*caccagccttgggtaatg
pac R1	a*g*tggtttccagaacatcgg
pac F1	t*a*tgccgccgagtttaatct
pSV272 aldh7 GF1	acaacaacctcgggatcgaggaaaacctgtattttcagggcatggaacgcaacttgtcgg
pSV272 aldh7 GR1	gtgcggccgcaagcttgtcgacggagctcgaattcggggatccttaaccggccagaacgc
pCW HisTEV aldh7 GF1	aacaggatccatcgatgcttaggaggtcatatgcatcatcatcatcatcatgagaatctc
pCW HisTEV aldh7 GR1	gctcagatctgctcatgtttgacagcttatcatcgataagcttttaaccggccagaacgc
sadh1 GF1	caggaaacaggatccatcgatgcttaggaggtcatatgttcatgaaaggtctgacgtatc
sadh1 GR1	aaaatcttctctcatccgccaaaacagccctcgagttacggcgtcagaataattttcatc
sadh2 GF1	ggaaacaggatccatcgatgcttaggaggtcatatgtctaagaactacaaaaagcaccag
sadh2 GR1	atcttctctcatccgccaaaacagccctcgagttagtattccggtttctggtaaatagag
sadh3 GF1	cacaggaaacaggatccatcgatgcttaggaggtcatatgaaaatctctctgacttccgc
sadh3 GR1	gctgaaaatcttctctcatccgccaaaacagccctcgagttaagcggtcggtactgcgac
sadh4 GF1	cacacaggaaacaggatccatcgatgcttaggaggtcatatgaagggtttcgctatgctg
sadh4 GR1	gctgaaaatcttctctcatccgccaaaacagccctcgagttacgccaggatgactaccgg
sadh5 GF1	cacacaggaaacaggatccatcgatgcttaggaggtcatatgaaaggcttcgctatgctg
sadh5 GR1	cttctctcatccgccaaaacagccctcgagttacagaattacaacagctttgatcagatc
sadh6 GF1	acaatttcacacaggaaacaggatccatcgatgcttaggaggtcatatgcgcgcactggc
sadh6 GR1	gctgaaaatcttctctcatccgccaaaacagccctcgagttagttggtcgcgtccagttc
sadh7 GF1	acacaggaaacaggatccatcgatgcttaggaggtcatatgaaagcagccgtgttcaaag
sadh7 GR1	gaaaatcttctctcatccgccaaaacagccctcgagttacggcttgatggctactttcag
sadh8 GF1	ttcacacaggaaacaggatccatcgatgcttaggaggtcatatgaccgcgatgatgaagg
sadh8 GR1	aggctgaaaatcttctctcatccgccaaaacagccctcgagttagtgcggtttgatggcg
sadh9 GF1	tttcacacaggaaacaggatccatcgatgcttaggaggtcatatgaccttcgaactgccg
sadh9 GR1	aatcttctctcatccgccaaaacagccctcgagttacagtttgtcgttgtattcaatcag
sadh10 GF1	ttcacacaggaaacaggatccatcgatgcttaggaggtcatatgatggcgaccatgaaag
sadh10 GR1	ctgaaaatcttctctcatccgccaaaacagccctcgagttacggggtgattgcaactttc
sadh11 GF1	ttcacacaggaaacaggatccatcgatgcttaggaggtcatatgtctaaagtggctgccg
sadh11 GR1	gctgaaaatcttctctcatccgccaaaacagccctcgagttaatgaaattgcatgccacc
sadh12 GF1	cacacaggaaacaggatccatcgatgcttaggaggtcatatgcgtgcgctggcctatttc
sadh12 GR1	ctgaaaatcttctctcatccgccaaaacagccctcgagttatttcatttcaccgtggttg
sadh13 GF1	tttcacacaggaaacaggatccatcgatgcttaggaggtcatatggaccgcatcatccag
sadh13 GR1	aggctgaaaatcttctctcatccgccaaaacagccctcgagttattcccattcctgcagg
sadh14 GF1	acacaggaaacaggatccatcgatgcttaggaggtcatatgatgaaagcgctgtgttacc
sadh14 GR1	ctgaaaatcttctctcatccgccaaaacagccctcgagttattcgtcacaggtgatcacc
sadh15 GF1	tttcacacaggaaacaggatccatcgatgcttaggaggtcatatgagcattccgtccagc
sadh15 GR1	tctctcatccgccaaaacagccctcgagttacgggttaaagacaacacgaccctcatacg
sadh16 GF1	ttcacacaggaaacaggatccatcgatgcttaggaggtcatatgctgcgcactaatctgc
sadh16 GR1	aatcttctctcatccgccaaaacagccctcgagttatttggtagtatccacgacgtaacg
adh15 GGF1	gatacaggtctcgtctagaagagatc
adh15 GGR1	tgtatcggtctctactagtttagctacg
adh16 GGF1	gatacaggtctcgtctagaaagagtaataag
adh16 GGR1	tgtatcggtctctactagtttaatcgttag
adh17 GGF1	gatacaggtctcgtctagatagccag
adh17 GGR1	tgtatcggtctctactagtttacagtgc
adh18 GGF1	gatacaggtctcgtctagacacgac
adh18 GGR1	tgtatcggtctctactagtttattttgc
adh20 GGF1	gatacaggtctcgtctagacataaagc
adh20 GGR1	tgtatcggtctctactagtttacagagc
adh22 GGF1	gatacaggtctcgtctagattaaaatcc
adh22 GGR1	tgtatcggtctctactagtttacaggg
aldh17 GF1	tgtgtggaattgtgagcggataacaatttcacacaggaaacagaattcaaaaaaggaggtaaaaaatgaacaaagatactaccatcagcg
aldh17.(adh2) GR1	ttgaccatcttatacctctttgtgtgtgaaagctctagattaaccagccagaacacagcg

Name	Sequence
(aldh17).adh2 GF1	tcacgcgtcagcgtcgctgtgttctggctggttaatctagagctttcacacacaaagagg
aldh17.(adh8) GR1	tcgtacattaaaagtcctccccccttgtgggttttctagattaaccagccagaacacagcg
(aldh17).adh8 GF1	caacttcacgcgtcagcgtcgctgtgttctggctggttaatctagaaaacccacaagggg
aldh18 GF1	atgtgtggaattgtgagcggataacaatttcacacaggaaacagaattcaaaaaaggaggtaaaaaatgaacaataacctgttcgt tagc
aldh18.(adh2) GR1	gaccatcttatacctctttgtgtgtgaaagctctagattaacccacgaagacacaacgac
(aldh18).adh2 GF1	ttactcgtcagcgtcgttgtgtcttcgtgggttaatctagagctttcacacacaaagagg
aldh18.(adh8) GR1	gtacattaaaagtcctccccccttgtgggttttctagattaacccacgaagacacaacgac
(aldh18).adh8 GF1	aactttactcgtcagcgtcgttgtgtcttcgtgggttaatctagaaaacccacaaggggg

Appendix 3: *Oligonucleotides used for sequencing*

Table A3.1 Oligonucleotides used for sequencing

Name	Sequence
pKD3 SF1	gcagaaggccatcctgacgg
pKD3 SR1	gatggcttccatgtcggcag
pKD3 SR2	cctctcaaagcaattttcagtgacacagg
sacB SF1	gccatataaggaaacatacggcatttccc
sacB SF2	ctgaagatggctaccaaggcgaag
pPro18 SF1	gccggtgatgccggccac
pPro18 SF2	cggcgcgacctccgcgtc
pPro18 SR2	ctgttttatcagaccgcttctgcg
pBBR2 KmR SF1	gacggcaggctgtcggcc
pBBR2 KmR SR1	ccagtcgcgagtttcgatcgg
pBBR2 KmR SF2	cttctatcgccttcttgacgagttcttctga
pCOLA SF1	cacctgaagtcagccccatacg
pCOLA SR1	gcattatgcggccgcaagc
pCOLA SF2	ccacagccaggatccgaattcg
pCOLA SR2	ccatgtgctggcgttcaaatttcg
bmoR SF1	cgaccttgccgatgttgcc
FNR SF1	cgtgatggctgtcacgcgg
FNR SR1	ccagttaatcaaatcacgcatacgcgc
pCOLA SF3	gtttacagagcaggagattacgacgatcg
pCOLA SR3	gccactcgaaccgggctc
RSF ori F1 SOE	gcggacacatacaaagttacccacagcttccgcttcctcgctcactg
RSF ori R1	gcatcctgaggaacggaatagctgttcgttgacttgatagacc
Clo DF13 SF1	gcatttatcagggttattgtctcatgagcgg
Clo DF13 SR1	gactaacatgagaattacaacttatatcgtatggggc
pCDF-TA SF1	caaaagattacgcgcagaaaaaaaggatctcaag
pCDF-TA SR1	cgcttatgtctattgctggtttaccgg
pCDF-TA SF2	gtcagagacatcaagaaataacgccgg
pCDF-TA SF3	cacaggaaacaggatcgatccatcgatg
pCDF-TA SF4	cggcgttctgcacaaggc
pCDF-TA SF5	ccggttgacgttgaagcgg
pCDF-TA SF6	gccggtaacaccccggcg
pCDF-TA SF7	cgccgtgaaaacatgctgtgg
pRSF SF1	caaacccgcgttacccgc
pRSF SR1	cgaatggcgctttgcctgg
pCDF-TA SF2	gcaaacccgcgttacccg
pCDF-TA SR2	ggcgctttgcctggtttcc
pCDF-TA SF3	ctcatactcttccttttttcaatattattgaagcatttatcaggg
pPro18 SF3	ggcaattgtggcacacccc
pTet SF1	gtagatcctctagagtcgactaagaaaccattattatcatgac
pTet SR1	gatgatgatggtcgacggcgc
pMDxxx SF1	gcggccgcaggtggcac
pMDxxx SF2	ggatggaggcggataaagttgcag
pMDxxx SF3	tcgggctgaacgggggg
pMDxxx SF4	tggagagcgccgccagag
pMDxxx SF5	ttcttctgccagatcggtaatcagcc
pMDxxx SF6	gagcaggaagacgataagcagtactcc
pMDxxx SF7	ccgtccaggcctggctg
pMDxxx SF8	tgcaggcccggcgccatatc
pMDxxx SF9	cacccgcataaccgctgatgtc
pMDxxx SF10	gtagcgcgctgttacgcg
pMDxxx SF11	ctgctggccggttgcgtc
pMDxxx SF12	gccaggttttcaccgtaacacgc
lpd SF1	caaagacatcgttaaagtcttcaccaagcg
lpd SR1	gccgtcttctttcgcttcaacgg
pBAD/BT Km SF1	ggcaccaataactgccttaaaaaaaattacgc
pBAD/BT Km SR1	ccggcatcaccggcgcc
pBAD/BT Km SF2	gcgcatgcccgacggcg
pBAD/BT Km SR2	tgcatcgatgataagctgtc
pMDxxx SF13	aatatatgtgtagaaactgccgg
pMDxxx SF14	gacatccttcggcgcg
pMDxxx SF15	caagacgcgtgctgaagtc
pMDxxx SF16	cgcgagatcttgcaacttttgg

Name	Sequence
pMDxxx SF17	gccgatgttgcctttgcg
pMDxxx SR1	cctttgatcttttctacgggg
EcADH SF1	ggttccccgatggacgcc
pMDxxx SR2	cgccgcacggcagttggggg
pMDxxx SF18	cggcgccgcggtgcaggtg
ALDH46 SF1	gaaaaggcgggtcgttccatc
pCWori SR2	gagcacctaagaaaccattattatcatgac
aldh46 SF2	catggaatccctggacgc
pMD13x SF1	gacagtcattcatctttctgcc
pMD13x SF2	gagcatggcaagcgctgc
pMD13x SF3	gccgcggtgccggccgag
pMD13x SF4	atcatcctgggggcgtcg
pMD13x SF5	gaaagcgttagagcggaatc
pMD13x SF6	ctgtccacacaatctgccc
pMD13x SF7	ctgcgccttatccggtaac
pMD13x SF8	ccagccggaagggccgag
pMD13x SF9	attattgaagcatttatcagggtt
pMD13x SF10	ggtcactacgacgctgaag
pMD13x SR1	gggcgagccgcactacagcg
aldh46 SR1	caacagcaaaggcaacacac
phaA SF1	cgcaagcggtctgaacg
bdhB SR1	agcaatagtcagaatggaagc
fucO SR1	ataaacgccacctgcgg
yqhD SR1	gtgtaggtataaaccggatcg
aldh46 SR2	catacctggaccaccgg
pBu2 SF1	gctttttatcgcaactctctactg
pBu2 SF2	aaggcgacccggtggcc
pBu2 SF3	gttgatcgcggcctggac
pBu2 SR1	ggctgatcattaactatccgc
pBu2 SR2	gacctcctaagcatcgatg
pBu2 SR3	gcccttacgatacagtgcc
pBu2 SR4	gccagcacctctttattctg
pBu2 SR5	gaagtactcttgggtttcgttg
pBu2 SF4	cattaacgaggaccaagttagc
aldh4 SF1	cgggtaatccgccggtg
aldh10 SF1	ccggctgcagctttgac
adhE2 SR1	ccgggtattcgtacagcag
aldh4 SR1	ggatgcacattgaacgtc
aldh10 SR1	gctcgtaaacgacgcgg
aldh10 SR2	cagtttcaagtatttgtggtcg
aldh10 SR3	gcgaaaagagtccaccatc
phaB SF1	gtgcatggctgtcttccg
HBD SF1	gcacacgctgctgaaaaag
pMut SF1	caattcagcgcattgtacgtc
pMut SF2	gctttacgcagacatgagc
pMut SF3	agcagggatatcattaacgc
pMut SF4	cgccctggcgcgttctg
aldh3 SF1	acgcaattatcaaacacccgtcc
aldh6 SF1	ggaagagccgtctattgagaacac
aldh7 SF1	gcacccgtacatcaagctgc
aldh9 SF1	gtgtcattggtgcgatcg
aldh10 SF1	cgcacccaggtgcgctg
aldh12 SF1	cgcacccgcgtgcgaag
aldh15 SF1	caattctgttgttttcgcgc
pMut SF5	cgacatccttcggcgcg
pMut SF6	caccataagcattttccctac
pMut SF7	gaaaagcccgtcacgggc
pMut SF8	cagtaacaattgctcaagcag
aldh46 SF3	gtcgttttcaatggccac
aldh46 SF4	cattgcaggttctaccgg
pRecA SF1	atggctatcgacgaaaacaaac
pRecA SF2	aaaatcgctgcgccgtttaaacag
pRecA SF3	gcgatcatggcgaccac
pRecA SF4	ggatcgctcgcggctcttac
pRecA SF5	tatttgccttgtgagttttcttttg
pRecA SF6	ataactcactactttagtcagttccg
pRecA SF7	caaaaccaacattgcgac
pRecA SF8	tttgcgcttcagccatac

Name	Sequence
cI857 SR1	acagtacccaatgatcccatgcaatgagag
cI857 SF1	tcgtagatttctctggcgattgaagggc
flp SF1	cagcgatattaagaacgttgatccgaaatc
flp SF2	acgataccccgcatggaatgggataatatc
mutS SF1	cgtatcgttacgccaggcaccatc
mutS SF2	acgtctggctttacgaactgctcg
mutS SF3	cggtagttgaacaagtactgaatgagcc
mutS SF4	gctctggaaaatcttgatccggattcac
bet SF1	gatgaatgccgccgcgaacc
TcR SR1	gatttcatacacggtgcctgactgcg
mutS SR1	cacccatccggtaaaacagcaggatc
aldh46 SF5	cgggtgtgggctatgaggc
pRecA SF9	tgaccacttcggattatcccgtgac
pRecA SR1	gaaagggcctcgtgatacgcc
tolC SF1	gtgttgaatgctattgacgttctttcctatacac
pRecA SF10	agaagaaagtacgtgagttgctgctg
pRecA SR2	gagtgacatgcaaagtaagtatgatctcaatg
adh2 SR1	gttttttgcgccatattcgttcag
adh8 SR1	cgctttctggccaaaatcagac
adh12 SR1	cgttcttgccatagtttttcagttc
aldh14 SF1	atcaaaatgctggttgctaccg
adh2 SR2	ccagttcggagaatttgatgc
aldh3 SF2	cgaaggctttaccacgttcac
aldh6 SF2	cttcactatcgcaggccc
aldh9 SF2	cttcaccattgccggtc
aldh14 SF2	gtttaccattgcaggccc
RFP SF1	tcgtcactccaccggtgc
rrnB SR2	cagaccgcttctgcgttctg
aldh7 SR1	gcgctccagcagaaccag
RFP SR1	cgggatgtcagccgggtg
aldh7 SR2	cacggtttcagtaggattggtg
RFP SR2	cccaacccatggttttttc
aldh7 SR3	cagaatcggcatcatcaattc
RFP SR1-2	gggaaggacagtttcaggtagtc
aldh7 SF2	agaaggctttacgaccttcacc
cas9 SR2	cacttgcgttaatagggttttcttc
RFP SF1	tcgtcactccaccggtgc
pGuide SF1	gccgaacgccctaggtctag
cas9 SR1	tccaccaaaaaagactcttcaag
pKD46 SF1	cacttccggcgtgaatgttac
pSV272 SF1	gtcgtcagactgtcgatgaagc
sadh1 SR1	ccttcacctcgccattcag
sadh2 SR1	tcagctcttttgcgatttttc
sadh3 SR1	ggttagtagtcagacgggcaac
sadh4 SR1	tatcagccagttcagcaccg
sadh5 SR1	cagccagctctgcaccg
sadh6 SR1	cagcaccagagcagattgg
sadh7 SR1	tggaccggagaaaccagtg
sadh8 SR1	ctccgcacctttgaaacc
sadh9 SR1	atcacccagcttaacctccg
sadh10 SR1	tgcgccagagaaaccag
sadh11 SR1	gactacctggtcacggtcagaaac
sadh12 SR1	tgcggaagaacctttttg
sadh13 SR1	ctcgaaccaagtggccag
sadh14 SR1	gccagagacaacgccc
sadh15 SR1	gtcggcggaaacgttatc
sadh16 SR1	aacagtgacaccagcgcac

Appendix 4: *Linear DNA used for plasmid and strain construction*

Table A4.1 Linear DNA used for plasmid and strain construction

dhaT G1
CCATTGCAGGTTCTACCGGTGAAGGTATCACGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCG
GCTAACTCGAGCCGGAATAAGGAGGGGACCTATGAGCTACCGCATGTTTGACTATCTGGTACCGAACGTTAACTT
CTTTGGTCCAAACGCAATCAGCGTTGTGGGTGAACGTTGTCAGCTGCTGGGCGGCAAAAAAGCGCTGCTGGTTAC
TGACAAAGGCCTGCGTGCGATCAAAGATGGTGCGGTGGATAAAACCCTGCACTACCTGCGCGAAGCAGGCATCG
AAGTCGCTATCTTCGACGGTGTAGAGCCGAACCCGAAAGATACTAACGTACGCGATGGTCTGGCTGTATTCCGTC
GCGAACAGTGTGACATCATTGTGACCGTTGGCGGTGGTTCCCCGCACGACTGCGGCAAAGGTATCGGCATTGCG
GCTACTCACGAAGGTGACCTGTACCAGTATGCCGGTATCGAAACTCTGACCAACCCGCTGCCTCCGATTGTTGCA
GTTAATACTACCGCCGGTACGGCGAGCGAAGTTACCCGTCACTGCGTACTGACCAACACTGAAACCAAGGTAAAA
TTTGTGATTGTCTCTTGGCGTAACCTGCCGTCTGTCTCCATTAACGACCCTCTGCTGATGATTGGCAAACCGGCCG
CACTGACCGCAGCAACTGGCATGGATGCTCTGACTCACGCG

dhaT G2
CCATTAACGACCCTCTGCTGATGATTGGCAAACCGGCCGCACTGACCGCAGCAACTGGCATGGATGCTCTGACTC
ACGCGGTGGAGGCTTACATCTCTAAAGACGCAAACCCGGTTACTGACGCCGCAGCTATGCAGGCTATCCGTCTGA
TTGCCCGTAACCTGCGCCAGGCTGTCGCTCTGGGCTCTAACCTGCAGGCTCGTGAATATATGGCTTACGCTTCTC
TGCTGGCGGGTATGGCGTTCAACAACGCGAACCTGGGTTATGTACATGCGATGGCCCACCAGCTGGGCGGCCTG
TACGACATGCCGCACGGCGTAGCAAACGCGGTACTGCTGCCGCATGTTGCACGTTATAACCTGATCGCAAACCCA
GAGAAATTCGCTGACATCGCCGAACTGATGGGCGAAAACATTACTGGCCTGTCTACGCTGGACGCGGCGGAAAA
AGCAATCGCGGCGATCACCCGTCTGAGCATGGACATTGGTATCCCGCAGCATCTGCGCGACCTGGGTGTAAAAG
AAACCGATTTCCCGTATATGGCTGAAATGGCGCTGAAAGACGGCAACGCGTTCTCTAACCCACGTAAAGGTAATG
AACAGGAAATCGCGGCTATTTTCCGCCAGGCGTTCTAAGAGCTCAAGCTTATCGATGATAAGCTGTCAAACATGAG
CAGATCTGGATCCGCCTAATGAGCGGATCTTTTTTTCAGATC

aldh46.x G1
ATCGCTTCGAACGTGAGATTGACACCACGATTTTTGTGAAAAACGCAAAAAGCTTTGCGGGTGTGGGCTATGAGG
CGGAAGGCTTCACCACCTTTACCATTGCAGGTTCTACCGGTGAAGGTATCACGAGCGCCCGTAACTTCACGCGCC
AACGTCGTTGTGTTCTGGCCGGCTAATCTAGATGAACAACTTCCGTTTCTGCACCCTACCGAATTACTAGTATCGA
TGATAAGCTGTCAAACATGAGCAGATCTGAGCCCGCCTAATGAGC

adh1 G1
CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGACTCGCCATATCCGACCCAC
CCAAGGACAACTCATATGAACAACTTCCGTTTCTGCAGCCCTACCGAATTCATTTTTGGTAAAAACACCATCTGTAA
AGTGGCTCAGCTGGTTAAACAGTATGGTGGCTCTAAAGTTCTGATCCATTACGGCAATAAATCTGCGAAAAAATCT
GGTCTGCTGACCCAGATCGAGAACTGCTTCCAGAACGAATTTATCGAATATGTCAAACTGGGTGGTGTTCAGCCG
AACCCGATCGACGAACTGGTCTACAAGGGTATCGAACTGGGCCGTAAAGAAAAAGTTAACTTCATCCTGGCTATC
GGTGGCGGTAGCGTTATCGACTCTGCTAAAGCAATCGCTGCGGGCATTCTGTACAACGGTGATTTCTGGAACTTT
TTCGAAGGCATCGTTACCATTAACCACGCCCTGCCAATTGCAACTGTTCTGACCCTGCCTGCTGCGGGCTCTGAG
GGTTCTCCGAACACTGTCATCACGAAAACCGACGGTATGCTGAAACGTGGCATCGGTTCTTCCTTCATCCGCCCA
GTCTTCTCTATCATGGATCCAGTGCTGACGTTCACCCTGCCGACCTGTCAGACCGTTTATGGCATCGCAGATATGA
TGGCCCACGTTATGGA

adh1 G2
ACCTGTCAGACCGTTTATGGCATCGCAGATATGATGGCCCACGTTATGGAACGCTACTTCACCCAGACCCAGGGT
GTGGATATTACTGACCGCATGTGCGAGTCTATCCTGCTGTCTATTATCCACAGCGCGAAAACTCTGATTCGCGAAC
CGGAAAACTACGACGCTCGTGCCAACATCATGTGGGCCTCCACGATCGCGCACAACGGTATCTGCGGCGTGGGT
CGTGAAGAAGACTGGGCGACCCATGCTCTGGAACATGAACTGTCCGCGCTGTATAACATCGCACACGGCGCCGG
CCTGGCTGTGATGTTTCCGGCGTGGATGCAATACGTATACACCGCGGGTATCGACCGTTTCGTGCAATTTGCTAC
CCGCGTTTGGAACATCGAAAACATCGGCTCTAAAAAAGAGATTGCCCTGAAAGGTATCCACGCTCTGAAAGACTTT
TTCTCCTCCATCAAACTGCCAATCAACTTTGAACAGCTGGGCGCACAGAAAAGCGATATTGACAAACTGATTGACA
CCCTGAAAATTAACACCAAAGGTAAACTGGGTAACTTCCTGCTGCTGGACATGAACGATGCTCGTGCAATCTACGA
AATTGCTGCTAAGCGTTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCCGCCTAATGAGC

adh2 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGAAGATTAAACTCTAAGCGAG
GAATACATGGTCAACTTTTCCTACTGCAATCCAACCCGTATCGAATTCGGCAAAGGTAAAGAAAACTCCATCGGTG
AATACCTGAACGAATATGGCGCAAAAAACGTGCTGATTCTGTTCGGCTCCGACCGCGTTAAAAAAGACGGTCTGTT
TGACAAAGCGACTGCGTCCCTGACCAAATTCGGCATCAAATTCTCCGAACTGGGTGACATTGTGAGCAATCCAGT
ACTGTCCAAAGTTTATGAAGCTATCAACCTGGCCCGCAAAAACGGCGTGGATAGCGTTCTGGCGATCGGCGGTG
GTTCTGTCCTGGATACTGCCAAATCCGTAGCAGCCGGTGCAAAATACGACGGTGACGTTTGGGATCTGTTCCTGG
CCAAAGCTCCGATTAAAGATGCTCTGATGGTTTTCGATATTATGACCCTGGCTGCAACTGGTAGCGAAATGAACAG
CTTCGCCGTTGTCACCAACGAAGACACTAAAGAGAAAATCTCTATCACCTCTTCCCTGGTGAACCCAAAAGTAAGC
GTAATCAATCCGGAACTGATGAAATCCATTTCTAAAAACTACCTGGTGTACTCCGCGGCCGACATCATCGCGCATT
CTATCGAAGGCTACCTGACCGCAACTCATCACCCGGAAATTATCTCCAAACTGGTTGAAGCGAATATCTCC

adh2 G2

CAACTCATCACCCGGAAATTATCTCCAAACTGGTTGAAGCGAATATCTCCACTATTATTAAAACGACCGAAATCCTG
CTGGCTGACCCAGACAACTACGACGCACGTGCGGAATTTGCGTGGGCAGCAACTTGTGCTCTGAACGGCACCAC
TTACGTTGGCGTTGGTGGTTACTCCTACCCGAACCACATGATCGAACATTCCATCTCTGCACTGTACGGTGTACCG
CATGGTGCGGGTCTGTCCGTAGTAATGCCGGCATGGATGAAATGGTATAAGGACAAAAATGAAGCCCAGTTCTCT
CGCTTCGCTAAAGTAATCTTCGGTAAAAACAGCGCTGATGAAGGTATTGAAGCCCTGAAGACGTGGTTCAAAAAAA
TCGGCACCCCGACCAAACTGCGCGACTTCGGCCTGGACATGTCCGTATCTGACATCACCACTGCTGCGCTGCAT
CACGCTAAAGCATTTGGTATCGCTGATATCTATACCAAAGACGTTCTGGAAGAAATTCTGAACCTGGCTTACTAAA
CTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCCGCCTAATGAGC

adh3 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGAGATACCTCTCCCTTAAGAG
CGAGGTCATTATGATTAACTTCGACTATTGCGTGCCGACTAAAGTTGTTTTCGGTCATGGTGTTGAATCTAACGTTG
GCAAATACGTAAAAGAGTTCGGTGGTACCAAAGCGATGATTCACTGGGGCGGTGACTATGTTCGCGATACGGGTC
TGCTGGACCGTGTCGAAAAATCTCTGTCCGCGGAAGGTATCGGCTACGTTGAGTTTGAAGGCGTCGTACCGAACC
CGCGCCTGTCCACCGCTAAAGAGGGCCTGGCTCTGGCGAAACGTGAAGGTGTAGATTTCCTGCTGGCTATCGGC
GGCGGTTCTGCAATCGATAGCAGCAAAACCATCGCATACGGTCTGGCGAACGATTTCGAGCTGGAAGACCTGTTC
CTGGGTAAAGTAAGCACTGACCGTATCGCGGGCCTGGGTGCGATCTCTACCCTGGCCGGCACCGGTTCTGAAAC
CTCTAACTCTACTGTTATCAACATCGATACGATGGGTGACGTCGAGCTGAAACGTAGCTACAACCACGAATGTGCC
CGTCCGAAATTCGCGATCATGGATCCGGAACTGACCTATACCGTTCCGGCATGGCAGACGGCCGCCGCTGGCTG
CGACATTATGATGCACACTA

adh3 G2

TTCCGGCATGGCAGACGGCCGCCGCTGGCTGCGACATTATGATGCACACTATGGAACGTTTCTTCACTACCGTTT
CTCATACGGAACTGATCGATCAAATGTCCCTGGGTCTGCTGCGTGCTGTCAAAACCGCGATTCCACTGGCTCTGG
CTGAGCCGGATGACTATGATGCACGCGCCACCCTGCTGTGGGCGGGCTCTCTGTCTCACAACGGTCTGACCGGC
ACCGGTCAGCAGGGTGACTTCGCATCCCATGCAATTGAACACGAAATGGGTGCTCTGTACAACTGCACCCACGGC
GCAGGTCTGTGCGCGATGTGGTCTTCCTGGGCTCGTTATGTCATTGATGTGCGTCCGGAACGTTTCGCACAGTTC
GGTGTGGAAGTCTTCGGTGTGGTAAACGACTACTCTGATCCGAAAGGTACCGGTCTGCGCGGTATCGAGGCTTG
GGAAAAATTCTGCAAATCTGTGGGTATGCCGGTACGTATGAGCGAACTGGCAATCAACCCGACTGATGAGGAGAT
CCGTCATATGGCTCAGGGCGCCATTGACGCCCGTGGTGGTGATCATTGCGGTTCTTTCATGGAACTGCGTGTTGA
TGACGTCGTAAAAATTCTGGAAATGGCCCGCTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAG
CCCGCCTAATGAGC

adh4 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGACCCACCTTCCAAAACTCCC
AGAGGTATTCATGCAGAAATTTGACTACTATACTCCGACCAAAGTTATCTTTGGCAAAGGCACCGAAAACAAAGTG
GGTAAAGAGATGAAAAAAGACGGTGCTAAGAAGGCTTATATCGTTTACGGCGGCAAATCCGCGAAAAAAAGCGGT
CTGCTGGACAAAGTGGAGAAATCTCTGAAAGACGAAAACATTGAATACAAAATGATCGGTGGCGTGAAACCGAAC
CCTCGCCTGTCTCTGGCTCGCGAAGGTGTGAAGGAAGCGAAGGAATTCGGTGCCGATTTTATTCTGGCGGTTGGT
GGTGGCTCTGTTATCGATACCGCAAAAGGCATCGCACATGGCGTAGCAAACCCTGACACTGACATCTGGGATTTC
TGGGAAGGTAAAGCCAAGGTTGAAAAATCCCTGCCTGTTGGCGTTATCCTGACCATTTCTGCTGCGGGTTCTGAA
ATGAGCAACTCCGCGGTGCTGACGAATGAAGAAACTGGCATGAAGCGTGGCCTGTCCACCGATTTCAACCGTCC
GAAATTCGCCATCATGGACCCGGAACTGACCTACACGCTGCCGGATTACCAGGTTGGTTGCGGTGTGGTAGACAT
CATGATGCACACCATGGATC

adh4 G2

ATTACCAGGTTGGTTGCGGTGTGGTAGACATCATGATGCACACCATGGATCGTTATTTCACTGACCTGACTGATTG
CCAGAACGATCTGACCGATGAAATCGCAGAGTCTCTGCTGCGTATCGTTATCAAAAACGGTCGTGTAGCTTGCAA
GAATAAAGAAGACTACCACGCTATGAGCGAAATCATGTGGGCAGGTTCCCTGTCCCATAACGGCCTGACCGGTCT
GGGCGCCCCGATGGACTTTGCAACGCACCGCCTGGGTCACTCTCTGTCCGCGAAATTTGATGTTGCACACGGTG
CGTCCCTGTCCGCCATGTGGCCGCACTGGGCTAACTACGTAAAACATAAAGACATCGAGCGTTTTGCACGCTATG
CGCGTAACGTTTGGGGCATTACGGAAGGCACCGATGAAGAACTGGCTGATAAAGGTATTGAAGCGACCGTGGAA
TTCTTCAAATCTATCAACATGCCGACCTGCTTTAGCGAACTGGGTATCGGCATCCAGGATGAGGATGGCCTGCGT
GAGCTGACCAACCGTTGCTTCTACGTGAAAGGTACCAAAGTAGGTAAACTGATTCCGCTGACCGAAGAAGATATTT
ACCCGATCTATGTATCTGCGAACAAATAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCCGC
CTAATGAGC

adh5 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGATAACACACCTATCAAGAAA
TAATTCAGAGGTCCCAATGGAAAACTTCACCTACTACAACCCGACCAAACTGATCTTCGGCAAAGGCCAGCTGGA
ACAGCTGCGCAAAGAATTTAAACGTTATGGTAAAAACGTTCTGCTGGTTTATGGTGGCGGCTCCATCAAACGCAAC
GGTCTGTACGACCAGGTCACCGGCATCCTGAAAGAGGAGGGCGCGGTGGTTCACGAACTGAGCGGTGTTGAACC
GAACCCGCGCCTGGCTACCGTGGAAAAGGGCATTGGTCTGTGCCGTGAACACGATATCGATTTTCTGCTGGCCG
TCGGTGGTGGCTCTGTCATTGACTGCACCAAAGCAATCGCGGCGGGTGTAAAATACGATGGTGACGCTTGGGATA
TCTTTTCCAAAAAGGTTACCGCCGAAGACGCTCTGCCGTTTGGCACCGTACTGACCCTGGCCGCTACCGGTTCCG
AGATGAACCCGGATTCCGTTATCACCAACTGGGAAACTAACGAAAAATTCGTCTGGGGTTCCAACGTTACCCACC
CGCGCTTCTCTATCCTGGACCCGGAAAACACCTTTACCGTACCGGAAAACCAGACAGTGTATGGCATGGTTGACA
T

adh5 G2

AAACACCTTTACCGTACCGGAAAACCAGACAGTGTATGGCATGGTTGACATGATGTCTCACGTTTTCGAACAGTAT
TTCCATAACGTAGAAAACACTCCGCTGCAGGATCGTATGTGCTTTGCTGTGCTGCAGACCGTCATCGAAACGGCT
CCGAAGCTGCTGGAAGACCTGGAAAATTACGAACTGCGTGAAACCATTCTGTACGCGGGTACCATTGCGCTGAAC
GGTACTCTGCAGATGGGTTACTTCGGTGATTGGGCGTCTCACACTATGGAACACGCAGTGAGCGCAGTGTACGAC
ATTCCGCACGCGGGCGGTCTGGCGATTCTGTTTCCGAATTGGATGCGTTACACGCTGGATACTAACGTGGGTCGT
TTCAAAAACCTGATGCTGAACATGTTCGATATCGATACGGAAGGCAAAACTGACAAGGAGATCGCCCTGGAAGGT
ATTGACAAACTGTCCGCATTTTGGACGAGCCTGGGCGCGCCGTCCCGTCTGGCCGATTACAACATCGGCGAAGA
AAAACTGGAGCTGATCGCAGACATTGCTGCGAAAGAGATGGAGCACGGCGGCTTCGGCAACTTTCAGAAGCTGA
ATAAAGACGACGTACTGGCGATCCTGCGTGCATCTCTGTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAG
ATCTGAGCCCGCCTAATGAGC

adh6 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGACCTCTCCCGGTACGATAAT
AAGGAGGCATCAATGAACAACTTCCTGTTCGAAAACAAAACCAAAGTATACTTCGGTAAGGGTGGTGTTAAAGAAT
ATCTGGGTTGTCTGCTGGAACATTATGGTGACACCGTTATGCTGGCCTATGGCGGCGGCTCCATCAAACATAACG
GTGTATATGATGAAATTGTGGGCATCCTGAACGCCGAAGGCAAACGCATCGTTGAATTCCCGGGTATCATGCCGA
ACCCGACGTATGCTAAGGTGCAAGAAGGTGCTAAACTGGCGCGTGAAAACCACGTAGACCTGATCCTGGCCGTT
GGCGGTGGTAGCGTTTCCGACTGCTGCAAAGTTGTGAGCGCGCAGGCAAAAGTAGATGAAGATCTGTGGGAGCT
GGAAAACACTAAACACACTCGCCCGACTGCATTCATTCCGCTGGGTACCATTGTGACCGTTTTTGGTACTGGCAG
CGAAATGAACAACGGCGCTGTAATCACCCACGAGGAGAAAAAAATTAAAGGTGCTCTGTGGGGCGCACAGGCGG
ACTTTGCATTCCTGGACCCGACTTATACTCTGTCCGTGCCGATGAAACAGGTTATTAGCGGTGCGTTCGACACTCT
G

adh6 G2

ACTCTGTCCGTGCCGATGAAACAGGTTATTAGCGGTGCGTTCGACACTCTGAGCCACGCTATGGAAACTTATTTC
GGCAAACCGGATGAGAACAATCTGTCCGACGACATCAACGAAGCGGTGATGCGTTCCGTTATCCGTAACATTCGT
GTGCTGCTGACCGACAAGGATAACTACGAAGCACGCTCCGAACTGACCTGGGCTTCTGCGATGGCAGAAAACGG
TATTCTGAAAATCGGTAAAGTAACTGACTTTCAATGCCACATGATCGAACATCAGCTGGGCGCATACACTAACTGT
AACCACGGCGCTGGTCTGGCGGTTATCCACCCGGTTCTGTATCGTCATCTGCTGCCGGCGAACACCGCACGTTT
CGCGCGTTTCGCTCAAAACGTTTGGGGCATCGATCCAGCAGGTAAATCCGAACTGAAACTGGCGCAGGCGGGTG
TGGAAGCTCTGGCGGCGTTTATCAAGGAAATTGGCATGCCGACTACCTTCGCTGAGCTGGGCGTTCCGGCGGAC
ACCGATCTGAAAGCCGTAGCTGACTCTACCGTCCTGACCGGTGGTTGTTGCAAAAAACTGTCTCGTGAAGAGCTG
CTGGACATCCTGAACGAATGTAAATAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCCGCCT
AATGAGC

adh7 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGATATCCAGCCATTCCCCAGG
AGAAACCACTATGCGTAACTTTACCTACCACAACCCGGTCCGTATCCTGTTCGGCGATCATGCTCTGGACCAGCT
GCCGGATCTGTTCCGTGAATTCCACGTGTCTAACCTGCTGCTGGTGTATTCTGGCGATTTTATTAAAGAACTGGGC
ATCTGGGATGCCGTTTACAACGCTTGCGCGGAAAATGGTATCGCATTTTACGAAGAAGGTGGTGTAGTCCCGAAC
CCGAAAATTGAACTGGTTCGTGAACTGGTCGCACTGGGCAAAAAAAAAAAAGATCGACTTCATTCTGGCTGTAGGC
GGTGGTTCTTCCATCGACACTGCTAAGGCTGTTGCCGCAGGCATCCCGTACGCCCACGACGTGTGGGACTTCTTC
GAATACACTGCGGTTCCGGAAACGGCGGTGCCGATCGGTGTAATCACCACGATCCCAGCGTCTGGTTCCGAATG
TTCTAATTGCAGCATTATCTCCAACGGTCTGCACAAATGCGGTATTGAGTACGATTGCATCATCCCACAGTTTGCC
ATCATGAACCCGGAGTACACCCGTACCCTGCCTGCGTACCAGACCTCCGCAGGCATCGCGGACATTCTGTCCCA

adh7 G2

GTACCCTGCCTGCGTACCAGACCTCCGCAGGCATCGCGGACATTCTGTCCCACATGCTGGAACGCTACTTCACGA
ACACTACTCACGTTGACACCACCGACTACATGCTGGAAGGTACCATGCAGGCTCTGATGGTCAACGCGCGCCGC
CTGATGAAACAGCCGGATGACATCCACGCGCGCGCAGAAGTTCAGTGTCTGGCTTTCCTGGCACATAACAACCTG
CTGGACATCGGTCGCGAATCTGACTGGGGCCCGCATCGTATTGAACACGAACTGTCCGCACAGTACGGCATTAC
CCACGGTGAAGGTATGGCAGTTGTAACCATCGCGTGGGCACGCTACATGGCTGCACACCACCCGGACAAACTGG
CACAGCTGGCCTCCCGTATCTTCGGTGCTGATCCGTTTGTACATTCCAAAGAGGATATGGCACTGCTGCTGGCTG
ACCACCTGGAAGAGTTTTTCAAATCCCTGCACCTGAAAACCACCCTGCACGAAATGGGTATCGACGATACCCACTT
TGAAGAGATGGCAAACCGTGCCACCAATAACGGTAAGGATTGTGTTGGCCACTACGTGGCTCTGAACAAACAGAT
CTTTATCGACATTCTGCACATGGCCCTGTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCC
GCCTAATGAGC

adh8 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGAGACTTAGTAGTCACACGCA
AGAGGAGGATTCCAGTATGTACGACTTCATGTTCCACGTACCGACCAAGATCTACTTCGGCCGCGGCCAGATCTC
TCACCTGGCAGAACTGTCTGATTTTGGCCAGAAAGCGCTGCTGGTTTACGGTGGCGGCAGCATCAAACGTAACGG
CATTTACGACGAAGCGATTCGTATTCTGACCCATGCGGGTATCGAAGTTGTAGAACTGAGCGGCGTTGAACCGAA
CCCGCGTATTGAAACCGTGCGTCGCGGTGTCGGTCTGTGCGCTCGCGAAGGTGTTGACATGGTTCTGGCTATCG
GCGGCGGTAGCACCATCGATTGCGCTAAAGTAGTTGCGGCCGGCGCGCGTTACGATGGCGACCCGTGGGACCT
GGTACTGGACGGTTCTAAGGCGGCTTCCGCGCTGCCAATCTTTTCTGTGCTGACCCTGTCCGCGACCGGTTCTGA
GATGGATGCATTCGCTGTCATCAGCGATATGAGCAAAAATGAAAAGTGGGGTACCGGCGCAGAGTGTATGAAACC
GACCATGTCTGTGCTGGACCCGTCTTACACCTTCAGCGTGAGCCCTAAACAGACCGCGGCTGGCACCGCCGATA
TGAT

adh8 G2

ACACCTTCAGCGTGAGCCCTAAACAGACCGCGGCTGGCACCGCCGATATGATGAGCCATACCTTCGAATCTTATT
TTTCCATGGACGAAGGTGCGTACGTCCAGAAGCGTCTGGCAGAAGGTCTGCTGGGCACTATGATCCACTTCGGC
CCGATTGCCCTGGCACATCCGGACGACTACGATGCGCGTGCGAACCTGATGTGGGCGGCTTCTCACGCAATTAA
CGGCCTGGTTTCTGATGGTTGTAGCCCTGCCTGGTGCGTTCACCCGATGGAACACGAGCTGTCTGCATTCTACGA
TATCACTCACGGCGAGGGTCTGGCGATCCTGACGCCGGCATGGATGGAGCACGTTCTGGATGCTCAGACTGCTC
CTCTGTTTGCTGCATACGGTTGCAACGTATGGGGTCTGTCCGGCGTAGATGACATGAAAGTTGCTCGTGAAGCAA
TCAGCCGCACTCGTGCGTTTTTTGTTGAAGCTATGCATCTGCCGGCAACCCTGCGCGAGGTCGGCATTACCGATG
AAAAAAACTTCGAAGTTATGGCTCGCAAAGCCGCCGATGGTTGCAAAGGCAGCTTCGTTGCGCTGTCTCAGGACG
ACATCGTAGAAATCTACCGTGCTGCTCTGTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCC
CGCCTAATGAGC

adh9 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGACTCCTTCAATAAGCCCAGG
GAGGATTAAAGCATGAATGATTTCCAGTTTCAGAACACTACCAAAGTTTATTTCGGTAAACATCAGCTGCAACACCT
GCACCAGGAAGTGCTGAAATACGGTCAGAAAGTGCTGATCGCTGATGGCGGTGAATTCATCCGTCAGTCTCCGCT
GTATGCTCAAGTTCTGAAAGAACTGACGGACAACGGCATCCAGATCTTCGAACTGGGTTCTGTGGAGCCGAATCC
GCGCCACACCACCGTTAACCGCGAAGTAAAACTGTGTAAAGGCAACAACATCCAGACCGTACTGGCCGTTGGCG
GCGGCTCCACGATTGACTGCTGTAAAGCGATCGCGGCGACCTCTTGCACCGACGAAGACGACGTTTGGACCCTG
ATCGAAAAACGTGAACCGATCAACCAAGCGCTGGCGGTTATCGCTATGCCGACCATCGCGTCCACGGGCTCTGA
AATGGACAAGAGCTGCGTGATTGCCAACGAAGAGCTGCACCTGAAAAAGGGTCTGAACGGCGAAGCTATCCGTC
CGAAAGCGGCTTTTCTGAACCCGGAAAACACCTTCACCGTTCCGGCACGTCAGACCGCGTGTGGTGGCTTCGAC
ATCATGATGCATCTGCTGGATAT

adh9 G2

CGTCAGACCGCGTGTGGTGGCTTCGACATCATGATGCATCTGCTGGATATGAACTATTTTGTAGACTCTGATAAAT
ATCCGCTGCAGTTCAATGTGGTAGAAACCCTGCTGCGCACTATTCGTGAGCAGCTGCCGATCGCGCTGCGTGAG
CCGGAAAACTACGAGGCTCGTGCGACCCTGCTGTGGGGTGCTTCCTGGGCGCTGAACTCTTTCTGTACCTCCGG
TTTCAAAACCGCACCGAGCAACCACGGTCTGGAACAATTCTCTGCGTTCTACGATCAGACGCATGGTCTGGGTCT
GGCTCTGGTGGTTACCAAATGGATGACCTACCTGCTGGAAAAGGACCCGACCGTGGCACCAGATTTCGCTCGTCT
GGGCACCAATGTGCTGGGCTGTCAGCCAGTTGACGATGTGATCGAGGGCGCAAAAAACGCTATCAAAGCCTTTG
ACGCATTCATTGTGAATGACCTGGGTCTGCCGCGTACCATGACTGAAATCGGTCTGAACGACTCTAAGCTGAGCG
AGATGGCTCATGCTGCGGTAACCGGTTATGGCGACGGCACGCTGAAGGGCTACCGTGAACTGACTGAAGCGAAC
TGCCTGGCCATTTATAAAATGTGCCTGTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCCG
CCTAATGAGC

adh10 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGAGCCTTAATCCCCGTAAGCA
CAGGAGATCCACAATGCAGAATTTCGTTTTTCACAACCCGACCCGTATCGTTTTCGGCCGTGACAAGACGGCGAG
CATCGGCAAGGCGACCCTGCCGTATGGTCGCCGCGTTCTGCTGCTGACGGGTCAGGGTTCCGTCGTGAAACACG
GTATCCTGGCGAAAGTGACCTCTTCCCTGTCTACTGCGGGTATCTCCTGGGTTGAGTGTAGCGGTGTGCAGCCGA
ACCCGGTTCTGGGCTTCGTGCGTCAGGCCATCGACACTTTCCGTCGTGAAAACCTGGACGCCATTGTAGCGGTTG
GCGGTGGCTCCGTGATCGACACCGCGAAGGCGGTGGCTGCGGGCGTTCGTTACGAAGGCGATGTTTGGGACTT
CTTTACCGGTAAAGCTAACGTCCTGGACGCGGCCCCGATCACTGTAGTGCTGACTCTGCCGGCGGCTGCATCCG
AGATGAACAGCGGCGGTGTTATCACTAATGAACAAACTCGTCAAAAATTCAACCTGGGCGGCGAACCGCTGTCTC
CGAAAGTTTCTATCCTGGACCCGGTCAACAGCTTTAGCGCCCCGGTGAATCACTCCCTGTACGGTGTTGTTGACG
CGAT

adh10 G2

ACAGCTTTAGCGCCCCGGTGAATCACTCCCTGTACGGTGTTGTTGACGCGATGGTTCATCTGCTGGAGGGCTACT
TCAACGGCTCTGACCCGTGGACTCCACTGCAGGACCGTTACGCGGAAGGTATCATTCGCACTCTGATGGAATGC
GCTGCCATTATTCGTGAACAGCCAGACCACTACGACGCACGTGCTAACATCATGTGGGGCGCGACTCTGGCTTTC
AACGGCCTGGCACCGTGCGGTATCGGCCCGGCAGGTTTTCCGATGCACATGATCGAACACAGCCTGTCTGCACT
GTATGATGTATCTCATGGTGCGGGTCTGGCGATGATCCTGCCGGGTTGGCTGAAGTACCACTCCGATTCCAGCCC
GCGCAAAGTTAACCAGTTTGGCCGTCGTATTTTTGAACTGGATCACCAGGATGATCGTCAGGGCGCTCAAGCAGC
CATTGCCGAGCTGGAACGTTGGCTGCGTTCCATGGATATCCCGGCATCCCTGCACGAAGGTGGCATCCCGATCG
ATGAGATCCCAGCAATTGCGGAGAACGCTGTGATGCTGGCGCAGAAATGGGGTCTGAAAGCTTACACTCAGGCC
GTTATCGAAGACGTTCTGCGTCGCGCTTCTCGCTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTG
AGCCCGCCTAATGAGC

adh11 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGATGATCCCTCCACAACTAAA
GGCGGTATTCAAATGAAAGACTTCAACTTCTACGCACCGACCCGTGTAGTGTTCGGCAAACAGAGCGAAGAGCAG
CTGCCGCGCCTGCTGAAAGAAGCGGGTGGTAAAAAGGTTCTGGTACACTATGGTGGCGGCTCTGCAAAACGTTC
TGGCCTGCTGGATAAAGTGTATGGTATGCTGGACGACGCGGGCATCGAACATGTAGGTCTGGGCGGTGTAGTAC
CGAACCCGCTGCTGTCCAAAGTAAACGAAGGCATTGACCTGTGCCGTCGTAAAGGTGTAAACTTCATTCTGGCTG
TAGGCGGCGGCTCCGTAATCGATAGCGCGAAAGCAATTGCGTATGGTGTGCCGTACGAGGGTGACGTTTGGGAT
TTCTGGAATGGTAAGCCGGCAACCGCTGCCCTGCCGGTCGGTGCAATGCTGACTATCCCGGCTGCTGGCTCTGA
AATGAGCAATTCTTGCGTGATTACTAAAGACGAAGGTGCTGTTAAACGTGGCTTCAACAACGATCTGTGCCGCTGT
AAATTCGCGATCATGAACCCAGAACGCACTTACACGCTGCCGCCGTACCAGACTGCCGCGGGTGCGACCGACAT
CATG

adh11 G2

CACTTACACGCTGCCGCCGTACCAGACTGCCGCGGGTGCGACCGACATCATGATGCACACCATGGAACGCTACT
TTTCCAAACATGAAGACATGACCCTGACCGACGCAATTGCGGAAGCCCTGCTGCGCACGGTTAAAGAAAGCACCT
TCGAAGTGCTGAAACACCCGGAGGACTACCGTAACCGCGCTCAGATTATGTGGGCCGGCTCCCTGTCTCATAAC
GATCTGACCGAATGTGGTCTGGAAAAGGATTTCGCGACTCACCGCCTGGAACACGAGCTGTCTGCGCTGTTCGG
CGTTACCCATGGCGCCGGCCTGGCAGCCGTGTGGCCTGCATGGGCGCGTTATGTGATGAAGAAACACATTTCCC
GCTTCGTTCAGTTCGCGGTCAACGTGATGGGCGTTCCGAACGACTTTTCTAACCCGGAAGCTACCGCTGAGAAAG
GTATCTGTCGTATGGAACACTTCTTCCACGCGATCGGTATGCCGACCTCCATCAAAGAACTGCTGGGTCATGATAT
CACCGAAGCGCAGATTGACGAAATGGTTGACAAATGCTCTCGTGGTGGTACTATCACTGTTGGTGCCATGGAGGT
GATTGCCCCAGACGACATGCGTGCGATCTACCGTATGGCACGCTAAACTAGTATCGATGATAAGCTGTCAAACAT
GAGCAGATCTGAGCCCGCCTAATGAGC

adh12 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGATGGTTCTACAATAATAGGA
GGACTCTACACATGCTGGGCGACTTTACCTACTCCAACCCGACGAAAATTTATTTCGGCGAGAACTCTCTGGACAA
CCTGTCTACCGAACTGAAAAACTATGGCAAGAACGTGCTGCTGGTATACGGTGGTGGTTCTATCAAAAAAAACGGT
ATCTACGATAAGGTTATCGACATTCTGAAAAAGTGTGATAAGACTATTATTGAGGATGCGGGCGTAATGCCTAATC
CGACTGTTGAAAAGCTGTATGAAGGTTGCAAACTGGCTCGTGAAGGTAACGTTGACCTGATTCTGGCGGTTGGCG
GTGGCAGCGTGTGTGACTACGCGAAAGCAGTTAGCGTCAGCACGTATTGCAACGAGGATCCGTGGGAAAAGTAC
TACCTGCGTATGGAGGACGTTGATAACAAAATTATCCCAGTTGGTTGTATCCTGACCATGGTTGGTACTGGTTCCG
AAATGAATGGCGGCTCTGTTATCACCAATCATGAACAGAAACTGAAAATTGGTCACGTTTTCGGCGACAATGTGTT
CCCGAAGTTCTCCATTCTGAACCCGACCTTCACCTACACGCTGCCGAAATATCAGATGATCGCTGGTTTCT

adh12 G2

AACCCGACCTTCACCTACACGCTGCCGAAATATCAGATGATCGCTGGTTTCTACGACATCATGTCCCATATCCTGG
AACAGTACTTTAGCGGTGAAGACGACAACACCTCTGATTATATCATGGAAGGTCTGCTGAAATCTCTGATCCATTC
TAGCAAAATTGCCGTGAACGATCCTACCAACTACGAGGCTCGTTCTAACATCATGTGGATTGCAACCTGGGCTCTG
AACACCCTGGTGGCTAAAGGCAAAACCACGGATTGGATGGTTCACATGATCGGCCAGAGCATCGGTGCTTACACC
GACGCCACGCATGGTATGACCCTGGCTGCCGTGTCCATTCCGTACTACAAGTACATTTGTCCATACGGCCTGAAC
AAATTCAAACGCTATGCGATTAACGTTTGGGATGTTCTGTCTGAAGGCAAAACTGACGAGCAGATCGCTAACGAAG
GTCTGGAATGTATGGAAAAATACATGCGTGACCTGGGTCTGGTAATGAACATTTCCGATCTGGGCGTCAAAGAAG
AGATGCTGGAGGGTATCGCTGAAGGTACGTTCATCATGAACGGCGGTTATAAAGTACTGACCAAAGACGAAATTA
TCACCATCCTGAAACAATCCATGAAATAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCCGC
CTAATGAGC

adh13 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGAACGTAAGGCCACTACATTA
ACTAAGGAGCAAAATATGGAAAATTTCGATTTCCACGTTACTACTGATATCCGCTTTGGCAAAGACCGTCTGGGTG
AACTGCCGCAGGTTCTGAACAACTTCGGCAAAAACGTGCTGCTGGTTTACGGTGGTGGCTCCATCAAGCGTAATG
GTCTGTACGACAAACTGTACGAACTGTTCAACCAGAACGACAATAACGTTGTTGAACTGGCGGGTGTAGACCCGA
ACCCGCGCATTGAAACCGTGCAAAAAGGTGTCCAGCTGTGTAAGGAACACGCGATCGACGTCGTGCTGCCGGTA
GGTGGCGGCTCTGTGATTGACTGCTCCAAAGCTGTGGCGGCTTGCGTCTTTGTTAGCGGTGACCTGTGGGAAAA
CTTCGTGCTGCAGAAAAACTATAAAGGCCCGGCACTGCCGATTGTCACCATTCTGACGCTGGCCGCTACGGGCTC
TGAGATGAACGGTACGTGCGTAATCTCTAACATGGATGCGCAGATTAAACTGGGCGTCCACGGTACCACCAACCT
GCTGCCAAAGGTATCCTTCCTGGATCCGACTAACACCTTCTCTGTTGGTGCATACCAGACTGCAGCTGGCTCCGC
TGACATCCTGAGCCACCTGAT

adh13 G2

TGGTGCATACCAGACTGCAGCTGGCTCCGCTGACATCCTGAGCCACCTGATGGAGAACTATTTCAACGCGACCGA
AGGCACCGAAGTTCAGGATGAAATCGCTGAAGGCCTGATGAAAACGGTGATCAAATATCTGCCGGTGGCGCTGG
ACGAACCGGACAACTATATTGCCCGTGCTAACCTGATGTGGGCCTCTACTCTGGCGCTGAACGGCCTGGTTGGCA
AAGGTAAAAAAGGCAGCTGGTCTTGTCATGCTATGGAACACGAACTGTCCGCTTTCTATGACATCACTCACGGCGT
CGGCCTGGCTATGCTGACCCCGCGTTGGATGGCACACATCCTGGACGAAGACACCCTGCCGAAATTTCAACGTTT
TGCTGAAGAGGTCTGGAATGTTAAAGAAAAGGAACCGAAACGTACGGCGGAGATCGGCATTCAGAAACTGTACGA
TTTTTTCGTCTCCTGCAACATCCCTATGACCCTGTCCGGTGTGGGCATCCAGACCGAAGAAAATTTTGAAGAAATG
GGTCAGCGTGCCGTTGCTCACTCCTCCATCTCTAATCAGGGCTTCGTACCGCTGCACGAGGACGACGTGGTCTC
CATCTATCGCGACTGCATGTCCGAGTCTTCTTTCGTCTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGA
TCTGAGCCCGCCTAATGAGC

adh14 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGAAGTATATTTCCCGCTCAAT
ATAAGGAGGAGTACATATGGAATCTTTCGATTTTTTCCGTCGCACTCGTATCATCTTTGGCCAGTCTGCGGACAAC
GAAGTAGGTCAGATTATCAAATATCAAGGTGGCACTCGTGTGCTGCTGCTGCACGGTGAAAAAGCAGCGATCAAG
TACGGTATTGTGGAGCGTATTGGTCGTACTCTGGACCGTTCCGGTCTGAAATACTTCTCCAAAGGCGGCATCAAG
AGCAACCCGCATATTGATAAAGTTTACGAATGCATTGAATTCTGCCTGTCCAACTCCATTAATTATATCCTGGCTGT
GGGTGGTGGTTCCGTGATCGACACCGCCAAAATCGTCGCGGCGGGCGTATTCTTCGACGGCGACATCTGGGACA
TGTTTGAAAAACATCGCGAACCGTACCGTTCCCTGCCGCTGGGCTGCGTAGTTACCGTTCCTGCAAGCGGTACTG
AATGCAGCAACTCTTCTTCCCTGATGCGTGAAAAAGACGGCCGCCGTGAAAAACTGATCGCGTATTCTAACAGCTT
CGTACCGGAGTTCGCCATTCTGAACCCGGACCTGACGCTGTCTCTGTCTCCGCGTGTGACCGCTAGCGGTTGCG
TTGATATGATTAACCATG

adh14 G2

CTCTGTCTCCGCGTGTGACCGCTAGCGGTTGCGTTGATATGATTAACCATGTCCTGGAAGGTTATTTCTCCAACTC
TACCGGTGTACTGCTGAGCGATAAGCTGTGTGAAGCGGTTCTGAGCTCTATTATCGAACTGCTGCCGCAGATCTA
TGAAGATCCGAATAACATTGATGCGCGCGCAAACCTGATGCTGGCAGCAACCCTGTCTCACAATGATATCTGCTG
CATGGGCCGCAAGTCCGACAACGTTATCACGAAACTGGCCAACCAGCTGGTGGTTGAAAACGATTGTCCGTTCGG
TGATGCACTGGCTGTTCTGATCCCGGCTTGGATGGAATATGTTGTTCAGTTTAACCCGCTGCGCATCGCACAATTC
TCCAACCGCGTTTTTGGTATCGCAATCAACTTCGAAGATCCGAAAATTACCGCGTATGACGGTATCAAAGCCCTGC
GCGCTTTTTTCAAAAATGTAAAACTGCCGTGCAACTTCGTTGAACTGGGTATCAAGACCGAAGCAATCGCGGACAT
CGTAAACGCTCTGGACCTGAAAGAAGGTAAAACTCTGGGTTCTTTTGTGCCGCTGGACGCTGTGGCCTGCGAAGC
AATCCTGTCCCTGGCCGCCAATTACTGCGAAGGTCGCGATATTTTCTAAACTAGTATCGATGATAAGCTGTCAAAC
ATGAGCAGATCTGAGCCCGCCTAATGAGC

dhaT2 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGATCCTTACCGCTTACCACGA
AATAAAGAGGGAAAAATGTCCGACTACTATGATTATATGCTGCCGACCGTTAACTTCATGGGCCCGGGCTGTGTG
GAGGTTGTGGGCGAACGTTGCAAGATCCTGGGTGCTAAAAAAGTTCTGATTGTTACCGATTCTTTCCTGCGTAACA
TGGAAGGCGGCCCGGTCGACCAGGTGGTCAAATACCTGAAAAAGGCGAATCTGAATTACGCTTTTTACGACGAGG
TTGAACCGAACCCGAAAGACGTTAACGTTTACGCCGGTCTGAAAATTTACGAACGTGAAAACTGCGATATGATCGT
TACCATTGGTGGTGGTTCCGCTCACGATTGTGGTAAAGCTATTGGTGTTGCTGCTACTCATGACGGTGATCTGTAC
AAAGACTACGCTGGCATCGAAAAGCTGGAAAACGAAACTCCGCCGATGGTATGTGTGAACACCACTGCTGGCACC
GCTTCTGAAGTTACTCGTCACACCGTGATTACTGACACCAGCCAGACGCCTAACGTCAAGTTCGTGATTGTATCCT
GGCGTAACACTCCAGATGTGTCTATCAACGATCCGGAACTGATGGTTGGTAAGCCGCCGGGTCTGACCGCAGCT
ACTGGCATGGACGCTCTGACTCACGCGGTTGAAACCTATGTGAGCACCAATGCGAAC

dhaT2 G2

CATGGACGCTCTGACTCACGCGGTTGAAACCTATGTGAGCACCAATGCGAACGCGCTGACTGATGCCGCGGCTA
TTAAGTCTATTGAACTGGTTGCTAATAACCTGCGTAAAGTTGTTAAAGATGGCCAGGACATCAAAGCGCGCGAAAA
TATGGCGAACGCTTCCGTTCTGTCTGGTTTCGCGTTTAATAACGGTGGTCTGGGTTACGTTCACGCTATGGCACAC
CAGCTGGGCGGCTTCTACGACATGCCTCACGGTATCGCAAATGCAATTCTGCTGCCGTACGTCGAGAAATTCAAC
CTGGGTACCGACGTTGAACGTTTCAGCAACATCACTGAAATCTTCGGCAAAGAACAGTCTAAAATCAGCAATAATC
CGGAAGCACAGGAAAGCATCAAAGCCATCAAAGACGAAATCGATAAACTGAAACGCTTTAAAAAAATTGCAGAAGT
GTTTGGCGTGGATACCTCTAACATGTCTACCCGCGAAGCTGCAGAGGCTAGCCTGGATGCCATCAAAGAACTGGC
CCGTGATATCGGTATTCCGAGCTCTCTGTCTGAGTCTAAATTCGACGTTAAACGTGACGATTTCGAGGAAATGGCT
AAACTGGCGCTGGAAGATGGTAACGCGGGCACCAACCCGCGTAAAGGTTCTGTAGAAGATATTGTTCGTATCTTC
GAGGATGCGTTCTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCCGCCTAATGAGC

dhaT3 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGATAATAGCAATACTAATAAA
GAGGTCTCCTAATGAGCTATCGCATGTTCGACTACCTGGTCCCGAACGTGAACTTCTTCGGTCCAAACGCAATCA
GCGTGGTGGGTGAACGCTGCCAGCTGCTGGGTGGTAACAAAGCTCTGCTGGTGACCGACAAGGGTCTGCGCGC
TATCAAAGACGGTGCGGTAGACAAGACCCTGCATTACCTGCGTGAGGCGGGTATCGAGGTTGCTATCTTCGATGG
CGTTGAACCTAACCCGAAGGATACTAACGTGCGCGACGGTCTGGCAGTTTTCCGCCGTGAACAATGCGATATTAT
TGTGACGGTTGGTGGCGGTTCCCCGCATGATTGCGGCAAAGGTATCGGCATCGCTGCTACGCACGAAGGTGCGA
TCTGCACTTCTATGCCGGAATCCCGTCCGTGGACCAACCCGCTGCCTCCGATCGTGGCCGTAAACACCACTGCA
GGTACCGCCAGCGAAGTGACCCGTCACTGTGTTCTGACCAACACCGAGAGCAAAGTAAAGTTCGTTATTGTCTCC
TGGCGTAACCTGCCGTCTGTGTCCATCAATGACCCGCTGCTGATGATCGGTAAACCGGCTGCCCTGACGGCAGC
CACCG

dhaT3 G2

TGACCCGCTGCTGATGATCGGTAAACCGGCTGCCCTGACGGCAGCCACCGGCATGGACGCTCTGACCCACGCA
GTGGAGGCATACATCTCCAAGGATGCCTCCCCGGTTACTGATGCAGCGGCTATGCAGGCAATCCGTCTGATCGC
CCGTAACCTGCGCCAGGCCGTTGCTCTGGGCAGCAACCTGCAGGCTCGTGAAAACATGGCGTACGCTAGCCTGC
TGGCTGGTATGGCATTTAATAATGCGAACCTGGGTTACGTGCACGCAATGGCCCACCAGCTGGGTGGTCTGTACG
ACATGCCACATGGCGTGGCGAACGCAGTTCTGCTGCCGCACGTTGCCCGCTACAACCTGATCGCGAATCCGGAA
AAATTCGCGGACATCGCGGAACTGATGGGCGAAAATATCACCGGTCTGTCTACTCTGGACGCGGCAGAGAAAGC
TATCGCAGCGATCACCCGTCTGAGCATGGATATTGGTATCCCGCAACATCTGCGTGACCTGGGTGTCAAAGAAGC
GGACTTCCCGTACATGGCGGAAATGGCACTGAAAGACGGTAATGCGTTCAGCAACCCGCGTAAAGGTAACGAGC
AAGAGATTGCGGCTATCTTTCGTCAAGCTTTTTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGA
GCCCGCCTAATGAGC

dhaT4 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGAATACGGAAACCTCACGTAT
AACCAAGGAGGTTACAATGCGCATGTATGACTACCTGCTGCCTAACGTTAACTTCATGGGCCCGGGCTGCATCAA
AGTTATCGGCGAACGTTGCAAACTGCTGGGCGCTAAAAAAGCGTTCATCGTAACCGGTAAACACATTGGTAGCAT
GGAAAACGGTCCGCTGCAGATCGTTGTAAAATACCTGACCGATGAAGGTATCGACTACGTTCACTTTAGCGGCTC
TGAACCGAATCCGAAAGACATTAACGTGCGTAAAGGTGTTGAACTGTTCAAAAAAGAAAACTGCGATATGATCATT
ACCATCGGCGGTGGTTCTGCACATGACTGCGGTAAAGGTATTGGCATCGGCGCGACGCACGAAGGTGACCTGTA
CGACTACGCAGGCATCGAGACTCTGACCAACCCACTGCCGCCTATTGTAGCGGTGAACACCACCGCGGGCACGG
GCTCCGAAGTTACCCGTCACTGTGTGCTGACGAACACCGAGAAAAAAAATCAAATTCGTAATCGTGTCTTGGCGTAA
TCTGCCGCAGGTGAGCATCAACGATCCGCTGCTGATGGTGGATATGAGCCCGCGTCTGACTGCGGCGACCGGTA
T

dhaT4 G2

TCCGCTGCTGATGGTGGATATGAGCCCGCGTCTGACTGCGGCGACCGGTATGGATGCACTGACTCACGCGATTG
AGGCCTACGTGAGCAAAGACGCAAACGTGGTTACGGACGCGGCGGCAATTCAAGCGATCAAACTGATCAGCAAA
AACCTGCGTAAAGCCGTCGCACTGGGCGAAAACCTGGAGGCCCGTGACAACATGGCAAACGCTTCTCTGCTGGC
AGGCATGGCCTTTAACAATGCGAACCTGGGTTATGTTCATGCGATGGCTCATCAGCTGGGCGGTCAGTACGACCT
GGCCCACGGTGTTGCAAACGCCATGCTGCTGCCGCATGTCGAACGCTACAACATTATCTCTAACCCAGAAAAATT
CCGTGACATCGCTGAGTTCATGGGTGAGAATATCGAGGGTCTGTCTGTCATGGAGGCTGCAGAAAAAGCAATCGA
TGCGATGTTCAAACTGGCCGAAGACATTGGCATTCCGCGTCGCCTGCGTGACGTAGGTGTCAAGGAAGAAGATTT
CGAATATATGGCGGGCAACGCGCTGAAAGATGGTAACGCATTTTCTAACCCTCGTAAGGGCACTGAAGAAGATAT
TGTCAACATTTTTAAAGCTGCATACTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCCGCC
TAATGAGC

dhaT5 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGAAATATTATCTCTAATCATAA
AAGGAGGTTCAAACTATGGCGGATTTCCTGTGCCCTGCTGTTAACTTTATTGGCGCGGGTACCGTAGCAGAAACT
GGCCCGCGTGCTGCGATGTTCGGTACGAAAGCTCTGATCGTGTGCGACGGCTTCCTGGCCAAACTGGAGGGCG
GCCCTGTAAGCAAAGTGAAAGATACGCTGACTTCCTCTGGCGTTGAGTTCGCAGTTTATGACGGCGTGGAACCGA
ATCCAAAAGATACGAACGTAGCAGCCGGCCTGAAAATTTATAAATCTGAAAACTGTGACCTGATCGTAACCGTTGG
TGGCGGCTCCTCTCATGATTGCGGTAAAGGCATCGGCATCGCAGCTACTCACGAGGGCGACCTGTACGAAGACT
ACGCCGGTATCGAAACCCTGACTAACGAACTGCCACCGATCATCGCGGTAAACACCACTGCTGGCACTGCGAGC
GAGGTTACGCGTCACTGCGTGATCACCAACACGGCGAAAAAGGTTAAGTTCGTTATCGTTTCTTGGCGTAACCTG
CCAAAAGTGTCTATCAATGACCCGGAGCTGATGGTTGCAAAACCGGCTGGTCTGACCGCCGCGACCGGTATGGA
CGC

dhaT5 G2

GCTGATGGTTGCAAAACCGGCTGGTCTGACCGCCGCGACCGGTATGGACGCACTGACCCACGCAGTTGAGTGTT
ACGTAACTAAGGACGCGAACCCGGCGACTGATGCTGTGGCTATCCACGCTATCAAACTGATCGGCAAGTACCTGC
GTCGCGCAGTGGCAAACGGTGAAGACCTGGAAGCGCGCGAAGGTATGGCTTACGGCTCTCTGCTGGCTGGTATG
GCTTTCAACAACGCGGGCCTGGGCTATGTTCACGCTATGGCGCACCAGCTGGGCGGTCTGCTGGATATGCCGCA
CGGCATCGCAAACGCGGTGCTGCTGCCGCATATCGAACGTTACAACCTGATGGTGAACCCGGAAAAATTCGCCG
ATATTGCGGAGGCGATGGGTGAAAACATCGATGGTCTGGGTAAAATGGAAGCTGCTGAAAAGGCTATCGATGCTA
TCGTTCGCCTGTCCATTGATGTTGGTATTCCGCAGCATCTGGCTGACCTGGGCGTTAAAGAATCCGATCTGGAGC
CGATGGCGAAACTGGCCATGCAAGATGGCAACGCAGGTACTAACCCGCGTGTTGGCAAAGTTGAAGATATCATCC
AACTGTTCAAAAACGCAATGTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCCGCCTAATG
AGC

dhaT6 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGATTCCATTTTTCTTCCACCAA
AAGGAGTTACTTCCATGTATTACGATTTCCTGAACCCGTCCGTTAACTTTTTTGGCCCGGGTTGTGTCTCCGTTGTA
GGCGAGCGTTGCAAGATCCTGGGCGGCAAAAAAGCTCTGATTGTTTGCGATCCGTTTCTGGCAAAGATGGAAGG
CGGTCCGGTGGAACAGGTACTGGGTTACGTTCAGGAAGCGGGCCTGAAAACCGTGGTCTTTGACGGTGTGGAAC
CAAACCCAAAAGACAAAAATGTACACGCAGGCCTGAAAGTGTTCAAAGAAGAAAAATGCGATATGATCATTACTGT
TGGTGGCGGTAGCGCGCACGATTGTGGCAAAGGTATTGGTATCGCGGCTACCCACCCGGGTGATCTGTACAAAG
ACTACGCGGGTATTGAAAAACTGACCAATCCGCTGCCACCGATTGTTGCAGTGAACACCACTGCTGGTACCGGCT
CCGAAGTTACTCGTCACTGCGTTCTGACTAACACCTCCACTTCCATCAAATTCGTCATTGTCTCTTGGCGCAACCT
GCCACTGGTTTCCATCAACGATCCGATGCTGATGCTGAAAAAGCCGGCGGGTCTGACTGCGGCCACTGGCATG

dhaT6 G2

TCCGATGCTGATGCTGAAAAAGCCGGCGGGTCTGACTGCGGCCACTGGCATGGATGCTCTGACGCACGCAGTCG
AATGCTATGTTACCAAAGCTGCAAACCCGGTGACGGACGCACTGTGTGCGCAATCCATCAAACTGATCGCGAATA
ACCTGCGTCAGGCGGTGGCGAACGGTGAAAACCTGACCGCTCGTGAAAACATGGCATATGCTAGCATCCTGGCG
GGCATGGCCTTCAACAACGCCGGTCTGGGCTATGTTCATGCAATGGCTCACCAACTGGGTGGCTACTATGACATG
GCGCACGGTGTGGCGAACGCTATCCTGCTGCCGCACGTGGCTCGCTTTAACCTGATCAGCAACCCACAAAAATTC
GCGGATATCGCAGTGTTTATGGGTGAAAACATCGAAGGTCTGTCCGTACGTGCGGCGGCAGAAAAAGCGATTGA
CGCCATTGTTCAGCTGTCCAAGGATGTGGGCATCCCATCTGGCCTGGCTGAGATGGGTGTAAAAGAGGAAGATTT
TGACAAAATGGCGAAAACGGCACTGGAGGATGGTAACGCTGGCTGCAACCCGATTGTGGGTACCCACCAGGATA
TTGTGAAGCTGTTCGCGGCTGCTATGTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCCG
CCTAATGAGC

dhaT7 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGAAGCCTAGGCTAAGAAGCAT
AAGGACCCCACTATGGAAGAGAACGAAATGCGTATGTATGACTACCTGGTTCCGAGCGTGAACTTCATGGGTGCT
AACAGCATTTCCGTTGTAGGCGAGCGTTGCAAAATCCTGGGTGGCAAGAAAGCTCTGATCGTTACCGACAAATTC
CTGCGTGGCCTGAAAGGTGGTGCAGTTGAACTGACGGAAAAATACCTGAAGGAGGCGGGCATTGAAGTAGCATA
CTACGACGGCGTAGAACCGAACCCGAAAGACACTAACGTGAAAGATGGTCTGAAGATCTTTCAGGACGAGAACTG
TGATATGATTGTAACCGTTGGTGGTGGTTCCTCCCACGACTGCGGCAAGGGCATTGGTATCGCGGCGACTCACGA
AGGCGATCTGTACGACTATGCAGGCATTGAAACCCTGACCAACCCGCTGCCTCCGATCGTGGCTGTAAATACTAC
CGCCGGTACGGCAAGCGAAGTTACCCGCCACTGCGTCATCACTAACACCAAAACTAAAGTTAAATTCGTGATTGTT
AGCTGGCGTAACCTGCCGCTGGTCAGCATTAATGATCCGATGCTGATGGTTGGTAAGCCGGCGGGCCTGACCGC
A

dhaT7 G2

CAGCATTAATGATCCGATGCTGATGGTTGGTAAGCCGGCGGGCCTGACCGCAGCAACTGGCATGGATGCTCTGA
CCCACGCAGTTGAAGCTTACGTTTCTAAAGATGCCAATCCTGTTACCGACGCGGCAGCAATTCAGGCAATTAAACT
GATTTCCTCCAATCTGCGTCAGGCGGTAGCCCTGGGCGAAAACCTGGTGGCTCGTGAAAACATGGCGTATGGCA
GCCTGCTGGCTGGTATGGCCTTCAACAACGCAAACCTGGGCTATGTGCACGCTATGGCGCACCAGCTGGGCGGC
CTGTACGACATGCCACACGGTGTTGCTAACGCGATGCTGCTGCCTCACGTGTGTAAATATAATCTGATCAGCAAC
CCGCAGAAGTTCGCTGATATCGCCGAATTTATGGGCGAAAACATCGAGGGCCTGTCCGTGATGGACGCGGCACA
GAAAGCGATCGACGCGATGTTCCGTCTGTCTACGGATATCGGCATCCCTGCTAAACTGCGCGATATGGGCGTTAA
GGAGGAAGACTTCGGTTACATGGCCGAAATGGCGCTGAAAGATGGCAACGCGTTTTCCAATCCGCGCAAAGGCA
ACGAACGCGATATTGTTGAAATTTTCAAAGCAGCTTTCTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGA
TCTGAGCCCGCCTAATGAGC

dhaT8 G1

CGAGCGCCCGTAACTTCACGCGCCAACGTCGTTGTGTTCTGGCCGGCTAATCTAGAAGCCTAACTTAAATTTAAAA
GGAGGCCCTTATGCGTTACTACGACTATCTGATGCCGTCTGTGAACTTTTTCGGTCCGGGTTGCCTGGAAGTGAT
CGGCGAACGTGCAAAAATCCTGAATGGCACTAAAGCTCTGATCGTTACGGACAAATTCCTGTCTTCCCTGAAAGG
CGGTGCTGTTGAAAAAACCCTGGAGTATCTGAAATCCGCTGGCGTTGAAGCCGTTGTATTTGATAACGTGGAACC
GAACCCAAAGGATACCAACGTTTACGAGGGTGCAAAAGTATATAAGGAAAACAATTGCGATATGATCATCACCGTG
GGCGGTGGTAGCCCACACGACTGCGGTAAGGGCATCGGTATTGCAGCGACCCATGACGGCGACATCTGTGATTA
TGCCGGTATTGAAACCCTGACCAACGCGCTGCCGCCGATCATCGCTGTTAACACCACCGCTGGCACTGCTTCTGA
AGTTACTCGTCACGCCGTCATTACTAACACTAAGACCAAAGTTAAATTTGTGATCGTGAGCTGGCGTAACCTGCCA
CAAGTTAGCATCAACGACCCGCTGCTGATGATCGGTAAACCGGCTGGTCTGACCGCAGCTACCGGTATGGACG

dhaT8 G2

GCTGCTGATGATCGGTAAACCGGCTGGTCTGACCGCAGCTACCGGTATGGACGCTCTGACCCATGCTGTGGAAG
CTTACATTTCCAAGGACGCGAACCCAATTACTGATGCTGCGGCAATCCAAGCGATCAAACTGATCGCCCAGAACC
TGCGTCTGGCCGTGGCCAACGGCGAGAACCTGAAAGCGCGTGAAAACATGGCGTACGCTTCTGTACTGGCCGGT
ATGGCATTTAACAACGGCAACCTGGGTTACGTTCACGCGATGGCGCACCAGCTGGGCGGCCTGTATGATATGCC
GCACGGCATCGCAAACGCAATGCTGCTGCCGCACGTGTGTAAATATAACATGATCTCTAATCTGGATAAGTTTGCA
GATATCGCAGAATTTATGGGTGAAAACGTTGATGGCCTGTCTAAATCTGAAGCGGCCGAAAAAGCGATCTCTGCG
ATGTTCCGTCTGTCTTCCGATCTGGGCATTCCGACCTCTCTGGAAGAAGCAGGCATCAAAGAATCCGACATTGAA
CTGATGGCTGAGAATGCCCTGAAGGATGGTAACGCGTTCTCTAACCCACGTAAAGGCAACGAGAAAGATGTTGAA
AACATTTTTAAAGCGGCGATGTAAACTAGTATCGATGATAAGCTGTCAAACATGAGCAGATCTGAGCCCGCCTAAT
GAGC

adh15

ATGCAGCCGTTCGTGTTTCATAACCCTACCGAAATCGTTTTCGGCGTTGACACTGCTGATAAAGTAGGCAAATACG
CCGCACGTCAGGGTGGTAAAGCCCTGCTGGTTTATGGCCGCAACTCTATCAAAACCACTGGCCTGTATGATCGCG
TAACTGCTAGCCTGCAGGCCGCCGGCCTGAGCTGGGTTGACCACGGCGGTGTCAAATCCAACCCGGTTCTGAGC
CACGTGCGTGAAGGCGTTGCAGTTGCAAAACGTGAACAAGTAGACGTTGTTGTCGCAGTAGGCGGCGGTAGCGT
TCTGGATGAGTCTAAAGCGATCGCTGCAGGCGCGCTGTGCGATCACGATGTTTGGGAGTTCTTCCTGCAGGCTAA
AGTCGAGAAAGCGCTGCCGCTGGTTACTGTTCTGACTCTGGCTGCAACGGGTTCCGAAATGAACAGCGGCGGCG
TTGTAACCAACGAGAACACCGCTCAGAAATTCAACATCGGCTCCCCGCTGCTGTTCCCGAAAACGTCTATTCTGGA
CCCGGCACTGACCTATACTGTTCCGGCTGATTACACCGCATATAGCGCCGTGGATGCGATCTCTCACATCATCGA
AGGTTATTTCACCTCCAACGATCAGGCAACTCCTCTGCAGGACCGTTTCGTAGAAGGCCTGGTGAAAACCATCAT
GGAGTCCACCGAGCAGATTCTGCAGCAACCGGATCACGCGGACGCGCGTGCTACCATGATGTGGTCTGCCACTT
GGGCTCTGAACGGCCTGTCCACTGCTGGTATTGGTCTGTATCAATTCCCGAACCACATGATCGAACACTCTCTGT
CCGCTATGTACGACATCGCCCATGGTGCCGGTCTGTCCATTGTAATCCCAGGTTGGATGGATTATGCGGCGACCC
AAAACCCGGCTAAATTCGCTCAGTTTGCACGTCGTGTTTTCGACTGTGAACTGAGCGATGACCTGGAATGCGCTC
GCTACGGCATCGAAGCGCTGAAAACGTGGTTTCATTCTATCGGCTCTCCGGTGTCCCTGGCTCAGGGTAACATTC
CTGATGAAGAGATCGGCGCTATCGCAGATAACGCTGTAATGCTGGCCCGCAAATGGGGTCTGAAAGCGTACACC
GCCGAAGTTATTGCCGACATCCTGAGCCGCTGTCGTAGCTAA

adh16

ATGGAAAACTTTGAGTTCTACAACCCGACGCGCATTATCTTTGGCAAAGAAGCCGAAAAGAAGATCGGCAAAATCC
TGGAGAAAGACGACGTTGAACGTGTGCTGTTCGTGTATGGTAAAAGCTCTATCAAAGAGACAGGCCTGTATGATC
GCATTGTTAAGGCTCTGAAAAAAGAAGGCATCGAATTTATTGAGCACGGTGGTGTGAAGCCGAATCCGGTGCTGT
CCCACACCCGTGAAGGTATTGAAAAAGCGAAGAAGCACAAGGTAGACGCGATTCTGGCTGTAGGCGGCGGTAGC
GTTATCGACGAAGGCAAAACTATTGCAGTCGGTACCAAAACCGACAAAGACGTTTGGGATTTCTTCAAACGTAACA
AAGAAATTAAAAAAGCACTGCCGATTTATGTGATCCTGACTCTGGCGGCAACCGGTTCTGAAATGAACGGTTTCGC
TGTAATCACCAACGAAGAAACCCAGGAAAAGCTGAGCATCTCTTCTGAACACATCTTCCCGCGTGTTTCTATCCTG
AACCCGGAACTGACCTTCACCGTATCCGCTAAATATCAAGCGTACGCGGCAGTAGACGCAATCGCGCACGTTATC
GAACACTATTTCTCCGGTTCTTACTGCCCGAACCTGCAGAACCGTTTCGTTGAGGGTCTGATCAAAACCGTAATGG
AAACCACCGAAATCATTCTGAAAGAACAGAAGAACTACAACGCTCGTGCAGAATTTATGTGGGCGGCTACCCTGG
CCCTGAACGGCCTGGCAAAGCTGGGCATCAAAGGCGGCTCCTTCCCGAACCACATGATTGCACATTCCCTGGGT
GCTATCTACGATCTGCCACATGGTGCTTGTCTGAGCATCGTAATTCCGGCGTGGATGAAATGGTACCAAGAAAAG
AACCTGATTCAGTTCGAACGCTTCGCCAAAGAAATCTTCGGCGTCAACACTGCTACCGAGGGTGTGTTTCAACTGA
AAGAATGGTTTCGCAAAATCGGTGCGCCGGTTTCTCTGAAAGAGGCAGGCATCTCCATTGGCGAGATCGATCGTA
TTGTAGACAATGCGTATAACATCGCGAAAGTCTGGCAGATGGAAAAAGACTACACTAAGGAGGTACTGACCGAAA
TTATCAAAAACGCTAACGATTAA

adh17

ATGGACAATTTTAGCTTTTATAACCCGACCCGTATCGAATTCGGTGCTGGCAAAGAACAGCTGATCGGTGAAATCA
TGGCTTCTCACGGCATTAAGAAAGTCCTGCTGTCCTATGGTAGCGAACGCATTAAAGATAACGGCCTGTTCGTGA
CCGTTTCCGACTCTCTGAACCGTCACGGCATCACTCTGATCGAATGCGGTGGTATCATCTCTAATCCGCTGATTTC
CAAAGTACGTGATGCGATCACCGTAGCTAAAGAGCAGAATGTTGACGCCATCCTGTCTGTTGGTGGTGGTTCTGT
GCTGGATTCCGCAAAAGCAATCGCAGCCGGTTCCCTGTATCAGGGTGACGTGTGGGACCTGTTCATCGGTAAGG
GCCAGATTGACGCAGCACTGCCAGTTTTCGCCATCCTGACTCTGGCGGCGACTGGTTCTGAAATGAACTCTGGTG
CGGTGGTGACTAACGACGACACCAAAGAAAAATTCGCAATCAACAGCGTACACATCTTCCCGAAAGTGTCCATTGT
AAACCCAGCTCTGATGCAAACCGTTTCCCGTGACTACCTGGTGTATTCTGCTGCAGACATCATTGCACATTCTATT
GAAGGTTACTTCACTGCGACCATTCAGCCGAAAATTCAGTCCCGTCTGGTTGAATCTGTCATTGCTACCGTTCTGG
AAACTACCGAACAACTGCTGGCTGACAGCGCAGACTATAACGCTCGCGCAGAATTCGCTTGGGCTGCGACCCTG
GCGCTGAACGGTCTGACCTATGCTGGCACCTCCGGTTTCGGTTACCCTAACCACATGATTGAACACGCGCTGTCT
GCACTGTTCAACGTACCGCATGGTGCTGGCCTGAGCGTTATCATGCCTGCATGGATGAAGTGGTTCCACTCCCAG
AACACTGCTCAGTTCGAACGCTTCGCCCAGCACCTGTTCGGCCTGAACACCGCGGAAGAGGGCATCGCTGCTCT
GGAGAAATGGTTCGATAAAATTGGTACCCCGACTCGTCTGAGCCAGCTGGGCATCAAAGCCACTGACCTGCCGG
CTATCCTGGACAACCTGGAAGGTAACGCGCGTTGGTTCGGTCTGGCAGAAACTTATACCGAGGAAGTACTGGCGA
CCATCCTGCGTCTGGCACTGTAA

adh18

ATGTATCCATTCTCCTTTCAGAATCCAACCCGTATCGAGTTCGGTCTGGATAAAGAAAAGGAAATGGGTAAATACA
TGCACGAATACGGCGCGAAGAAAGCGCTGATCATCTATGGCAGCGAGCGTGTGAAACAATCCGGCCTGTTCGAA
GACGTTACCAAGTCTCTGCGTGAACACGGCATCGAGTACATCGAATGTGGCGGCGTTAAAAGCAACCCGACCATC
AGCAAAGTCCGTGAAGCTGTTGCAATGGCCAAAGCATTCGGCGCAGATAGCGTGCTGTCTATTGGTGGTGGTAGC
TGCCTGGACAGCGCGAAAGCGATCGCAGCCGGTGCGTGCTACGATGGCGATACTTGGGACTTCTTCAAGGGCAC
CCTGGTTCAAAAGGCGCTGATGATCTTCGATGTAATCACCCTGGCGGCTACGGGTTCCGAAATGAACTGGGGTAG
CGTGATCACCAACGAGGAAACCCAACAGAAATACTCCATTCACAACAACCATCTGTTCCCTAAAGTTAGCGTAATC
AACCCGAAGCTGCAGGCGACTGTGTCTCGTGATTACCTGGTTTATTCTGCTGCCGATATCATCGCACACTCTATCG
AAGCGTATTTCACCGCAGAATACCGCCCGGAGATCATCGACTTTCTGGTAGAGAGCAACATCAAAACTGTCATCC
GTACCACGGAAATCCTGCTGAACGACCCGCAGGATCTGAACGCGCGCGGCGAATTTGCATGGGCCGCTACGCTG
GCACTGAATGGCCTGACCCACCTGGGTATCAGCCCGTACGGCTTTCCGAACCACATGATCGAACACAGCATGTCT
GCGATTAGCGACGTTCCGCACGGTGCAGGTCTGAGCGTGATCATGCCGGCGTGGATGCAGTGGTATCAGAGCCA
GCGTCCGGCGCAGTTCAAACGTTTTGCCAAAGAAATCTTCGGTCTGGAGAACGCGGAAGAGGGTATCCAGGCGC
TGAAAACTTGGTTCGACAAAATTGGTACTCCAACGCGTCTGGAACAGCTGGGTATTGACGACAAAACCCTGTTTGA
AATTGTTGATAATGCAGTTCAGACTGCCATCCGTGCCAAAGTTGAAAAAACCTACACCAAAGAGGCGATTAAAGCG
ATCTTCGCCCTGGCAAAATAA

adh19

ATGCAGAACTTCCGTTTTTACGTTCCGACCGACATTCGTTTTGGCACGGACCGTCTGAGCGAACTGCCAGAAGCG
CTGTCCAGCATCGGCAAACGCGTTCTGCTGGTATACGGCGGCGGCTCTATCAAAAAGAGCGGTCTGTATGATAAA
GTCCAGCAGCAGCTGGTTAAGGGCGGTTTCGAAGTCGTAGAACTGAGCGGTATCGAACCGAATCCGAAAATTCAC
AGCGTTCGTTCCGGTGTTAAACTGATCCGCAAACACCAGCTGGAAGTAATCCTGGCGATCGGTGGTGGCAGCGT
GATCGATACTGCTAAAGTAATCGCGGCTGGCGTATTCTATGAAGGCGACCCGTGGGACCTGGTTGTCGACAGCA
GCAAAATCAAACAAGCTCTGCCGATCGTTGACATCCTGACCCTGGCGGCAACCGGCACCGAAATGAACCGCAATG
CGGTGATTTCTAACGCTGATACCAAAGAAAAACTGGGTACCGGCGGCGCTGAACTGCTGCCGCACGTTTCTTTTC
TGGACCCTACCACCACTTTCTCCGTTTCCAAATGGCAGACCGCGGCTGGTGCGGCAGATATCCTGTCCCACCTGT
TCGAACAGTACTTCAACCGTACCAAAGCTGTCCAAGTTCAAGACAACATTGCGGAAGCCCTGATGAAGGTAGTAAT
CGATTTCGCTCCGGTGGCCTTTGAAAATCCTGACGACTATGGCGCACGTGCGAACCTGCTGTGGGCTAGCACCCT
GGCTCTGAACGGCCTGGTTGGCAACGGTCGCTCCGGTGGCTGGACTTGCCACCCGATCGAACATGAGCTGTCTG
CTTTCTATGATATCACTCATGGCATTGGTCTGGCTATCCTGACCCCGCGCTGGATGGCTTACTGTATCACCCACGA
CCCAAGCACCCATGAGAAATTCGCTAAGTATGGTGAAGCTGTGTGGGGCCTGACGGGTGACTCCCAGGCAGAAA
TCGCTCGTAAGGCGGTACGTACTACTTACGAATTTTTCGAAGACCAGCTGGAAATTCCGATGACCCTGCAGAAAGT
GGGTATCGAAACGACGGAACTGGTGGACGAAATGTCTCAGCAGGCGGTGGTGCATGGCGAGCTGAACACTGACG
GCCCGTTCGTTCGCCTGGACGAAGAAGCGGTTAAAAGCATCATTACCAGCTGTTTCGAAGAAATGACTCTGAACT
AA

adh20

ATGGAAAATTTCAACTACTCTATTCCGACCGAAATCTATTTCGGTAAGGGTCAGATCAAAAACCTGGGTAATGCCAT
CAAAAAGTACGGTAGCAAAGTGCTGGTTGTGTATGGCGGCGGTAGCATCAAACGTATCGGCCTGTACGACGACAT
GATGAAAATCCTGAAAGACAACAACATCTCTTACGTTGAACTGTCCAACATCGCACCAAACCCGCGCATCGAATCT
GTGCGTGACGGCGTGAAACTGTGCCGTGATAACGACGTTGAAGTCGTTCTGGCAGTCGGTGGCGGCAGCACGAT
CGACTGCGCTAAAGTAATCGCAGCAGGTGTAAACTATGACAAAGACCCGTGGGACCTGGTTCTGGACAGCTCTAA
AATGACCACGGTTCTGCCAGTTATCACTATTCTGACCCTGAGCGCAACCGGTTCTGAAATGGACCCATTCGCAGT
GATCTCTGACATGAGCAAAAACGAAAAGGTAGGCGTGGGTAACGATAAAATGAAACCGAAAGTGTCTATCCTGGA
CCCGGAATACACCTACAGCGTTCCGAAAAACCAAACCGCGGCGGGCACGGCTGATATCATGTCCCACATCTTCGA
AAACTACTTCAACAACACGAAAGGCGCCTTTATCCAGGCACGCACTGCTGAAGGTCTGCTGAAAGCTTGTATGAAA
TATGGCAAGATCGCCATCGAGGAACCAAACAACTACGAAGCTCGTGCAAACCTGATGTGGGCATCCAGCCTGGC
CATTAACGGCCTGATCTCTTACGGCACCGCGGGTGCGTGGAGCGTCCACCCGATGGAACACGAACTGTCTGCAT
TCTACGACATTACGCATGGTGTCGGTCTGGCGATCCTGACCCCGCATTGGATGCGCTACGTCCTGAACGAAGATA
CTCTGGACAAATTCGTTGAATATGGCATTAACGTGTGGGAGCTGGACAAAAACCTGGACAAATATACCATCGCGAA
CACTGCGATCGATAAAACGGCAGAATTTTTCAAGGAAATGGGTATTCCGAGCACTCTGCGCGAAGTGGGTATTGG
CGAGGAGAACTTCAAAATCATGGCGCAAAAAGCGGTTAAAAGCGGCCTGGAATACGGTTTCAAACCGCTGGCTCC
GGAAGATGTGGTTAACATCTACAAAGCTGCTCTGTAA

adh21

ATGACCGAAGCGATGGAAAACTTCATTTACGATATCCCGACCAAAGTTTACTTTGGCAAAGGTCAGCTGAACCAGC
TGGCAGATATCGTGCAGGTCTATGGCCAGCGTGTTCTGCTGGTTTATGGTGGCGGTTCCATTCAACGTAACGGCA
TTTATGACGCAGCAGTTGCACAACTGAAAAAGGCGGGTAAGAAGTACACCGAACTGTCTGGTGTTGAACCGAATC
CGTCCATCCACACGGTCGAAAAAGGTGTTATGCTGTGCCAGCGTGAACAGATCGACATGCTGGTAGCTATTGGTG
GCGGCAGCGCTATCGATTGTGCCAAGGTGATCAGCGCCGCCGCCTGCTCCACCCGTCGTCCGTGGGAACTGGTT
ACGCACCCAGAGGAAATTCAGCGCGCGCTGCCGGTTATCGCCGTTCTGACCATCGCTGCAACCGGTAGCGAAAT
GGATCACATTGCTGTGATCACGAATCCGCAAACCAAAGAGAAAATTGGCACCCGTCACCCGCTGCTGCGTCCGAA
AGCCGCTATCCTGGATCCGAGCTTCACTTTCTCTGTTAACGCTTACCAAAGCGCCTGTGGCGTGGCTGACATCAT
GTCTCATACGATGGAGTCCTATTTTGCCCGTAAAGAAGCAGCCCTGCAAGATCGTTTCGCAGAGGGTATTCTGAA
GATCTGCCTGACCTACGGCCCGATCGTGCTGCAGCAGCCGGATAATTACGAGGCGCGTAGCAACCTGATGTGGG
CAGCTTCTTGGGCTATCAACGATCTGCTGAAACTGGGTCACATGACTCAGTGGTCCGTGCACCCGATGGAACACC
CGCTGAGCGCGTTCTACAGCGTTACTCACGGCGAAGGTCTGGCGATTCTGACCCCGCATTGGATGGATTACGTC
CTGTCTGAAGCTACCGTTGGTAAATTCGCGTGTTTCGCTCGCGAAGTATGGCAGGTTCGTGAAATGGATCCATGG
GACATGGCTCGTGAGGGCATCGAACGCCTGCGTGGTTTCTACAAACAGCTGCAGCTGCGCAGCAGCCTGGGCGA
ACTGGGCATCGACGAAACCCACTTTGACGCTATGGCAGCAGACGCAGCCAAACAAACCGTTAACGGTTACGTTGC
ACTGCTGGCAGAAGATGTCAAGAACATCTACCGTAACTCCCTGTAA

adh22

ATGAATAACTTCACCTACAGCATCCCGACTAAGATTCACTTCGGCAAAGGTCAGATCTCCCACCTGTCTGAGCTGT
CCGAGTCTGGCAACAAAGTTCTGCTGTGCTATGGCGGCGGCAGCATCAAGAAAGCGGGCATCTATGACGAAGCC
GTGAAAATCCTGAAAGAAGAAGATATGGAAATCTTCGAACTGTCTGGTATTGCACCGAACCCGAAAATCGAAAGCG
TACGTGAGGGTGTGAAACTGTGCAAGGAGAACTCCATTGATATGGTTCTGGCTATCGGCGGCGGCTCCGTAATCG
ATTGTGCTAAAGTCGTAGCCGCTGGTGCCTGTTATGACGGCGACCCGTGGGACCTGGTTATTACCCCGCGTTGGA
TCAAAAAGGCACTGCCGATTTACTCCGTACTGACTCTGAGCGCGACCGGCTCCGAGATGGATAAATTCGCAGTAA
TCTCTGACATGTCTAAAAACGAAAAATGGGGTACTGCGTCTGATCACATGAAGCCGAAAATGTCTATTCTGGACCC
GGAATACACGTATTCTGTTAGCAAAAAACAAACTGCCGCGGGTACCGCGGATATCATCTCTCATATCTGCGAAAAC
TACTTCACGAACGTTAAAAACGCAGATGTGCAGGCTCGCTTCGCGGAAGGTCTGCTGAAAAATTGCTTTAAGTATG
GCCCGGTTGCCCTGGAAGAACCGGATAATTATGACGCTCGTGCGAACCTGATGTGGACCGCATCTATGGCGATTA
ACGGTATGATTCAGTACGGCGCTGAAGTGGCGTGGTGTGTACATCCGATGGAACATGAACTGTCCGCGTTCTATG
ATATCACCCACGGTGAAGGTCTGGCAATTCTGACCCCGCATTGGATGGAGTTTGCGCTGAACGATGATACCGCTT
ATAAATTCGCTGATTACGCACGCAACGTTTGGGACGTTGTTAACGACGACGACATGGCCGCAGCAAAAGAGGGTA
TCGCGTACACTCGCGAATACTTTAAAAAGATGGGTCTGCCGCAGACCCTGACCGACGTGGGTATCGATAAAGAAT
ATTTCGACATCATGGCTCAGAAAGCTGCGGATGGCTGCAAGGGCAGCTTCGTGCCACTGAGCAAAGAAGACATC
GTTTCCATCTATGAAGCTGCCCTGTAA

sadh1

ATGTTCATGAAAGGTCTGACGTATCTGAAGCCGGGCATTGTATCCTGGCAAAGCATCCCGAAACCGGTTCTGAAA
AAACCGACTGACGTGATTGGTAAAGTTGTCACCACTACCATCTGCGGCTCCGACCTGCACATTCTGAAAGGTGAT
GTTCCTGAGACCACCGCGCTGGCGGCCACTACGGGTCATGGTGTGGTCCTGGGCCACGAAGCCATCATCGAAAT
TGAATCTGTAGGCGACGCTGTGAAAAACTTTAGCAAAGGCGACGTTTGTATCGTATCCTGCATCACCTCTTGTGGT
AAATGCTACTACTGCAAACGCAACCTGCAGTCTCACTGTACCGGCCACATGGAAGGCACTTCCGGTTGGGTGTTC
GGTCATGAAATCGATGGCACCCAGGCCGAATACGTGCGTGTTCCGTGCGCTGATTATGGCCTGTACAAGGTTCCT
GAAGGCGTTGCGTACGAAAAACTGCTGATGCTGTCCGACGCCATTCCAACCTCCTACGAAATCGGCATCCTGAAT
GGCGAGGTGAAGGAAGGCGACTCCGTTGCTGTTGTCGGCCTGGGTCCGGTAGGTCTGAGCGCACTGCTGACCG
CAATTAACAAAAAACCGAAACAACTGATCGCAATCGACATGGACGAGAACCGTCTGGACCTGGCTAAGCAGCTGG
GTGCGACGCATATCATCAACTCCACCAACATGCCGAACGAAGAGGTCGCGAAGAAAGTGCAGGAGATCAGCAAA
GACCTGGAACCGGGCCGTGAATCCGGCGTGGACGTCGCTATTGAATGTGTGGGCGTCCCTCCGACCTTCGAGCT
GTGTGAAGACCTGATCGCATCTGGTGGCACTATTGCTAACGTGGGCGTCCACGGTGCGAAAGTAGACCTGAAACT
GCAAGAACTGTGGATCAAGAACTGTAAGATTACCACGGGTCTGGTGAGCACCTACTCCATCCCGGACCTGCTGAA
CCAGGTTGCAGACGGTTCCCTGGACCCTAGCCCAATCATTACCCATCACTTCAAATTTGACGAATTCGAGAAAGCT
TACCAGGTTTTCAAAGACGCTAAAAATACCAAAGCGATGAAAATTATTCTGACGCCG

sadh2

ATGTCTAAGAACTACAAAAAGCACCAGGCATATATCGCAGGTCGCGACAAAATGAAAGTTGCTGTTATTACTGGTG
CTTCTCGTGGCATCGGCGAAGCTATTGCTAAAGCCCTGGCGGAAGATGGCTACTCTCTGGCGCTGGGTGCGCGC
TCCGTTGACCGTCTGGAAAAAATCGCAAAAGAGCTGAGCGAGAAACACGGTGTGGAAGTCTTTTACGATTATCTG
GACGTGTCCAAACCGGAATCTGTAGAAGAATTTGCACGCAAAACGCTGGCACACTTCGGTGACGTTGATGTTGTT
GTGGCGAACGCCGGTCTGGGTTATTTCGGTCGCCTGGAAGAGCTGACCGAAGAACAGTTCCACGAAATGATCGA
AGTTAACCTGCTGGGCGTTTGGCGTACCATCAAAGCATTCCTGAACTCCCTGAAACGTACTGGTGGCGTTGCGAT
CGTCGTTACCTCCGATGTCTCCGCTCGTCTGCTGCCGTACGGCGGCGGTTATGTCGCAACGAAATGGGCAGCTC
GCGCGCTGGTACGCACGTTCCAGATCGAAAACCCTGACGTGCGTTTCTTCGAACTGCGTCCGGGTGCTGTTGAC
ACCTACTTCGGCGGTAGCAAAGCCGGCAAACCGAAGGAGCAGGGCTACCTGAAACCGGAAGAAGTTGCGGAAGC
AGTTAAATACCTGCTGCGTCTGCCGAAGGACGTTCGTGTTGAGGAGCTGATGCTGCGCTCTATTTACCAGAAACC
GGAATAC

sadh3

ATGAAAATCTCTCTGACTTCCGCACGTCAGCTGGCCCGTGATATCCTGGCCGCGCAGCAGGTGCCGGCCGATAT
CGCGGACGACGTGGCTGAACATCTGGTTGAATCTGACCGTTGTGGCTATATCTCCCACGGCCTGTCTATCCTGCC
GAACTATCGTACTGCGCTGGACGGCCACTCTGTTAACCCGCAAGGTCGTGCAAAATGTGTGCTGGATCAGGGTAC
CCTGATGGTCTTCGACGGCGATGGTGGTTTTGGTCAACACGTAGGCAAATCTGTTATGCAGGCTGCGATTGAACG
CGTTCGTCAGCACGGTCACTGTATCGTTACCCTGCGTCGTTCCCACCACCTGGGTCGCATGGGCCACTACGGCG
AAATGGCAGCTGCTGCGGGCTTTGTTCTGCTGTCCTTCACCAATGTGATCAACCGTGCGCCAGTTGTGGCGCCGT
TCGGCGGTCGTGTTGCCCGTCTGACTACTAACCCGCTGTGCTTTGCTGGCCCGATGCCAAACGGCCGCCCGCCT
CTGGTAGTTGACATCGCAACCAGCGCCATCGCCATCAACAAAGCTCGCGTTCTGGCCGAAAAAGGCGAACCAGC
ACCAGAAGGCTCTATCATCGGTGCGGATGGTAACCCGACCACCGATGCCTCTACCATGTTTGGTGAACACCCGG
GTGCTCTGCTGCCATTCGGTGGTCACAAAGGTTATGCACTGGGTGTTGTAGCTGAACTGCTGGCAGGTGTGCTGA
GCGGCGGCGGTACGATCCAGCCGGACAACCCGCGTGGCGGTGTGGCGACTAACAACCTGTTCGCTGTTCTGCT
GAACCCGGCGCTGGATCTGGGCCTGGACTGGCAGAGCGCTGAAGTTGAGGCTTTCGTTCGTTACCTGCATGATA
CCCCGCCGGCCCCGGGTGTTGACCGTGTTCAGTACCCGGGCGAATATGAAGCAGCGAACCGTGCGCAGGCTTC
CGACACGCTGAACATCAACCCGGCGATCTGGCGCAACCTGGAGCGCCTGGCGCAGTCTCTGAACGTCGCAGTAC
CGACCGCT

sadh4

ATGAAGGGTTTCGCTATGCTGTCCATCGGCAAGGTGGGCTGGATTGAAAAAGAGAAACCGGCGCCTGGTCCGTT
CGACGCCATTGTTCGCCCGCTGGCCGTAGCGCCGTGTACCTCTGATATCCACACTGTGTTCGAGGGTGCGATTG
GCGAACGCCACAACATGATCCTGGGCCACGAGGCGGTCGGCGAAGTAGTGGAAGTTGGTAGCGAGGTAAAGGA
TTTCAAACCGGGCGATCGCGTCGTTGTGCCGGCAATCACTCCGGATTGGCGTACTAGCGAAGTACAGCGCGGTT
ACCATCAGCACAGCGGCGGCATGCTGGCGGGCTGGAAATTCAGCAATGTTAAAGACGGTGTTTTCGGTGAATTTT
TCCACGTGAACGACGCCGATATGAACCTGGCGCACCTGCCGAAAGAAATTCCGCTGGAAGCAGCGGTTATGATC
CCGGACATGATGACTACTGGCTTCCACGGTGCTGAACTGGCTGATATTGAACTGGGTGCGACCGTAGCAGTACTG
GGCATCGGTCCGGTGGGCCTGATGGCAGTGGCCGGCGCTAAACTGCGTGGTGCGGGTCGCATCATCGCCGTTG
GCTCCCGCCCGGTGTGCGTCGACGCTGCGAAATACTATGGCGCGACGGATATTGTTAACTACAAAGACGGCCCG
ATTGAATCTCAGATCATGAACCTGACTGAAGGTAAAGGTGTTGACGCTGCTATCATCGCCGGCGGCAATGCAGAC
ATCATGGCGACTGCCGTGAAAATTGTGAAACCGGGTGGTACCATCGCGAACGTTAACTATTTCGGTGAGGGTGAA
GTTCTGCCAGTACCGCGTCTGGAGTGGGGTTGTGGCATGGCCCATAAAACCATCAAAGGTGGCCTGTGTCCGGG
CGGTCGTCTGCGCATGGAACGTCTGATTGACCTGGTCTTCTATAAACGTGTGGACCCGTCTAAACTGGTCACCCA
CGTATTCCGTGGCTTCGATAACATTGAAAAAGCTTTCATGCTGATGAAAGACAAACCGAAAGACCTGATTAAGCCG
GTAGTCATCCTGGCG

sadh5

ATGAAAGGCTTCGCTATGCTGGGTATCAACAAACTGGGTTGGATCGAAAAGGAACGTCCGGTGGCAGGCTCTTAC
GATGCTATTGTTCGTCCACTGGCTGTCTCTCCGTGCACCTCCGATATCCACACTGTCTTTGAGGGCGCGCTGGGT
GATCGTAAAAACATGATCCTGGGCCACGAAGCAGTCGGCGAGGTCGTTGAAGTTGGTTCTGAGGTGAAAGACTTC
AAACCAGGCGATCGTGTTATCGTTCCGTGTACCACCCCGGACTGGCGCTCTCTGGAAGTACAAGCAGGCTTCCAG
CAGCATAGCAATGGTATGCTGGCAGGCTGGAAGTTCAGCAACTTTAAGGACGGTGTGTTCGGCGAATACTTCCAC
GTTAACGACGCAGATATGAACCTGGCGATCCTGCCGAAAGATATGCCGCTGGAAAACGCGGTTATGATCACCGAC
ATGATGACTACCGGCTTCCACGGTGCAGAGCTGGCTGACATCCAAATGGGTTCTTCTGTCGTGGTGATCGGCATC
GGCGCTGTTGGCCTGATGGGCATCGCTGGCGCTAAGCTGCGCGGCGCCGGCCGTATTATCGGCGTTGGCAGCC
GTCCGATCTGCGTGGAAGCTGCTAAATTCTATGGCGCCACCGACATTCTGAACTACAAAAATGGTCACATCGTGG
ATCAGGTGATGAAACTGACCAACGGTAAAGGTGTAGACCGTGTAATCATGGCAGGTGGTGGTTCTGAAACCCTGA
GCCAGGCGGTATCTATGGTGAAACCGGGTGGCATCATCTCCAACATCAACTATCACGGCTCTGGCGATGCGCTG
CTGATCCCGCGTGTGGAGTGGGGTTGCGGTATGGCTCATAAAACCATCAAGGGCGGCCTGTGTCCGGGCGGTC
GTCTGCGCGCAGAAATGCTGCGCGATATGGTGGTATATAACCGTGTAGACCTGAGCAAGCTGGTTACCCATGTTT
ATCACGGTTTCGACCACATCGAGGAAGCCCTGCTGCTGATGAAAGACAAACCTAAAGATCTGATCAAAGCTGTTGT
AATTCTG

sadh6

ATGCGCGCACTGGCTTATTTCGGCAAGCAGGATATCCGTTACACCAAGGATCTGGAAGAACCGGTTATCGAGACC
GACGACGGCATCGAAATTGAAGTCAGCTGGTGCGGTATCTGTGGCTCCGACCTGCACGAATATCTGGATGGTCC
GATTTTTTTCCCTGAAGATGGTAAAGTGCATGACGTTAGCGGTCTGGGCCTGCCACAGGCGATGGGTCACGAGAT
GTCTGGCATCGTATCTAAAGTGGGCCCGAAAGTTACCAACATCAAAGCTGGTGACCACGTGGTCGTCGAGGCAAC
TGGCACTTGTCTGGATCATTACACCTGGCCGAACGCGGCTCATGCGAAAGATGCTGAATGCGCGGCGTGCCAGC
GTGGCTTTTACAACTGCTGCGCCCACCTGGGTTTCATGGGTCTGGGTGTTCACAGCGGCGGTTTCGCGGAAAAA
GTGGTGGTGAGCGAGAAACATGTTGTTAAGATCCCGAACACCCTGCCGCTGGACGTTGCAGCTCTGGTCGAACC
AATTTCTGTCTCCTGGCACGCGGTTCGTATCAGCAAGCTGCAAAAAGGCCAATCTGCTCTGGTGCTGGGCGCTGG
CCCAATTGGCCTGGCCACCATCCTGGCGCTGCAGGGTCACGGTGCAAGCAAAATCGTTGTATCCGAACCAGCGG
AAATCCGTCGCAATCAAGCAGCAAAACTGGGCGTTGAAACGTTCGATCCGTCCGAACACAAAGAAGACGCGGTTA
ACATCCTGAAGAAACTGGCACCGGGTGGTGAGGGTTTCGATTTCGCCTACGACTGTAGCGGTGTCAAACCTACCT
TTGATACTGGTGTACACGCTACCACCTTCCGCGGTATGTACGTGAACATCGCAATTTGGGGTCATAAACCGATCGA
TTTCAAACCGATGGACGTGACTCTGCAGGAGAAGTTCGTCACGGGTTCCATGTGCTACACCATTAAGGATTTCGAA
GATGTGGTTCAGGCTCTGGGTAACGGCTCCATCGCCATCGACAAGGCGCGCCACCTGATTACTGGCCGCCAGAA
AATTGAAGATGGCTTCACCAAAGGCTTCGACGAACTGATGAACCATAAAGAAAAAAACATCAAGATCCTGCTGACT
CCTAATAACCACGGCGAACTGGACGCGACCAAC

sadh7

ATGAAAGCAGCCGTGTTCAAAGGTAAGAACCGTATCGTACTGGATGAAAAGCCGGTGCCGGTGCCAAAACATAGC
GAAGCGCTGATTAAAATCACTACCACTACCATCTGCGGTACCGACATTCACATCCTGAAGGGCGAATACCCGGTA
GCGGAAGGCCTGACCATTGGCCACGAACCGATCGGTGTGATTGAGAGCTTCGGCGACGGTGTTACCGGCTTCAA
GAAAGGCCAGCGTGTTGTTATTGGCGCTATCACGCCGTGCAGCTCTTGTAGCAGCTGTATGGAAGGCATCCGTTC
TCAGTGTGGCAGCAAACTGATGGGCGGCTGGAAATTCGGTAATACCATCGATGGCTCTCAGGCGGAATATCTGAT
CGTGCCGGACGCTGCAGCGAACATGTACCCTATTCCGGATGGTATTACCGATGAACAGGTGCTGATGTGTCCGG
ACATTATGTCCACTGGTTTCTCCGGTCCAGAATCTGCCGGTGTGAAAGTGGGTGACACCGTCGTTATTTATGCTCA
GGGTCCGATCGGTCTGGGCGCTACCGCGGGTGCCAAAATGATGGGCGCAACGAAAGTAATTGTTGTAGATCGTT
TCCCGGAGCGTCTGGCGCTGGCAAAAAAGCTGGGCGCGGATTATACGCTGGATTTCACCAAATGTAACCCTATTG
AAGAAGTGATGCGTCTGACTGGTGGTCGTGGCGTTGACGTTGCAATTGAAGCGCTGGGTCTGCAGTCCACCTTTG
AGTCTTGCCTGCGTTGCCTGAAACCGGGCGGTGTGCTGAGCAGCCTGGGTGTCTATTCTGATGACCTGCGTCTG
CCGAACGACGCTTTCGCTGCGGGCCTGGGCGACTTCAAGATCGTTACTACCCTGTGTCCTGGCGGCAAAGAGCG
TATGCGTCGTCTGCTGTCTGTAATCGAGTCTGGCCGTGTTGACATGCGCCCGATGGTGACTCACACCTTTAAGCT
GGATGAAATCGAAAAAGCCTACGACCTGTTCGGTAACCAGCGCGATGGCGTTCTGAAAGTAGCCATCAAGCCG

sadh8

ATGACCGCGATGATGAAGGCAGCGGTATTTGTTGAACCTGGTCGTATTGAACTGGCGGATAAACCGATTCCGGAT
ATCGGTCCGAACGACGCACTGGTTCGCATTACCACCACCACTATTTGCGGCACCGACGTTCACATCCTGAAGGGC
GAATATCCGGTTGCTAAGGGCCTGACCGTGGGCCACGAGCCGGTTGGCATCATCGAGAAACTGGGCAGCGCAGT
AACGGGCTATCGCGAAGGTCAGCGTGTAATCGCAGGTGCGATCTGTCCTAACTTCAACTCTTACGCGGCGCAGG
ATGGTGTTGCGTCTCAGGATGGCAGCTACCTGATGGCTTCTGGCCAGTGCGGCTGTCATGGTTACAAAGCAACCG
CCGGCTGGCGTTTCGGTAATATGATTGACGGTACCCAGGCTGAATATGTTCTGGTACCGGATGCGCAGGCCAACC
TGACCCCGATTCCGGACGGCCTGACGGACGAACAGGTCCTGATGTGTCCGGATATTATGTCCACCGGTTTCAAAG
GTGCGGAGAACGCAAACATTCGCATTGGTGACACCGTTGCGGTGTTTGCGCAGGGTCCGATCGGTCTGTGCGCG
ACTGCTGGTGCGCGTCTGTGCGGCGCAACTACTATCATCGCTATCGACGGCAACGATCATCGTCTGGAAATCGCG
CGCAAAATGGGTGCGGACGTTGTGCTGAACTTCCGTAACTGCGACGTGGTCGACGAGGTGATGAAACTGACCGG
CGGTCGTGGTGTGGATGCGTCTATTGAAGCACTGGGTACCCAGGCGACCTTCGAACAATCTCTGCGTGTTCTGAA
GCCGGGTGGTACGCTGTCTTCTCTGGGTGTCTACTCCTCTGATCTGACTATCCCGCTGAGCGCCTTCGCAGCCG
GTCTGGGCGACCACAAAATTAACACCGCCCTGTGCCCGGGTGGCAAAGAACGTATGCGTCGTCTGATTAACGTTA
TCGAATCCGGTCGCGTGGACCTGGGTGCGCTGGTAACCCACCAGTATCGCCTGGATGATATCGTTGCGGCGTAC
GACCTGTTCGCTAACCAGCGTGATGGTGTTCTGAAAATCGCCATCAAACCGCAC

sadh9

ATGACCTTCGAACTGCCGAAAACTATGAAAGCGTTTGCGATGCGCAAGATCGGTGAAGTCGGCTGGATTGAAAAG
CCGGTGCCAGAATGCGGTCCGAACGATGCAATCTGCCGTCCGCTGGCTCTGGCTCCGTGCACCTCTGACATTCA
CACTGTGTGGGCGGGTGCGATCGGTGAACGCCATGACATGATCCTGGGCCATGAAGCCGTTGGCCAAGTTGTTA
AAATCGGTTCCGAAGTTAAAAACCTGAAAGTTGGCGATAAGGTTCTGGTTCCGGCAGTAACTCCGGATTGGGGTA
GCGAAGCAGCACAGGAAGGCTTCCCGGCACACTCCGGCGGTATGCTGGGTGGTTGGAAATTTTCCAACTTCAAA
GATGGCGTTTTTGCTGAATACTTCCACGTCAACGAAGCCGACGCTAACCTGGCCAAACTGCCGGAAGGTCTGACG
CCGCGTGATGTAATCATGTGCAGCGACATGATGACTACCGGTTTCCATGGTGCAGAACTGGCGGAGGTTAAGCTG
GGTGATATCGTAGTAGTTATCGGTATCGGTCCGGTTGGTCTGATGTCTGTACGTGGCGCTGCTCTGATGGGCGCA
AGCCGTATCTTCGCAGTCGGTTCCCGTCCGCACTGCTGCGACACCGCAGTCCAGTACGGTGCCACTGACATCAT
CAACTATAAAAATGGTGATATCGTGGAACAGATCCTGAAAGCTACTGGCGGTAAAGGCGTTGATCGTGTGATTATC
GCGGGTGGCGACGTCCATACTTTCGCGCAAGCTATTAAAATGATCCGTGCTGGCGGCGTTATCGGCAACGTGAA
CTACCTGGGCGAAGGTGAAATGATCGACGTACCGCGTGTAGAATGGGGTGTTGGCATGGGTCACAAATTCATCCA
CGGTGGTCTGACCCCGGGTGGTGCACTGCGTATGGAAAAAATGGCTAACCTGATCAAGTATAAAAAAGTCGATCC
GACGAAACTGATTACCCATGAATTCAAGGGTCTGGAAAAAGTTGAAGATGCTCTGATGCTGATGAAAGACAAACCG
GTTGACCTGATCAAACCGGTTGTTCTGATTGAATACAACGACAAACTG

sadh10

ATGATGGCGACCATGAAAGCAGCGATCTTTGTTGAAAAAAACCGTATTGTTCTGGAAGACAAACCAATCCCGGAAG
TTGGCCCGCTGGATGCACTGATCCGTATCACCACCACTACGATTTGCGGCACCGATGTTCACATCCTGCGCGGTG
AGTATCCGGTAGCGAAGGGTCTGACGATCGGTCATGAACCAGTAGGTATTATCGAACGTCTGGGTTCCCAGGTTC
GTGGTTTCGTAGAAGGTCAGCGTGTTATTGCAGGCGCGATCACCCCGTCTGGCCAAAGCTACGCATGCCTGTGT
GGCTGTGCCAGCCAGGACGGTCCGGATACCCGTCACGGTTTTCGTGCGACCGGCGGCTGGAAATTCGGCAACAT
TATCGACGGCTGCCAGGCGGAGTATGTACTGGTGCCAGACGCGCTGGCGAACCTGTGCCCAATTCCGGATGGCC
TGAGCGACGAACAAGTTCTGATGTGCCCGGACATCATGTCCACTGGTTTCTCTGGCGCAGAACGTGGCGAAATTA
ACATTGGTGATACTGTTGCGGTATTCGCACTGGGTCCGATCGGCCTGTGTGCTGTGGCGGGCGCCCGTCTGAAG
GGTGCGACCACCATCATCGGTGTGGACGCAGTGGCTCAGCGTATGTCTGTTGCACGTCAGCTGGGCGCCACCCA
CGTGGTAAACTTCAAAGAGGCGAACGTTGTGGAACAGATTATGGCGCTGACGGACGGCCGTGGTGTTGATGTATC
TATCGAAGCACTGGGCACCCAGGGCACCTTCGAATCTGCTCTGCGTGTCCTGCGCCCGGGCGGTCGCCTGTCCT
CCCTGGGTGTTTATTCCAGCGACCTGCGTATCCCGCTGGACGCCTTCGCCGCAGGCCTGGGCGATTACTCCATC
GTCACTACCCTGTGTCCGGGTGGCAAAGAGCGTATGCGTCGCCTGATGGCCGTAGTTCAGAGCGGCGCGGTCGA
CCTGTCTCCGCTGGTCACTCACCACTTCAAGCTGGATGACATCGAAGCGGCATATGAACTGTTTGCGAACCAGCG
TGATGGTGTAATGAAAGTTGCAATCACCCCG

sadh11

ATGTCTAAAGTGGCTGCCGTGACCGGTGCAGGTCAGGGCATTGGTTTTGCAATCGCGAAACGCCTGTATAACGAC
GGTTTCAAAGTAGCCATCATCGATTACAACGAAGAAACCGCTCAGCAGGCTGCGAAAGAACTGGGTGGTGAATCT
TTCGCGCTGAAGGCGGATGTTTCTGACCGTGACCAGGTAGTCGCCGCGCTGGAAGCTGTTGTTGAGAAATTCGG
TGATCTGAACGTGGTAGTAAACAACGCGGGTATCGCCCCGACTACCCCGATCGAAACGATCACCCCGGAACAGTT
TCACCAGGTGTACAACATCAATGTTGGTGGTGTGCTGTGGGGTACCCAGGCTGCTACTGCTCTGTTCCGTAAACT
GGGCCACGGCGGTAAGATTATCAACGCCACCTCCCAGGCTGGTGTGGTCGGTAACCCGAACCTGATGCTGTACA
GCAGCTCCAAATTCGCTGTCCGCGGCATGACCCAGATCGCAGCACGTGACCTGGCAGAAGAGGGTATCACCGTC
AACGCCTACGCTCCAGGCATTGTGAAAACCCCGATGATGTTCGATATCGCTCATCAGGTGGGTAAGAACGCCGGC
AAAGACGACGAGTGGGGTATGCAGACCTTCGCTAAAGACATCGCGATGAAACGTCTGAGCGAGCCGGAAGATGT
AGCAAACGTTGTTTCCTTTCTGGCCGGCCCGGATTCCAACTACATCACCGGCCAGACTATCATTGTAGACGGTGG
CATGCAATTTCAT

sadh12

ATGCGTGCGCTGGCCTATTTCAAGAAAGGTGACATTCACTTCACTAATGACATCCCACGTCCGGAAATCCAGACG
GACGACGAAGTTATCATCGATGTGTCTTGGTGCGGCATCTGTGGTTCTGACCTGCACGAATACCTGGATGGCCCG
ATCTTTATGCCGAAAGACGGTGAGTGCCACAAACTGAGCAACGCCGCTCTGCCTCTGGCTATGGGCCACGAAATG
AGCGGTATCGTTTCCAAGGTTGGTCCGAAAGTTACTAAAGTAAAAGTGGGTGATCACGTAGTTGTTGATGCGGCC
AGCTCCTGCGCCGACCTGCACTGCTGGCCGCACTCCAAATTTTATAACTCTAAGCCGTGTGACGCATGTCAGCGT
GGTTCCGAAAACCTGTGCACCCACGCGGGCTTTGTAGGTCTGGGCGTCATTAGCGGTGGCTTCGCGGAGCAGGT
AGTTGTATCTCAACACCACATCATCCCGGTTCCGAAGGAGATCCCGCTGGACGTTGCAGCTCTGGTTGAACCGCT
GAGCGTTACTTGGCATGCAGTGAAAATTTCTGGTTTCAAAAAAGGTTCTTCCGCACTGGTTCTGGGCGCAGGTCC
GATCGGTCTGTGCACCATCCTGGTTCTGAAAGGCATGGGTGCATCCAAGATTGTCGTATCCGAAATTGCGGAGCG
CCGTATTGAAATGGCGAAAAAGCTGGGTGTTGAAGTATTCAACCCGTCTAAACACGGCCACAAATCCATTGAGATC
CTGCGTGGCCTGACCAAAAGCCATGATGGCTTTGACTACAGCTATGATTGCTCTGGCATCCAGGTGACTTTCGAA
ACTTCTCTGAAAGCGCTGACTTTCAAAGGTACCGCTACCAACATTGCAGTGTGGGGCCCGAAACCGGTTCCATTC
CAGCCGATGGATGTCACCCTGCAGGAAAAAGTAATGACCGGCTCCATTGGTTACGTTGTGGAAGATTTCGAAGAA
GTTGTGCGTGCTATTCATAACGGCGATATTGCTATGGAAGATTGTAAGCAGCTGATCACCGGTAAACAGCGTATCG
AAGATGGTTGGGAGAAAGGCTTTCAGGAACTGATGGATCACAAAGAATCTAACGTAAAAATTCTGCTGACTCCAAA
CAACCACGGTGAAATGAAA

sadh13

ATGGATCGTATCATTCAGTCTCCTGGCAAATACATCCAGGGCGCAGATGTGATCAACCGTCTGGGTGAATATCTGA
AACCGCTGGCAGAGCGTTGGCTGGTGGTTGGCGATAAATTCGTGCTGGGCTTTGCTCAGTCTACCGTGGAAAAAT
CCTTCAAGGATGCGGGCCTGGTGGTTGAAATTGCACCGTTCGGCGGTGAATGTAGCCACAACGAAATTGACCGTC
TGCGTGGCATCGCTGAAACTGCCCAATGTGGCGCGATTCTGGGCATCGGTGGCGGCAAGACCCTGGACACTGCC
AAAGCGCTGGCGCATTTCATGGGTGTCCCAGTTGCGATCGCTCCGACCATCGCGTCTACCGCTGCGCCGTGCTC
TGCCCTGTCCGTGATCTACACGGACGAAGGCGAGTTCGACCGCTACCTGCTGCTGCCAAACAACCCGAATATGGT
GATTGTGGACACCAAAATCGTTGCAGGTGCGCCGGCGCGTCTGCTGGCGGCTGGCATTGGTGACGCACTGGCG
ACCTGGTTCGAAGCACGCGCGTGCTCCCGTTCTGGTGCGACGACCATGGCTGGCTGCAAATGCACCCAGGCAGC
GCTGGCGCTGGCGGAACTGTGCTATAACACCCTGCTGGAGGAAGGCGAAAAAGCAATGCTGGCGGCGGAACAG
CACGTCGTAACCCCTGCTCTGGAACGTGTAATTGAAGCAAATACTTACCTGAGCGGCGTGGGTTTCGAATCCGGT
GGCCTGGCGGCTGCGCACGCTGTTCATAACGGCCTGACCGCGATCCCTGACGCGCACCACTACTATCACGGCGA
GAAAGTGGCTTTCGGCACCCTGACTCAACTGGTACTGGAAAACGCGCCGGTCCAGGAAATCGAAACCGTGGCGG
CGCTGTCCCATGCAGTAGGCCTGCCGATCACCCTGACCCAGCTGGACATCAAAGAAGACGTGCCAGCAAAAATG
CGTATCGTGGCCGAAGCGGCATGCGCAGAAGGTGAGACCATCCATAACATGCCGGGCGGCGCTACCCCGGATC
AGGTCTACGCCGCCCTGCTGGTAGCAGACCAATATGGCCAGCGTTTTCTGCAGGAATGGGAATAA

sadh14

ATGATGAAAGCGCTGTGTTACCTGGGTTCTAAAACTATCAAATGGCAGACCGTAGCCAAACCGACCCTGAAATCC
CCAACCGACGTTATCGGTAAACTGTCTGCACTGACCCTGTGTGGCTCTGATCTGCACATTATTGCGGGTCACGTA
AAAGAAAGCACCGATATCGCCGAGTCCCAGCCGGGCCGTGGCCTGATCCTGGGTCACGAAGGTATCATCAAAGT
GGAAGAGGTAGGCGATGCTGTCAAAAACTTCAAACGTGGTGACGTTTGCATCGTCTCTTGCATCACCAGCTGCGG
TGAATGCTACTACTGCAAACGTGATCTGCAGTCCCATTGCAACCGTACCGAAGGTACCTCCGGCTGGATCCTGGG
CCACGAAATTGATGGTACCCAGGCCGAATACGTCCGTATCCCGTACGCCGACCAGTCTCTGTACAAAGCTCCGGA
GAACGTGCCGATCGAGAGCCTGCTGATGCTGAGCGACATCCTGCCGACCGCGTACGAAGTGGGCGTTGTCTCTG
GCAATGTGAAAAAAGGTGATACCGTCGCCATCGTAGGTCTGGGCCCTGTTGGCCTGTCCGCGCTGCTGTCCGCT
AAAGCCCTGGACCCGGCTAAAATCATTGCAATTGATATGGACGACTCTCGTCTGGAGGTTGCACGTCGTCTGGGT
GCGCACGAAACTATCAATCCGGGCAAACAGGATGCTGCAAAGCTGGTTCATGAACTGACCGCGAGCGAAGGTAA
GTCTTCTGGTGTAGACGTTGCAATCGAATGTGTTGGCGTGCCGGCCACCTTTGAAATGTGCGAAGACCTGCTGTG
TCCGGGTGGCCACCTGTCTAACGTTGGCGTTCACGGTTCTAGCGTCGAACTGAAACTGCAGGAGCTGTGGATCAA
AAACATTTCTATCAGCACCGGTCTGGTATCTGCTTATTCCACCGAAACCCTGCTGCAGAAAGTTATCGACAAAAAA
CTGGACCCGACCCCTCTGGCAACCCACCATTTCAAGCTGAGCGAAATCGAAAAGGCGTACGATGTCTTCTCTCAC
GCCGCAGATAACCAGGCCATCAAAATGGTGATCACCTGTGACGAA

sadh15

ATGAGCATTCCGTCCAGCCAGTACGGCTTCGTGTTCAACAAACAGTCCGGCCTGAACCTGCGCAACGATCTGCCT
GTTCATAAACCGAAGGCGGGTCAGCTGCTGCTGAAAGTTGATGCTGTTGGTCTGTGCCATTCCGATCTGCACGTA
ATTTATGAGGGCCTGGATTGCGGCGATAACTACGTGATGGGCCACGAAATTGCAGGTACCGTGGCTGCGGTAGG
CGACGACGTTATCAACTACAAAGTAGGCGACCGTGTTGCTTGTGTGGGCCCGAACGGTTGCGGTGGTTGCAAGT
ATTGCCGTGGTGCAATCGACAACGTTTGCAAGAACGCATTTGGTGACTGGTTTGGTCTGGGCTACGATGGTGGTT
ATCAGCAGTATCTGCTGGTAACCCGTCCGCGTAACCTGAGCCGTATCCCAGATAACGTTTCCGCCGACGTAGCGG
CGGCATCTACCGACGCGGTACTGACTCCGTATCATGCTATTAAGATGGCACAGGTGTCTCCTACCAGCAACATCC
TGCTGATCGGTGCGGGCGGCCTGGGTGGTAACGCGATCCAGGTAGCGAAAGCCTTCGGTGCAAAAGTTACTGTA
CTGGATAAAAAAAAAGAGGCACGCGATCAAGCTAAAAAACTGGGTGCAGATGCGGTCTACGAAACTCTGCCGGAA
TCTATCTCTCCTGGTTCCTTCAGCGCTTGCTTCGATTTCGTCTCTGTACAGGCGACGTTCGATGTTTGCCAGAAAT
ATGTTGAGCCAAAAGGTGTAATTATGCCGGTTGGTCTGGGCGCGCCTAACCTGTCTTTTAACCTGGGTGATCTGG
CGCTGCGCGAAATTCGTATTCTGGGTTCCTTCTGGGGTACCACTAACGATCTGGACGACGTCCTGAAACTGGTTT
CCGAGGGCAAAGTGAAACCGGTTGTTCGCAGCGCAAAACTGAAAGAGCTGCCGGAGTATATCGAAAAACTGCGT
AACAACGCGTATGAGGGTCGTGTTGTCTTTAACCCG

sadh16

ATGCTGCGCACTAATCTGCGCACGTTCGCTCGTCCGCAGTTTATTCGCGGCCTGGCGACTGCGCCGGTAATCCC
GAAAACCCAAAAAGGCGTTATTTTCTACGAAAACGGTGGTGAACTGCAGTACAAGGACATCCCGGTACCGGAGCC
AAAACCGAACGAAATCCTGGTTAATGTTAAGTACAGCGGTGTTTGTCACACCGATCTGCACGCCTGGAAGGGTGA
CTGGCCACTGCCAGTGAAACTGCCGCTGGTGGGTGGTCACGAGGGCGCCGGCATCGTGGTAGCTAAGGGCAGC
GAAGTAAAAAACTTTGAGATCGGTGATTATGCGGGCATCAAATGGCTGAACGGCAGCTGTATGAGCTGTGAGCTG
TGCGAAAAAGGTTACGAATCCAACTGCCTGCAGGCCGATCTGAGCGGCTATACCCATGACGGTAGCTTTCAGCAA
TACGCTACCGCGGATGCGGTTCAGGCGGCTCAGATCCCGAAGAACGTAGACCTGGCTGAGATCAGCCCTATCCT
GTGCGCTGGTGTCACTGTTTATAAGGCGCTGAAAACCGCGGATCTGGCTCCGGGCCAGTGGGTCGCTATCTCTG
GCGCGGCTGGTGGTCTGGGTTCTCTGGCGGTGCAATACGCTAAAGCTATGGGTCTGCGTGTCCTGGGCATCGAC
GGTGGTGCAGAAAAAGAAAAGCTGTTTAAAAGCCTGGGTGGCGAGATCTTCATCGATTTCACCAAAGAAAAAAACA
TCGTAGAAGCCATTCAGGAGGCAACTAATGGCGGCCCGCATGGCGTCATCAACGTATCCGTTGCGGAAGCAGCG
ATCTCTCAGTCCACCGAGTATGTTCGTCCGACCGGTACCGTAGTACTGGTGGGTCTGCCAGCAGGTGCTGTCTGC
AAAAGCGAAGTGTTCTCTCACGTGGTAAAATCTATCTCTATCAAAGGCAGCTACGTGGGTAATCGTGCGGACACC
CGTGAGGCGATCGACTTCTTCGAACGTGGTCTGGTGCGTAGCCCGATCAAGATCGTGGGTCTGAGCGAACTGCC
GGAAGTGTACAAACTGATGGAGCAGGGCAAAATCCTGGGCCGTTACGTCGTGGATACTACCAAA

sadh17

ATGAAAGCGCTGCAATACACCGAAATCGGTAGCGAACCGGTGGTCGTTGATGTCCCGACCCCTGCGCCGGGTCC
GGGTGAAATTCTGCTGAAAGTAACCGCTGCGGGTCTGTGTCACAGCGACATCTTCGTTATGGACATGCCGGCTGA
ACAGTATATCTACGGCCTGCCGCTGACCCTGGGCCATGAAGGCGTGGGCACCGTTGCGGAACTGGGTGCGGGC
GTGACCGGTTTCGAAACTGGCGACGCTGTTGCAGTTTACGGCCCTTGGGGCTGTGGCGCTTGCCACGCATGCGC
ACGTGGCCGTGAGAACTACTGCACCCGTGCGGCAGAGCTGGGCATCACTCCTCCGGGCCTGGGCTCTCCGGGT
AGCATGGCTGAGTACATGATCGTTGACTCCGCTCGTCACCTGGTCCCGATCGGTGACCTGGACCCGGTGGCAGC
AGTGCCGCTGACCGATGCTGGCCTGACCCCGTACCACGCCATCAGCCGTGTACTGCCGCTGCTGGGCCCAGGTT
CTACGGCTGTTGTAATTGGCGTCGGTGGTCTGGGTCACGTTGGCATTCAGATTCTGCGCGCCGTTAGCGCCGCG
CGTGTTATCGCGGTTGATCTGGACGACGACCGCCTGGCGCTGGCGCGTGAAGTTGGTGCAGATGCGGCTGTTAA
ATCTGGTGCGGGTGCGGCGGATGCTATTCGTGAACTGACTGGTGGCGAAGGTGCTACTGCTGTTTTCGATTTTGT
GGGTGCGCAGTCTACCATCGACACCGCACAGCAAGTCGTGGCAATCGATGGCCACATCAGCGTTGTCGGCATCC
ACGCCGGTGCACACGCCAAAGTTGGTTTCTTCATGATCCCATTCGGTGCTAGCGTTGTTACCCCGTATTGGGGTA
CTCGCTCCGAGCTGATGGACGTTGTTGATCTGGCTCGTGCAGGCCGCCTGGATATTCATACCGAAACCTTCACCC
TGGACGAAGGCCCTACTGCATACCGCCGTCTGCGCGAGGGCTCTATCCGTGGCCGTGGCGTGGTTGTACCGGG
TTAA

Appendix 5: *Complete list of mutations identified in EMS-mutagenized strains*

Table A5.1 Complete list of mutations identified in EMS-mutagenized strains

Position	Mutation	Annotation	Gene	Description
28,255	G→A	Q35Q (CAG→CAA)	*yidF* →	putative Cys-type oxidative YidJ-maturating enzyme
31,135	G→A	R53H (CGC→CAC)	*ilvB* →	acetolactate synthase 2 large subunit
69,053	G→A	R73C (CGC→TGC)	*slmA* ←	nucleoid occlusion factor, anti-FtsZ division inhibitor
112,343	G→A	E109K (GAG→AAG)	*yibI* →	DUF3302 family inner membrane protein
238,834	C→T	A85V (GCC→GTC)	*prlC* →	oligopeptidase A
247,075	G→A	L241L (TTG→TTA)	*yhiN* →	putative oxidoreductase
260,536	C→T	G1346S (GGT→AGT)	*rhsB* ←	Rhs family putative polymorphic toxin, putative neighboring cell growth inhibitor
444,515	C→T	S87S (AGC→AGT)	*yhdN* →	DUF1992 family protein
541,570	C→T	G310D (GGC→GAC)	*kdsD* ←	D-arabinose 5-phosphate isomerase
1,460,131	T→C	I699V (ATC→GTC)	*pta* ←	phosphate acetyltransferase
1,647,157	C→T	L228L (CTC→CTT)	*dusC* →	tRNA-dihydrouridine synthase C
1,670,169	C→T	A332A (GCG→GCA)	*yehM* ←	uncharacterized protein
1,670,535	C→T	W210* (TGG→TGA)	*yehM* ←	uncharacterized protein
1,702,408	G→A	intergenic (+608/-1115)	*gatZ* → / → *gatB*	D-tagatose 1,6-bisphosphate aldolase 2, subunit/galactitol-specific enzyme IIB component of PTS
1,706,510	C→T	S61F (TCT→TTT)	*gatR* →	pseudogene, repressor for gat operon; interrupted by IS3; split galactitol utilization operon repressor, fragment 2; split galactitol utilization operon repressor, interrupted
1,707,736	G→A	R29C (CGC→TGC)	*insE1* ←	IS3 transposase A
1,708,920	C→T	G142E (GGA→GAA)	*yegS* ←	phosphatidylglycerol kinase, metal-dependent
1,727,249	C→T	A60V (GCG→GTG)	*yegI* →	protein kinase-related putative non-specific DNA-binding protein
1,731,667	C→T	D1042N (GAT→AAT)	*yegE* ←	putative diguanylate cyclase
1,736,966	C→T	Q175* (CAA→TAA)	*asmA* →	suppressor of OmpF assembly mutants; putative outer membrane protein assembly factor; inner membrane-anchored periplasmic protein
1,794,024	G→A	R426C (CGC→TGC)	*sbcB* ←	exodeoxyribonuclease I; exonuclease I
1,802,207	G→A	S357F (TCC→TTC)	*yeeR* ←	CP4-44 prophage; putative membrane protein

Position	Mutation	Annotation	Gene	Description
1,803,787	G→A	F910F (TTC→TTT)	*flu* ←	CP4-44 prophage; antigen 43 (Ag43) phase-variable biofilm formation autotransporter
1,817,088	C→T	intergenic (+49/-53)	*nac* → / → *cbl*	nitrogen assimilation regulon transcriptional regulator; autorepressor/ssuEADCB/tau ABCD operon transcriptional activator
1,818,599	G→A	G79R (GGA→AGA)	*yeeO* →	putative multdrug exporter, MATE family
1,819,846	G→A	S494S (TCG→TCA)	*yeeO* →	putative multdrug exporter, MATE family
1,833,695	C→T	R237H (CGT→CAT)	*mtfA* ←	anti-repressor for DgsA(Mlc)
1,837,987	G→A	G197G (GGC→GGT)	*yedY* ←	membrane-anchored, periplasmic TMAO, DMSO reductase
1,840,215	C→T	P105S (CCT→TCT)	*yedV* →	putative sensory kinase in two-component regulatory system with YedW
1,847,326	C→T	H64Y (CAT→TAT)	*vsr* →	DNA mismatch endonuclease of very short patch repair
1,855,825	A→G	I121T (ATA→ACA)	*fliP* ←	flagellar biosynthesis protein
1,856,899	C→T	L23L (TTG→TTA)	*fliN* ←	flagellar motor switching and energizing component
1,865,124	G→A	P28P (CCG→CCA)	*fliE* →	flagellar basal-body component
1,867,917	C→T	W106* (TGG→TGA)	*yedK* ←	DUF159 family protein
1,870,651	G→A	P417S (CCG→TCG)	*amyA* ←	cytoplasmic alpha-amylase
1,897,367	G→A	D309N (GAT→AAT)	*otsA* →	trehalose-6-phosphate synthase
1,980,323	C→T	R408R (CGG→CGA)	*sdaA* ←	L-serine dehydratase 1
1,981,882	C→T	A143T (GCA→ACA)	*nudL* ←	putative CoA pyrophosphohydrolase, weak 3-phosphohydroxypyruvate phosphatase
1,996,465	C→T	W34* (TGG→TGA)	*dmlA* ←	D-malate oxidase, NAD-dependent; putative tartrate dehydrogenase
2,045,484	C→T	R189R (AGG→AGA)	*xthA* ←	exonuclease III
2,065,339	G→A	intergenic (+282/-7)	*cedA* → / → *ydjO*	cell division modulator/uncharacterized protein
2,099,459	G→A	A95T (GCG→ACG)	*ydiP* →	putative DNA-binding transcriptional regulator
2,112,098	C→T	S769L (TCG→TTG)	*ydiJ* →	putative FAD-linked oxidoreductase
2,114,247	C→T	G52G (GGC→GGT)	*sufA* →	Fe-S cluster assembly protein
2,139,085	G→A	A136T (GCG→ACG)	*ydhB* →	LysR family putative transcriptional regulator

Position	Mutation	Annotation	Gene	Description
2,170,172	C→T	R463H (CGC→CAC)	*rsxC* ←	SoxR iron-sulfur cluster reduction factor component; putative membrane-associated NADH oxidoreductase of electron transport complex
2,199,444	C→T	G203D (GGC→GAC)	*ydgH* ←	DUF1471 family periplasmic protein
2,206,367	G→A	T51I (ACC→ATC)	*ydgD* ←	putative peptidase
2,216,797	C→T	G376R (GGG→AGG)	*ynfF* ←	S-and N-oxide reductase, A subunit, periplasmic
2,220,216	G→A	A65V (GCA→GTA)	*ynfE* ←	putative selenate reductase, periplasmic
2,225,491	C→T	T310M (ACG→ATG)	*rspB* →	putative Zn-dependent NAD(P)-binding oxidoreductase
2,246,986	C→T	T265I (ACC→ATC)	*ydfJ* →	pseudogene, MFS transporter family; interrupted by Qin prophage;Phage or Prophage Related; putative transport protein; GO_component: GO:0009274 - peptidoglycan-based cell wall; GO_component: GO:0019866 - organelle inner membrane
2,269,767	G→A	L104L (TTG→TTA)	*yneE* →	bestrophin family putative inner membrane protein
2,294,711	G→A	G65S (GGC→AGC)	*safA* →	two-component system connector membrane protein, EvgSA to PhoQP
2,336,669	G→A	R261H (CGT→CAT)	*narZ* →	nitrate reductase 2 (NRZ), alpha subunit
2,346,145	G→A	A173V (GCG→GTG)	*ydcC* ←	H repeat-associated putative transposase
2,365,108	C→T	G185D (GGT→GAT)	*ydcT* ←	putative spermidine/putrescine transporter subunit
2,397,799	G→A	G514G (GGC→GGT)	*ynbC* ←	putative esterase
2,418,453	C→T	E324E (GAG→GAA)	*paaH* ←	3-hydroxyadipyl-CoA dehydrogenase, NAD+-dependent
2,420,907	C→T	A25A (GCG→GCA)	*paaF* ←	2,3-dehydroadipyl-CoA hydratase
2,455,253	C→T	V6I (GTC→ATC)	*rzpR* ←	pseudogene, Rac prophage; Bacteriophage Rz lysis protein family;Phage or Prophage Related; putative Rac prophage endopeptidase; completely contained in another CDS
2,462,358	C→T	Q423* (CAG→TAG)	*recE* →	Rac prophage; exonuclease VIII, 5'-> 3' specific dsDNA exonuclease

Position	Mutation	Annotation	Gene	Description
2,468,535	C→T	D145N (GAT→AAT)	*dbpA* ←	ATP-dependent RNA helicase, specific for 23S rRNA
2,515,419	G→A	Y9Y (TAC→TAT)	*puuB* ←	gamma-glutamylputrescine oxidoreductase
2,542,029	G→A	F606F (TTC→TTT)	*acnA* ←	aconitate hydratase 1
2,560,214	C→T	P109L (CCG→CTG)	*trpC* →	fused indole-3-glycerolphosphate synthetase/N-(5-phosphoribosyl)anthranilate isomerase
2,567,122	G→A	P112P (CCG→CCA)	*yciB* →	IspA family inner membrane protein
2,569,179	C→T	intergenic (+71/-229)	*yciI* → / → *kch*	putative DGPF domain-containing enzyme/voltage-gated potassium channel
2,574,921	C→T	M1M (ATG→ATA) †	*oppD* ←	oligopeptide transporter subunit
2,592,988	G→A	L118L (CTG→TTG)	*narJ* ←	molybdenum-cofactor-assembly chaperone subunit (delta subunit) of nitrate reductase 1
2,605,982	C→T	intergenic (-11/+147)	*chaC* ← / ← *chaB*	cation transport regulator/cation transport regulator
2,610,098	C→T	P72P (CCG→CCA)	*kdsA* ←	3-deoxy-D-manno-octulosonate 8-phosphate synthase
2,631,525	C→T	L142L (CTG→TTG)	*treA* →	periplasmic trehalase
2,640,097	G→A	P271S (CCG→TCG)	*dadA* ←	D-amino acid dehydrogenase
2,652,321	G→A	intergenic (-443/-77)	*ycgJ* ← / → *minC*	uncharacterized protein/cell division inhibitor
2,689,682	C→T	P131S (CCG→TCG)	*phoP* →	response regulator in two-component regulatory system with PhoQ
2,719,508	C→T	P441S (CCA→TCA)	*fhuE* →	ferric-rhodotorulic acid outer membrane transporter
2,721,331	C→T	Q180Q (CAG→CAA)	*ptsG* ←	fused glucose-specific PTS enzymes: IIB component/IIC component
2,728,832	C→T	V80M (GTG→ATG)	*fabG* ←	3-oxoacyl-[acyl-carrier-protein] reductase
2,738,087	C→T	Q906* (CAA→TAA)	*rne* →	fused ribonucleaseE: endoribonuclease/RNA-binding protein/RNA degradosome binding protein
2,773,785	G→A	A45A (GCC→GCT)	*ymdB* ←	O-acetyl-ADP-ribose deacetylase; RNase III inhibitor during cold shock; putative cardiolipin synthase C regulatory subunit
2,801,506	G→A	Y103Y (TAC→TAT)	*putP* ←	proline:sodium symporter

Position	Mutation	Annotation	Gene	Description
2,824,748	C→T	T37T (ACG→ACA)	*torT* ←	periplasmic sensory protein associated with the TorRS two-component regulatory system
2,846,324	G→A	H514H (CAC→CAT)	*hyaB* ←	hydrogenase 1, large subunit
2,857,479	G→A	E82K (GAA→AAA)	*yccS* →	putative transporter, FUSC superfamily inner membrane protein, tandem domains
3,027,076	C→T	A133T (GCT→ACT)	*ybiR* ←	putative transporter
3,296,632	Δ1 bp	intergenic (+193/+57)	*ompT* → / ← *appY*	DLP12 prophage; outer membrane protease VII (outer membrane protein 3b)/global transcriptional activator; DLP12 prophage
3,668,919	+T	coding (358/372 nt)	*fabZ* ←	(3R)-hydroxymyristol acyl carrier protein dehydratase
4,008,766	G→A	P328S (CCA→TCA)	*yjhB* ←	putative transporter
4,042,942	C→T	Y17Y (TAC→TAT)	*ridA* →	enamine/imine deaminase, reaction intermediate detoxification
4,091,141	C→T	E199K (GAG→AAG)	*ulaE* ←	L-xylulose 5-phosphate 3-epimerase
4,156,547	G→A	G195D (GGT→GAT)	*cadA* →	lysine decarboxylase, acid-inducible
4,274,025	G→A	H41Y (CAT→TAT)	*psiE* ←	phosphate starvation inducible protein
4,327,181	G→A	P647L (CCG→CTG)	*rpoC* ←	RNA polymerase, beta prime subunit
4,345,823	C→T	noncoding (7/2904 nt)	*rrlB* ←	23S ribosomal RNA of rrnB operon
4,366,690	G→A	E180K (GAA→AAA)	*yijO* →	AraC family putative transcriptional activator
4,376,036	C→T	A201V (GCA→GTA)	*gldA* →	glycerol dehydrogenase, NAD+ dependent; 1,2-propanediol:NAD+ oxidoreductase
4,418,380	C→T	V453V (GTC→GTT)	*rhaB* →	rhamnulokinase
4,437,580	G→A	P249S (CCC→TCC)	*yihY* ←	BrkB family putative transporter, inner membrane protein
4,446,052	G→A	A287T (GCG→ACG)	*yihQ* →	alpha-glucosidase
4,454,983	G→A	R361C (CGT→TGT)	*typA* ←	GTP-binding protein
4,459,774	C→T	L195L (CTG→TTG)	*glnG* →	fused DNA-binding response regulator in two-component regulatory system with GlnL: response regulator/sigma54 interaction protein

Position	Mutation	Annotation	Gene	Description
4,460,481	G→T	E430D (GAG→GAT)	*glnG* →	fused DNA-binding response regulator in two-component regulatory system with GlnL: response regulator/sigma54 interaction protein
4,479,495	G→A	A123V (GCC→GTC)	*hemG* ←	protoporphyrin oxidase, flavoprotein
4,490,061	G→A	R138H (CGC→CAC)	*rfaH* →	transcription antitermination protein
4,498,965	C→T	A222V (GCT→GTT)	*ysgA* →	putative carboxymethylenebutenolidase
4,563,558	G→A	S118L (TCG→TTG)	*ilvG* ←	pseudogene, acetolactate synthase 2 large subunit, valine-insensitive; acetolactate synthase II, large subunit, cryptic, interrupted
4,606,833	C→T	A244V (GCG→GTG)	*pstB* →	phosphate transporter subunit

Appendix 6: *Python scripts used for data analysis*

Script A6.1 Python script for analysis and plotting of metabolite production data
bar_graph_maker_v0.71.py

```
import sys
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# read data file and ask for file formate, title, and labels
file_name_in = sys.argv[1]
file_format_out = 'pdf'
#file_format_out = raw_input('Figure output format: ').lower()
title = raw_input('Figure title: ')
ylab = raw_input('Y-axis label: ')
xlab = raw_input('X-axis label: ')

# read data into dataframe, set 'name' column to string in case
they are numerical
data = pd.read_csv(file_name_in, dtype = {'name':object})

# get column headers for grouping/labeling later
headers = data.columns
#labels = headers[1][::3]

# create mean/stdev dataframes, grouping by sample names which are
identical within triplicates
# but unique outside triplicates, do not sort the dataframes when
creating them, calculat the mean/stdev
# create new column in each dataframe to hold the name of the
triplicate
mean = data.groupby(headers[1], sort = False).mean()
mean[headers[1]] = mean.index
stdev = data.groupby(headers[1], sort = False).std()
stdev[headers[1]] = stdev.index

# get number of samples used for bar spacing, set bar width, adjust
figure size to number of samples
sample_num = np.arange(len(mean.index))
bar_width = 0.35
fig_width = (len(sample_num) / 2) + 4

# because all things must be Arial
plt.rcParams['font.sans-serif'] = 'Arial'

# begin to make a figure of appropriate sizes, add first set of
bars, set labels and attributes
fig, ax = plt.subplots(figsize = (fig_width, 6))
rects1 = ax.bar(sample_num, mean[headers[2]], bar_width, yerr =
stdev[headers[2]], color = 'red',
      label = headers[2], linewidth = 1.5,
```

```
      error_kw = dict(elinewidth = 1.5, capthick = 1.5, ecolor =
'black', capsize = 4))

# if there is a second dataset make bars for it
# set attributes for X-tick labels and adjust label placement if
there is 1 dataset vs 2
if len(data.columns) == 4:
      rects2 = ax.bar(sample_num + bar_width, mean[headers[3]],
bar_width, yerr = stdev[headers[3]],
            color = 'grey', label = headers[3], linewidth = 1.5,
            error_kw = dict(elinewidth = 1.5, capthick = 1.5, ecolor
= 'black', capsize = 4))
      plt.xticks(sample_num + (bar_width), mean[headers[1]],
fontsize = 16, rotation = 45)
else:
      plt.xticks(sample_num + (bar_width / 2), mean[headers[1]],
fontsize = 16, rotation = 45)

# increase fontsize and linewidth for title, axis labels, and ticks
# define tick placement, force axis minimums to set values while
maximums will autoscale to data
plt.suptitle(title, fontsize = 20)
plt.ylabel(ylab, fontsize = 20)
plt.xlabel(xlab, fontsize = 20)
plt.yticks(fontsize = 16)
plt.tick_params(axis = 'y', length = 5, width = 1.5, right = False)
plt.tick_params(axis = 'x', length = 0)
plt.gca().set_ylim(bottom = 0)
plt.gca().set_xlim(left = -0.5)

#uncomment the line below to set an arbitrary y-axis maximum
plt.gca().set_ylim(top = 2000)

# increase linewidth for axes surrounding the figure
for axis in ['top','bottom','left','right']:
  ax.spines[axis].set_linewidth(1.5)

# create legend, automatic location, no box, change rectangles to
squares and adjust spacing
leg = ax.legend(loc = 1, frameon = False, handlelength = 0.70,
      handletextpad = 0.5, labelspacing = 0.25, fontsize = 30)

# readjust legend font size independent of square size, readjust
spacing
for txt in leg.get_texts():
            txt.set_fontsize(20)
            txt.set_va('bottom')

# save figure in desired format with name from input data, tightly
crop the figure
plt.savefig((file_name_in[0:-3] + file_format_out), bbox_inches =
'tight')
```

Script A6.2 Python script for analysis and visualization of DNA shuffled sequences
DNA_shuffling_analysis_v0.5.py

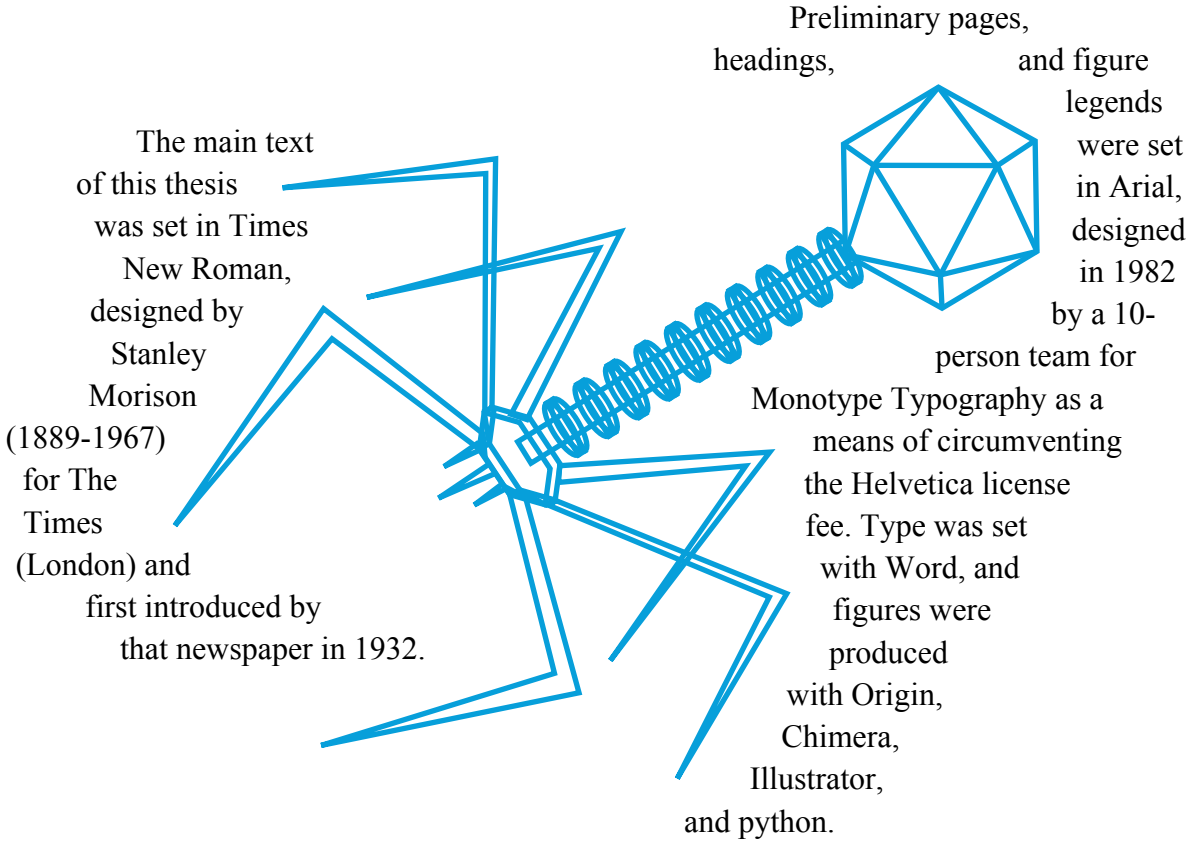
```
import subprocess as sp
import svgwrite
from operator import itemgetter
import sys
import os

# color lookup table to color code by parental sequence
color_palette = {
                        'aldh1':'rgb(245,138,94)',
                        'aldh2':'rgb(250,172,97)',
                        'aldh3':'rgb(255,239,134)',
                        'aldh4':'rgb(248,211,169)',
                        'aldh5':'rgb(177,255,103)',
                        'aldh6':'rgb(117,198,169)',
                        'aldh7':'rgb(183,230,215)',
                        'aldh8':'rgb(133,218,233)',
                        'aldh9':'rgb(132,176,220)',
                        'aldh10':'rgb(158,175,210)',
                        'aldh11':'rgb(199,176,227)',
                        'aldh12':'rgb(255,156,205)',
                        'aldh13':'rgb(214,178,149)',
                        'aldh14':'rgb(213,150,135)',
                        'aldh15':'rgb(180,171,172)',
                        'aldh16':'rgb(198,201,209)'}


def fasta_length_parser(fastafilename):
      #current_gene = ""   # Start with an empty string, just in
case
      genes = { }         # Make an empty dictionary of genes
      try:
            fh = open(fastafilename, 'r')
      except IOError:
            print 'Could not find file with filename %s' %
(fastafilename)
            result = 'Please verify that your filename is correct
and try again.'
            return result
      for lineInd, line in enumerate(fh.readlines()):
            if lineInd == 0:
                  if not line[0] == '>':
                        print 'File does not conform to FASTA
format.'
                        result = 'Please try again with FASTA
formatted file.'
                        fh.close( )
                        return result
                  else:
```

```
                  pass
            else:
                  pass
            line = line.strip()  # Clear out leading/trailing
whitespace
            line = line.upper()  # Deals with whatever case the
                                          # sequence is by making it
all upper case
            if len(line) > 0 and line[0] == ">":   # This one is a
new gene
                  seq_name = line[1:]
                  #genes[current_gene] = ""
            else:                 # Add onto the current gene
                  seq_length = len(line)
      fh.close()

      seq_info = seq_name, seq_length
      return seq_info


def shuffled_blocks_analysis(fasta_file, seq_name, seq_length):
      # some arguments for running BLAST
      # aldh.fsa database comprised of aldh1-16
      # ungapped and mismatch penalty of -15 give desired blast
alignments so far
      program = 'blastn'
      queryseq = fasta_file
      database = 'aldh.fsa'
      gap_mode = '-ungapped'
      penalty = '-15'
      outfmt = '6'

      # run the blast search as a process and capture the output
      proc = sp.Popen([program, '-query', queryseq, '-db', database,
gap_mode, '-penalty', penalty,
      '-outfmt', outfmt], stdout=sp.PIPE)
      output = proc.communicate()

      # split the blast output by newlines and remove the empty
final line
      outlist = output[0].split('\n')[:-1]

      # empty list to hold the aligned sequence blocks
      seq_blocks = []

      # read the blast output line by line, split into a list by
tabs, and capture the parental sequence,
      # length, and start/end position of each block
      for line in range(len(outlist)):
            out_line = outlist[line].split('\t')
            #seq_name = out_line[0]
            seq_blocks.append([out_line[1], int(out_line[3]),
int(out_line[6]), int(out_line[7])])
```

```
      # sort the seq_blocks by size, small to large
      seq_blocks = sorted(seq_blocks, key = itemgetter(1))

      print 'sorted by size'
      for x in range(len(seq_blocks)):
            print seq_blocks[x]
      print '\n'

      blocks_to_filter = []

      # save the start and end of a block (block A), starting with
the largest
      # then loop through the blocks a second time (block B), if
block A encompasses block B,
      # B is added to a list to be filtered
      for block in range(len(seq_blocks)):
            start = seq_blocks[block][2]
            end = seq_blocks[block][3]
            for block in range(len(seq_blocks)):
                  if start == seq_blocks[block][2] and end ==
seq_blocks[block][3]:
                        continue
                  elif start <= seq_blocks[block][2] and end >=
seq_blocks[block][3]:
                        blocks_to_filter.append(seq_blocks[block])
                  else:
                        continue

      # create a filtered blocks list by subtracting out blocks from
the filtered list
      seq_blocks_filtered = [x for x in seq_blocks if x not in
blocks_to_filter]

      # sort blocks from 5' to 3' to prepare for resolving overlaps
      seq_blocks_filtered = sorted(seq_blocks_filtered, key =
itemgetter(2))

      print 'sorted by position'
      for x in range(len(seq_blocks_filtered)):
            print seq_blocks_filtered[x]
      print '\n'

      # loop through the blocks (making sure to stop at the last
block)
      # if block A ends after the start of block B, update the
start/end of each block to be the average of the overlap
      # also update the length of the block (used for making the
figure)
      for block in range(len(seq_blocks_filtered)):
            #end = seq_blocks_filtered[block][3]
            if block < len(seq_blocks_filtered) - 1:
                  if seq_blocks_filtered[block][3] >
seq_blocks_filtered[block + 1][2]:
```

```
                        junction_position =
(seq_blocks_filtered[block][3] + seq_blocks_filtered[block + 1][2])
/ 2
                        seq_blocks_filtered[block][3] =
junction_position - 1
                        seq_blocks_filtered[block][1] =
seq_blocks_filtered[block][3] - seq_blocks_filtered[block][2] + 1
                        seq_blocks_filtered[block + 1][2] =
junction_position
                        seq_blocks_filtered[block + 1][1] =
seq_blocks_filtered[block + 1][3] - seq_blocks_filtered[block +
1][2] + 1

      # initialize the svg file with a filename and resolution
      svg_document = svgwrite.Drawing(filename = seq_name +
"_v0.5.svg",
                                                      size =
(str(seq_length + 100) + "px", "48px"))

      # add a black bar to represent the full length gene
      svg_document.add(svg_document.rect(insert = (100, 15),
                                                        size =
(str(seq_length) + "px", "16px"),
                                                        fill =
"black"))

      # draw a rectangle of the correct size and shape for each
sequence block
      offset = 0
      for block in range(len(seq_blocks_filtered)):
            svg_document.add(svg_document.rect(insert =
(seq_blocks_filtered[block][2] + 100, offset),
                                                              size
= (str(seq_blocks_filtered[block][1]) + 'px', "48px"),

      stroke_width = "1",

      stroke = "black",
                                                              fill
= color_palette[seq_blocks_filtered[block][0]]))
            #offset += 12

      svg_document.add(svg_document.text(seq_name, insert = (5,
30)))

      svg_document.save()
      print '%s done' % fasta_file


cwdfiles = os.listdir('.')

for cwdfile in cwdfiles:
      if cwdfile.endswith('.fasta'):
            #print cwdfile
```

```
            seq_info = fasta_length_parser(cwdfile)
            shuffled_blocks_analysis(cwdfile, seq_info[0],
seq_info[1])
```


The main text of this thesis was set in Times New Roman, designed by Stanley Morison (1889-1967) for The Times (London) and first introduced by that newspaper in 1932.
Preliminary pages, headings, and figure legends were set in Arial, designed in 1982 by a 10-person team for Monotype Typography as a means of circumventing the Helvetica license fee. Type was set with Word, and figures were produced with Origin, Chimera, Illustrator, and python.