Title
Harnessing evolution to study cellular regulation of metabolism using synthetic pathways for production of C4 monomers

Author
Yu, Vivian Yaci

Publication Date
2018

Peer reviewed|Thesis/dissertation

Harnessing evolution to study cellular regulation of metabolism using synthetic pathways for production of C_4 monomers

by

Vivian Yaci Yu

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, Co-Chair
Professor Jamie H. D. Cate, Co-Chair
Professor Nicolas T. Ingolia
Professor John E. Dueber

Spring 2018

Harnessing evolution to study cellular regulation of metabolism using synthetic pathways for production of C_4 monomers

by Vivian Yaci Yu

Abstract

Harnessing evolution to study cellular regulation of metabolism using synthetic pathways for production of C_4 monomers

by

Vivian Yaci Yu

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Michelle C. Y. Chang and Professor Jamie H. D. Cate, Co-Chair

The ability of living systems to carry out the tasks needed to sustain life relies on the existence of a dynamic and complex network of chemical reactions within each cell. Indeed, it is the cell's capacity for chemistry that allows it to intake simple carbon sources and transform them into the thousands of molecules needed to drive and coordinate the fundamental processes that are the hallmarks of life. Thus, cells possess an enormous synthetic potential that can be engineered for targeted chemical synthesis. By mixing-and-matching enzymes to construct synthetic metabolic pathways, the potential of natural metabolism can be harnessed to achieve multi-step synthetic routes in a single fermentation step in green conditions. As such, these approaches have expanded contributions of biological systems in new areas of the chemical, beauty, fashion, and food sectors as well as providing innovative solutions for sustainability. A major challenge in the development of cell-based chemical synthesis is the re-routing of carbon through a metabolic network that has evolved robust mechanisms to ensure coordination at the local- and system-level for the native function of cell growth and maintenance. In particular, central carbon pathways, such as glycolysis and the tricarboxylic acid cycle (TCA), form many connections with the rest of the network and are difficult to manipulate as their behavior is affected by multiple inputs and outputs and subject to strong homeostatic control.

In this work, we combine rational design and adaptive evolution to achieve a high carbon flux to synthetic pathways by coupling cell growth with product titers. We demonstrated a hybrid approach via the design of synthetic pathways in *Escherichia coli* to selectively produce three industrially-relevant C_4 monomers, 2-hydroxybutanone, 1,3-butanediol, and *n*-butanol, as bioproduct precursors to methyl vinyl ketone, 1,3-butadiene, and 1-butene. Using a genetic selection, these pathways could be evolved from theoretical yields of 7-20% to near quantitative yield. Genome sequencing of the evolved strains showed that global RNA processors, *rpoB/rpoC*, *pcnB*, and *rne*, were found mutated in the most successful daughter cells, giving rise to the hypothesis that changes in metabolism were related to transcriptional remodeling. Subsequent characterization of these mutations demonstrates that they are sufficient to capture the majority of the evolved phenotype. Further cell profiling experiments show that large-scale shifts do indeed occur in both the transcriptome and metabolome between the parent strains and evolved strains. Notably, we observed that a 25-fold increase in the central building block, acetyl coenzyme A (CoA), could be attained through adaptive evolution. Taken together, these results highlight the

possibility of synthetic pathways to be used not only for scalable chemical production but also as a platform for discovery and study of cellular function.

A similar strategy was developed for the eukaryotic host, *Saccharomyces cerevisiae*, with the goal of exploring metabolic compartmentalization and eukaryotic regulatory mechanisms. Towards this goal, a synthetic *n*-butanol pathway in yeast was constructed and optimize by a combination of promoter and terminator engineering, enzyme screening, and gene knockout to alter redox balance and cellular regulation of transcription and translation. These efforts yielded a 5-fold increase from ~120 mg L^{-1} to ~550 mg L^{-1}. In conjunction of the pyruvate dehydrogenase bypass pathway for production of cytosolic acetyl-CoA, we explored the effect of the deletion of *GCN5*, which consumes acetyl-CoA through its histone acetylase activity. Combining these approaches, we also achieved a 5-fold increase in *n*-butanol production titer (from ~100 mg L^{-1} to ~500 mg L^{-1}). Through the knockout of 7 redundant alcohol dehydrogenases for ethanol production, we have initiated the preliminary implementation of adaptive evolution in this system.

Table of Contents

Table of Contents *i*

List of Figures, Schemes, and Tables *iii*

List of Abbreviations *vii*

Acknowledgments *xi*

Chapter 1: Introduction

1.1	*Introduction*	*2*
1.2	*Acetyl-CoA: A highly-regulated and central building block*	*2*
1.3	*Studying and engineering natural systems that store acetyl-CoA*	*3*
1.4	*Bioinspired engineering of acetyl-CoA availability*	*9*
1.5	*Exploring new pathways for improving theoretical yields*	*10*
1.6	*Examining redox regeneration*	*14*
1.7	*Engineering other cellular processes*	*19*
1.8	*Conclusion and thesis organization*	*27*

Chapter 2: Evolution of cellular chemistry using synthetic pathways for production of C4 monomer

2.1	*Introduction*	*34*
2.2	*Materials and methods*	*36*
2.3	*Results and discussion*	*41*
2.4	*Conclusion*	*58*
2.5	*References*	*60*

Chapter 3: Characterizing the systems-level changes in *Escherichia coli* strains evolved for C_4 monomer production

3.1	*Introduction*	*64*
3.2	*Materials and methods*	*64*
3.3	*Results and discussion*	*70*
3.4	*Conclusion*	*89*
3.5	*References*	*91*

Chapter 4: Engineering *Saccharomyces cerevisiae* for the production of *n*-butanol

4.1	*Introduction*	*94*
4.2	*Materials and methods*	*97*
4.3	*Results and discussion*	*114*
4.4	*Conclusion*	*149*
4.5	*References*	*154*

Appendices

Appendix 1	*Complete list of constructs*	*157*
Appendix 2	*Strains, plasmids, oligonucleotides, sequences, and genome sequencing results for Chapter 2*	*163*
Appendix 3	*Strains, plasmids, oligonucleotides, and sequences, RNA-sequencing results, and metabolomics data for Chapter 3*	*194*
Appendix 4	*Strains, plasmids, oligonucleotides, sequences, and RNA-sequencing results for Chapter 4*	*230*

List of Figures, Schemes, and Tables

Chapter 1

Figure. 1.1	*Acetyl-CoA exists at the crossroad of metabolism and global cellular regulation*	*4*
Figure. 1.2	*Routes for cytosolic acetyl-CoA biosynthesis*	*5*
Figure. 1.3	*Regulation of acetyl-Co-A under high and low glucose availability*	*6*
Figure. 1.4	*Acetyl-CoA pools in Y. lipolytica*	*8*
Figure. 1.5	*Rewiring acetyl-CoA metabolism for farnesene production*	*11*
Figure. 1.6	*Biosynthesis of acetyl-CoA from oxidative glycolysis (EMP) vs. non-oxidative glycolysis (NOG)*	*12*
Figure. 1.7	*Production of BDO from lignocellulosic sugars through nonphosphorylative metabolism*	*13*
Figure. 1.8	*Self-redox balancing system in S. cerevisiae*	*15*
Figure. 1.9	*Programming redox pools*	*16*
Figure. 1.10	*Improving lipogenesis by overexpressing NADPH dependent enzymes*	*17*
Figure. 1.11	*Improving 2,3-BDO production by introducing a NAD^+ generation system*	*18*
Figure. 1.12	*Malonyl-CoA regulating hybrid promoters*	*20*
Figure. 1.13	*Quorum sensing circuit*	*21*
Figure. 1.14	*Compartmentalization of the isobutanol pathway in S. cerevisiae*	*23*
Figure. 1.15	*Controling non-genetic cell to cell variation*	*24*
Figure. 1.16	*Production of complex molecules by microbial partnership*	*25*

Chapter 2

Figure. 2.1	*Synthetic pathways for production of C_4 monomers*	*35*
Figure. 2.2	*Anaerobic fermentation pathways can operate at near quantitative yields in the absence of O_2*	*42*
Figure. 2.3	*Fermentation pathways of E. coli and gene knockouts*	*43*
Figure. 2.4	*Production of C_4 monomer precursors in engineered E. coli*	*44*
Figure. 2.5	*Introduction of a sADH to increase BDO selectivity*	*46*
Figure. 2.6	*Development of a genetic selection for n-butanol production*	*46*
Figure. 2.7	*Characterization of adaptive evolution of n-butanol strains under anaerobic conditions*	*47*
Figure. 2.8	*Characterization of adaptive evolution of BDO and HB strains under anaerobic conditions*	*49*
Figure. 2.9	*Production titers of C_4 monomers compared to parent strains with high glucose loading*	*50*
Table. 2.1	*Strains isolated from evolutions*	*51*
Table. 2.2	*Genome sequencing of evolved strains*	*53*
Figure. 2.10	*High C4 monomer producing strains were isolated from adaptive evolution*	*54*
Figure. 2.11	*Cell lysate enzyme activities of n-butanol pathway enzymes for parent and evolved strains*	*55*
Figure. 2.12	*RNA-Seq profile of evolved BDO producing strain*	*56*
Figure. 2.13	*Validating mutations that arose from evolved strain*	*57*

Chapter 3

Table. 3.1	*Strains characterized by RNA sequencing*	*71*
Figure. 3.1	*RNA-Seq profile of evolved HB and BDO producing strain*	*72*
Table. 3.2	*Strains characterized by metabolomics*	*74*
Figure. 3.2	*Metabolomics analysis between parent strains and evolved strains*	*75*
Figure. 3.3	*Physiology studies of the parent strains and evolved strains*	*80*

Figure. 3.4 *Cell growth and production profiles under different carbon sources* *81*
Figure. 3.5 *Validating mutations arose from evolved strain* *83*
Figure. 3.6 *Physiology studies of the parent strains and pcnB(R149L) and rpoC(M466L) mutants* *84*
Figure. 3.7 *Production profile with key mutants from evolved strains* *85*
Figure. 3.8 *BDO production with strains that carried mutations from evolved pathways* *87*
Figure. 3.9 *BDO production with the NNK library of rpoC M466* *88*

Chapter 4

Figure. 4.1 *n-Butanol pathway assembled from three different organisms.* *95*
Figure. 4.2 *Schematic of post-transcriptional processing of eukaryotic mRNAs* *96*
Figure. 4.3 *n-Butanol production titer and pathway enzymatic activities under different hosts* *116*
Figure. 4.4 *Protein abundance and translation efficiency under different media conditions* *117*
Figure. 4.5 *Optimization of TdTer UTR sequences* *118*
Figure. 4.6 *n-Butanol production with TdTer driven by different promoters* *119*
Figure. 4.7 *n-Butanol titer with different coding sequences of ter and adhE2* *120*
Figure. 4.8 *Increased TdTer activity correlates with increased n-butanol titer* *121*
Figure. 4.9 *Integrating optimization of promoters, terminators, and UTRs* *123*
Figure. 4.10 *Production of n-butanol with different ALDH-ADH pairs* *124*
Figure. 4.11 *Production of n-butanol with various enoyl-CoA reductase* *126*
Figure. 4.12 *Production of n-butanol production with different thiolases* *127*
Figure. 4.13 *Examining the abundance of the TdTer transcript compared to TDH3* *129*
Figure. 4.14 *RNA sequencing to compare changes in the transcriptome with and without the n-butanol pathway* *130*
Figure. 4.15 *RNA-Seq profiles of host only, host with empty vectors, and the n-butanol pathway* *131*
Figure. 4.16 *Poor correlation between protein levels and transcript levels in S. cerevisiae under different media conditions* *135*

Figure. 4.17	*5'-cap assays for transcripts*	*136*
Figure. 4.18	*Gel analysis of TdTer transcript 5'-cap assay*	*136*
Figure. 4.19	*Polysome profiles for cells expressing TdTer or the n-butanol pathway compared to an empty vector control*	*137*
Figure. 4.20	*n-Butanol production with co-expression of candidates from RNA-Seq data*	*138*
Figure. 4.21	*n-Butanol production with a reduced copy number plasmid for PhaA-Hbd-Crt*	*141*
Figure. 4.22	*SDS-PAGE of TdTer and AdhE2 protein purification*	*142*
Figure. 4.23	*Western blots for TdTer, AdhE2, and TDH3*	*143*
Figure. 4.24	*Ter activity and n-butanol production in vacuole protease knockout hosts*	*144*
Table 4.1.	*Selected knockouts for Ter expression screening*	*145*
Figure. 4.25	*Analysis of the effect of protein quality control gene knockouts on Ter expression and chaperone co-expression of n-butanol production*	*147*
Figure. 4.26	*n-Butanol production with single and double knockout hosts*	*148*
Figure. 4.27	*Approaches to improve cytosolic acetyl-CoA pool in S. cerevisiae*	*150*
Figure. 4.28	*Analysis of the effect of gcn5 deletion*	*151*
Figure. 4.29	*Cell growth and n-butanol profiles for the adaptive evolution culture with the BY4741Δ7 host*	*152*

List of Abbreviations

2KG	2-Ketoglutarate
ACC	Acetyl-CoA carboxylase
Acetyl-CoA	Acetyl coenzyme A
ACL	ATP citrate-lyase
Acp	Acetyl phosphate
ACS	Acetyl-CoA synthetase
ADA	Acetaldehyde dehydrogenase acylating
ADH	Alcohol dehydrogenase
ADP	Adenosine-5'-diphosphate
AHL	3-Oxohexanoylhomoserine lactone
AL	L-Arabinolactonase
ALDH	Aldehyde dehydrogenase
AMP	Adenosine monophosphate
AMS1	Vacuolar α-mannosidase
ANB1	Translation elongation factor eIF-5A
AP	Antarctic phosphatase
APE4	Cytoplasmic aspartyl aminopeptidase
ATG9	Autophagy transmembrane protein
ATP	Adenosine-5'-triphosphate
BDO	1,3-Butanediol
Cat2	Peroxisomal/mitochondrial carnitine acyltransferase
Cb	Carbenicillin
CCW12	Cell wall mannoprotein
CDC19	Pyruvate kinase
Cm	Choloramphenicol
COG	Cluster orthologous group categories
DBP2	ATP-dependent RNA helicase of the DEAD-box protein family
DCW	Dry cell weight
DHAP	Dihydroxyacetone phospahte
dNTPs	Deoxynucleotides
DTT	Dithiothreitol
dUTP	Deoxyuridine triphosphate
E4P	Erythrose-4-phosphate
EMP	Embden–Meyerhof–Parnas
ENO1	Enolase I
F6P	Fructose-6-phosphate

FAD	Flavin adenine dinucleotide
FAS	Fatty acid synthase
FBA1	Fructose 1,6-bisphosphate aldolase
G3P	Glycerol-3-phosphate
GAL	Galactokinase
GC-FID	Gas Chromatography - Flame Ionization Detection
GC-MS	Gas Chromatography - Mass Spectrometry
GCN5	Histone acetyltransferase
GCPR	G-coupled protein receptors
GD	D-galactarate dehydratase
GO	Gene ontology
GPD	Glyceraldehyde-3-phosphate
GPM1	Tetrameric phosphoglycerate mutase
GRAS	Generally Regarded As Safe
HB	4-Hydroxy-2-butanone
HMGR	3-Hydroxy-3-methylglutaryl-CoA reductase
HPLC	High Performance Liquid Chromatography
HSP	Heat shock protein
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
KdaD	L-arabonate dehydratase
KDC	Keto acid decarboxylase
KdxD	2-Keto-3-deoxy-D-xylonate dehydratase
KGSADH	2-Ketoglutarate semialdehyde dehydrogenase
Km	Kanamycin
LB	Luria Broth (Miller)
LHS1	Chaperone of the endoplasmic reticulum lumen
ME	Malic enzyme
NAD^+	β-Nicotinamide adenine dinucleotide
NADH	β-Nicotinamide adenine dinucleotide (reduced)
$NADP^+$	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced)
NAT	Streptothricin sulfate
NOG	Non-oxidative glycolysis
NoxE	NADH oxidase
OD	Optical density
PBR1	Putative oxidoreductase
PcnB	Poly(A) polymerase
PDC	Pyruvate decarboxylase
PDHc	Pyruvate dehydrogenase complex
PEP4	Vacuolar aspartyl protease (proteinase A)

PFK1	Phosphofructokinase (α-subunit)
PFL	Pyruvate formate lyase
PGI1	Glycolytic enzyme phosphoglucose isomerase
PHA	Poly(hydroxyl)alkanoate
PHB	Poly(hydroxyl)butyrate
PK	Phosphoketolase
PMSF	Phenylmethanesulfonyl fluoride
Pnp	Polyribonucleotide nucleotidyltransferase
PntA/B	NAD(P) transhydrogenase α/β subunits
PopQC	*In vivo* population quality control
POS5	NADH kinase
PPP	Pentose phosphate pathway
Pta	Phosphotransacetylase
PYK2	Pyruvate kinase
RKR1	RING domain E3 ubiquitin ligase
RLI1	Essential Fe-S protein
Rne	Ribonuclease E
RPN4	Regulatory particle non-ATPase
RpoB	RNA polymerase ß subunit
RpoC	RNA polymerase ß' subunit
RPS14B	Protein component of the small (40S) ribosomal subunit
RT-PCR	Real Time-Polymerase Chain Reaction
SAN1	Ubiquitin-protein ligase
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SLX8	Subunit of Slx5-Slx8 SUMO-targeted ubiquitin ligase (STUbL) complex
Sp	Spectinomycin
SSA1	Stress-Seventy subfamily A
SSB1	Stress-Seventy subfamily B
SSM4	Membrane-embedded ubiquitin-protein ligase
STE3	Receptor for a factor pheromone
T4PNK	T4 Polynucleotide linase
TB	Terrific Broth
Tc	Tetracycline
TCA	Tricarboxylic acid
TDH3	Glyceraldehyde-3-phosphate dehydrogenase
TMA10	Protein of unknown function that associates with ribosomes
TPI1	Triose phosphate isomerase
Tris	Tris(hydroxymethyl)aminomethane
UDH	Uronate dehydrogenase
UMP	Uridine monophosphate

UMP1	Chaperone required for correct maturation of the 20S proteasome
UTP	Uridine-5'-triphosphate
UTR2	Chitin transglycosylase
UTs	Untranslated egions
VSV	Vesicular stomatitis virus
XD	D-xylonate dehydratase
XDH	D-xylose dehydrogenase
XL	D-xylonolactonase
XRN1	5'-3' Exoribonuclease 1
YDJ1	Type I HSP40 co-chaperone
YHL001W	Ribosomal 60S subunit protein L14B
YLR075W	Ribosomal 60S subunit protein L10

Acknowledgments

First and foremost, I would like to thank my advisors, Professor Michelle Chang and Professor Jamie Cate. I am extremely grateful for the opportunities and tremendous support Michelle has given me throughout the graduate career. Michelle, thank you for your scientific mentorship, and always being super patient and encouraging when science doesn't go as planned or when I tried to learn a new technique. Thank you for assembling an incredible team and setting a great working culture one could ask for. Beyond the scientific trainings, I am thankful for all the time and effort Michelle has devoted to help me developing my professional skills, including writing and presentation skills. I will strive to make figures that are "Michelle standard"! I am grateful for the scientific guidance and support that Jamie has given me. Thank you for all the insightful and encouraging scientific discussion throughout these years, Jamie! Your passion about science is contagious! It has been a great learning experience watching Jamie leading the team and getting exciting science accomplished. Thank you for giving me opportunities to work with and learn from a group of incredible post-doctoral scholars.

Next, I would like to thank Dr. Matthew Davis for believing in me to continue his Ph.D. legacy project. Thank you for believing in me and dealing with my gChat spams, Matt! It has been a great pleasure and an honor that I got to work with Matt. He is truly a great teacher and an incredible lab citizen with great sense of humor! I still remember vividly about my first ever Unix lesson from Matt. Matt, thank you for all the civic engagement services you have done for us.

I would also like to thank members of my thesis committee, Professor Nicolas Ingolia, Professor John Dueber, and Professor Danielle Ercek. They all have been providing valuable scientific insights. Thank you for all the fun and fruitful discussion during thesis committee meetings! Nothing is more encouraging than the continuing discussion via emails after a long thesis committee meeting.

I have been extremely lucky that I got to work with an incredible team of scientists from both the Chang Lab and the Cate Lab. They have a lot to do with both my personal and professional development during graduate school. I owe a special debt to Dr. Michael Blaisse. Mike is one of the brightest people I ever met. I can always count on Mike for a clear explanation for any concepts that I am clueless about. I am grateful for all the valuable scientific advices I have gotten from Mike, whether that was from our shared old office space or at dinner tables in random restaurants. From a rotation mentor, to a colleague, and a friend, Mike, has been the best that anyone could ask for. Mike, thank you for adapting me to your friend zone, and always there to hang out, listen to my concerns, and offer great advice. Thank you for involving me on every Thanksgiving dinner, birthday celebration, and other special holidays, Mike! This thesis won't be written without the tremendous support from Mike. Mike, thanks for keeping me grounded through this entire thesis writing process! I also owe a special debt to Dr. Joseph Gallagher and Dr. Jonathan McMurry. Joe has literally been like an older brother to me. I am extremely fortunate that I got to work with Joe, and became one of his many friends. Joe, thank you for all those time that you have spent listening to me, calming down, and offering me valuable advice, on both work-related matters and personal concerns. Thank you for making late nights in lab fun and going on all sort of life adventures, Joe, from food exploration in Chinatown, attending dancing and yoga classes at RSF, to the most touristy adventures (biking across the Golden Gate Bridge?...)! It has been an honor that I get to claim Jon as a friend. He is just like the Wikipedia. Regardless what scientific questions I have,

Jon always has an answer. Jon, thanks for offering your wisdom, encouragement, and dealing with all my ignorant questions. I am also extremely grateful for this incredible friendship. Jon, thank you for being a very welcoming friend and opening the door for me to share my unusual family background. Thank you for sharing your professional knowledge on bikes, photography, and many more. Jon, I will forever admire the curiosity you have on literally everything! I am also extremely thankful for the friendship with my MCB classmate and colleague, Jase Gehring. We have taught together, shared a bench, and gone through quite a bit of debate on many different issues. Jase, I will never forget the "plate wall" with the giant "No Vivian" sign that you built between our benches. That's a perfect representation for our friendship. Graduate school has been a touch journey, but I have enjoyed every conversation we have had, both high and low. Jase, you have encouraged me to be more open minded person. Of course, I will never forget all the exciting science discussion we had! Jase's drawings on the white board made life in the Tan basement much better; no windows, no problem, illustrating on a giant iMac is just fun!

I am also very fortunate that I had worked with other incredible scientists from the Chang Lab, Dr. Miao Wen, Dr. Michiei Sho, Dr. Jeffrey Hanson, Dr. Mark Walker, Dr. Brooks Bond-Watts, Dr. Ben Thuronyi, Dr. Stephanie Jones, Dr. Maggie Brown, Dr. Amy Weeks, and Dr. Omer Ad. Miao always has brilliant suggestions. I am extremely grateful that I got to know Miao. I have enjoyed every lunch and Yogaland dates that we have had together! Dr. Michiei Sho helped me starting the yeast project in the Chang Lab. Dr. Mark Walker taught me Python and wrote many scripts for me. Mark, thank you for all you do to make everyone's life better in the lab. I have also gotten a lot of science advice from Dr. Jeffrey Hanson. I also enjoyed our late-night conversations in the Chang Lab, Jeff. To the later crew of the Chang Lab, thank you for being so awesome! Jorge Marchand, thank you for always being the leader and creating the enthusiastic working environment! Monica Neugebauer, thanks for being an awesome co-worker and friend! Thank you for dealing with all my urgent QQQ requests and listening to me both in- and out- of lab, Monica! Sasilada Sirirungruang, thank you for being a great friend! I have truly enjoyed our hikes together. Kersh Theva, thank you for the fun conversations on science and education! Jason Fang, thank you for always doing an excellent job on lab duties! I also enjoyed all our random conversations, from science to random memes from Facebook! I am also glad to get to meet our newest Chang Lab member, Ed Koleski. I am also very fortunate that I got to work with Dr. Chia-I Lin, Dr. Ningkun Wang, and Dr. Hongjun Dong. Chia-I, thank you for your all your scientific and career development advice! It has been a joy watching you draw those incredible complex molecules on ChemDraw. Thanks for being my coffee buddy! Ningkun, thank you for being an awesome co-worker and friend. I am glad that we were once CSP buddies, and thank you for all your valuable advice! Hongjun, thank you for all the great discussion we have had on science, education, future career, and many more! I am also very grateful that I got to work with you on the same project. I would also like to thank my collaborator, Charles Berdan from the Nomura lab and Professor Dan Nomura for their help on the metabolomics experiments. Finally, I am grateful that I got to work with a group of excellent undergraduates, Beverly Fu, Malcolm Smith, and Sean Yuan. Beverly, thanks for being the most awesome undergraduate one could ask for! It has been fun working with you and getting to know you, Beverly. I also have enjoyed hanging out with other undergraduates in the lab, Nicolas Lue, Joel Sax, Madeleine Ing, Kiera Sumida, and Milia Moore. Thanks for all the fun conversations, Boba and ice cream adventures guys!

As a joint student, I am also fortunate enough to get to learn from another group of fantastic scientists from the Cate Lab. First, I would like to thank my rotation mentor, Dr. Xin Li. Xin has been an excellent mentor. He is super knowledgeable and kind. It has been extremely rewarding learning and working with Xin! I am also very grateful that I got to work with Dr. Yuping Lin. Yuping, thank you for all your scientific help and the unconditional support a friend could ask for! Thank for being an older sister to me while I am away from home! I also have gotten a lot of help from Dr. Amy Lee, Dr. Ligia Acosta Sampson, Dr. Owen Ryan, Dr. Duane Smith, Dr. Nathanael Lintner, Dr. Kulika Chomvong, and Dr. Arto Pulk. Toward the end of my graduate journey, I am also get to know and work with the newer members of the Cate Lab, Mia Pulos, Fred Ward, Luisa Arake de Tacca, Annsea Jungmin Park, Angelica Gonzalez-Sanchez, Raissa Estrela Curado, Snigdha Poddar, Dr. Nadège Liaud, Dr. Wenfei Li, Dr. Audre Lapinaite, Dr. Christine He, Dr. Chris Brown, and Dr. Das De Silva. Thank you for being great coworkers! Finally, I would like to thank the lab manager, Bruno Martinez. Bruno, thank you for all your help and always creating a very welcoming environment!

I also owe a special debt to those whom have gone above and beyond to take care all the administration affairs. Thank you everyone from the MCB GAO office! I would like to give a special thanks to Eric Buhlis and Berta Parra. Lissette Garcia also has done a tremendous amount of work to make my life much easier. I also want to thank the Chemistry Receiving Department team, Carl Lamey and Roy Washington, for always making sure I get my primers on time!

Finally, I would like to thank those who have helped me make the graduate school journey possible. I am extremely grateful for the mentorship and supports from my undergraduate research advisor, Professor Robert Doyle and my mentor, Dr. David Dixson. I also want to give a special thanks to my incredible teachers and mentors, Ms. Rae-Ann Fischer, Mrs. Donna Turturro, Mr. Irving Reich, Dr. Stephen Wright, Mr. Henry Jankiewicz, and Ms. Christabel Sheldon. I would never make it to graduate school without their help and support. Lastly, I would like to thank all my friends and family for their continuous love and support. I am grateful for the incredible friends who keep me grounded. Thank you to my mom, sister, brother, and sister-in-law for supporting me through this journey.

Chapter 1. *Introduction*

1.1. Introduction

The ability to control molecular structure has transformed society in diverse areas, from art and agriculture to medicine and new electronic devices. Like traditional synthetic chemistry, the biochemistry of living organisms also offers thousands of chemical reactions that can be manipulated to produce molecule targets of interest. By taking advantage of genetic methods to mix-and-match enzyme catalysts to build synthetic metabolic pathways, living organisms can serve as the host for scalable processes to produce a broad range of small molecules in a single-stage reactor under green conditions. Indeed, synthetic biology systems have made an impact on non-conventional biological industries such as chemicals, beauty, fashion, and food sectors, to provide new approaches and solutions in renewability and sustainability. However, it remains challenging to rewire the metabolism of the cell for this purpose, given the sophisticated regulation that has evolved to coordinate the large number of metabolic pathways required to support the cell. In this section, efforts in controlling and rewiring metabolism to increase yields and productivity of synthetic metabolic pathways are reviewed.

1.2. Acetyl-CoA: A highly-regulated and central building block

Acetyl coenzyme A (CoA) is a key building block for the production of a variety of target compounds, including commodity chemicals, such as short- to long-chain hydrocarbons, and fine chemicals, such as polyketides, isoprenoids, flavonoids, and some alkaloids (*Figure 1.1*). It exists at the intersection of catabolic and anabolic pathways, serving as a central node for glycolysis, the tricarboxylic acid (TCA) cycle, and fatty acid synthesis (*Figure. 1.1*). Cellular acetyl-CoA levels are therefore subject to many layers of regulation to ensure both robust homeostasis as well as a sensitive dynamic response to the environment. Beyond its role as a metabolic building block, acetyl-CoA is also the acyl group donor for protein acetylation, controlling both transcriptional and post-transcriptional regulation. Acetylation is a ubiquitous protein modification in both prokaryotes and eukaryotes and alters protein–protein interactions, protein localization and stability, and transcriptional and enzymatic activities [*1*]. Acetylated proteins are involved in almost all cellular processes, including cell cycle, RNA metabolism, redox state, and metabolism. In particular, acetylation has been found to be especially important in controlling metabolic flux through primary metabolic pathways such as glycolysis, gluconeogenesis, TCA cycle, and the pentose phosphate pathway [*1*, *2*] As such, advancing our understanding of factors regulate the partitioning of acetyl-CoA pool between different metabolic outcomes or organelles is key to engineering high-flux acetyl-CoA dependent biosynthetic pathways [*2*].

Given the central position of acetyl-CoA in metabolism, it is not surprising that there are several pathways for its production (*Figure 1.1*). Under aerobic conditions, where flux to acetyl-CoA is highest due to high rates of cell growth, acetyl-CoA is mainly produced from pyruvate by the pyruvate dehydrogenase complex (PDHc) in both prokaryotes and eukaryotes. However, in prokaryotes the PDHc is localized to the cytosol whereas it is found in the mitochondrial matrix in eukaryotes [*3*]. Thus, cytosolic processes that use acetyl-CoA, such as fatty acid biosynthesis rely either on the transport of acetyl-CoA from the mitochondria or the use of the alternative PDHc bypass pathway. In the PDHc bypass, pyruvate is decarboxylated to acetaldehyde, followed by oxidation of acetaldehyde to acetate and ligation of CoA to produce acetyl-CoA. Obligate

anaerobes and other prokaryotes utilize pyruvate formate lyase (PFL) to convert pyruvate is to acetyl-CoA and formate by a radical-dependent mechanism [*4 , 5*] (*Figure. 1.2*).

In addition to multiple pathways for its biosynthesis, acetyl-CoA also has dual roles as a building block as well as a regulator. These two roles in metabolism and regulation are tightly coupled through protein acetylation, which has been found to regulate central carbon flux in both prokaryotes and eukaryotes through the action of acetyl transferases and deacetylases whose activities are also altered by the availability of their co-substrates [*6*]. Under high carbon availability, acetyl-CoA is abundant in the cell and hence protein acetylation is high [*7, 8*]. Overall, greater protein acetylation results in higher metabolic flux via direct regulation of protein activity. In one case, biochemical studies have shown that acetylation of malate dehydrogenase, which converts malate to oxaloacetate in the TCA cycle, greatly increases its enzymatic activity (50%) [*9*]. The global protein acetylation state is further affected by regulation by the cellular redox state through the NADH pool. Protein deacetylases belong to NAD^+-dependent sirtuin family and their activity is low when the NADH:$NADH^+$ ratio is high. As a result, the high acetyl-CoA and NADH levels found under conditions of high glycolytic flux act synergistically to increase protein acetylation and decrease protein deacetylation such that global acetylation is amplified.

In eukaryotes, histone modification serves as another major mechanism for metabolic flux control. Under conditions of carbohydrate abundance, both cytosolic and nucleocytosolic acetyl-CoA concentration are high. The high cytosolic acetyl-CoA is targeted for energy storage in the form of fatty acids (*Figure. 1.3*). This cytosolic chemistry is controlled by the first committed and rate-limiting step in *de novo* fatty acid synthesis, which is the carboxylation of acetyl-CoA to produce the malonyl-CoA extender unit by the acetyl-CoA carboxylase (ACC). Expression of ACC is activated under high nucleocytosolic acetyl-CoA, along with genes involved ribosome biogenesis and cell growth, which results in increased cell growth. When carbon is depleted, cell viability is prioritized over cell growth with low cytosolic and nucleocytosolic acetyl-CoA levels. Carbon flux is directed to mitochondria as pyruvate, which is then converted to acetyl-CoA by the pyruvate dehydrogenase (PDH) complex so that it can enter the TCA cycle for ATP production[*10*]. In conjunction with shifting acetyl-CoA pool from fatty acid synthesis pathway to the TCA cycle, the low nucleocytosolic acetyl-CoA level leads to low global acetylation that induces the expression of autophagy genes repressed by acetyl-CoA, such as *ATG7* [*10*]. Nucleocytosolic acetyl-CoA is synthesized by acetyl-CoA synthase. Acetylation on the acetyl-CoA synthase inhibits its activity, which serves as a negative feedback mechanism in response to high acetyl-CoA pool.

1.3. Studying and engineering natural systems that store acetyl-CoA

The role of acetyl-CoA is complex and so are the factors that control its availability for downstream biosynthetic pathways. One approach to elucidating strategies for engineering high-flux pathways come from the study of native hosts that store acetyl-CoA equivalents in the form of polymers, such as poly(hydroxyl)alkanoates (PHAs) [*11, 12*] or lipids [*13, 14*]. These hosts include bacteria, fungi, and algae [*15, 16*]. Since yeasts are preferred hosts for industrial processes and also have to solve the challenge of acetyl-CoA compartmentalization, much effort has been focused on

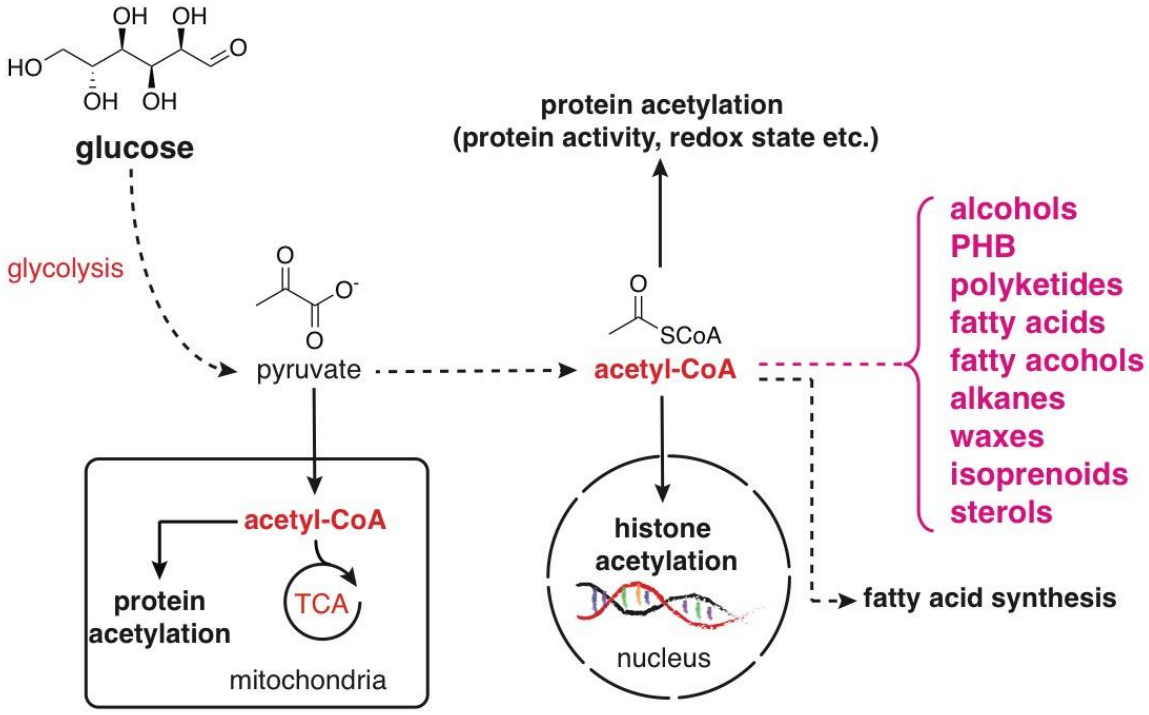


Figure 1.1. Acetyl-CoA exists at the crossroad of metabolism and global cellular regulation. The acetyl-CoA pool partitions between the cytosol and other organelles. Acetyl-CoA is the precursor for both the fatty acid synthesis (cytosol) and TCA cycle (mitochondria). In addition to its role as a metabolic building block, acetyl-CoA is the donor for protein acetylation, which takes place in the cytosol, mitochondria, and the nucleus. Compound families in pink represent reported bioproducts of acetyl-CoA. Dotted lines represent multiple steps. CoA, coenzyme A; TCA cycle, tricarboxylic acid cycle; PHB, poly(hydroxyl)butyrate.

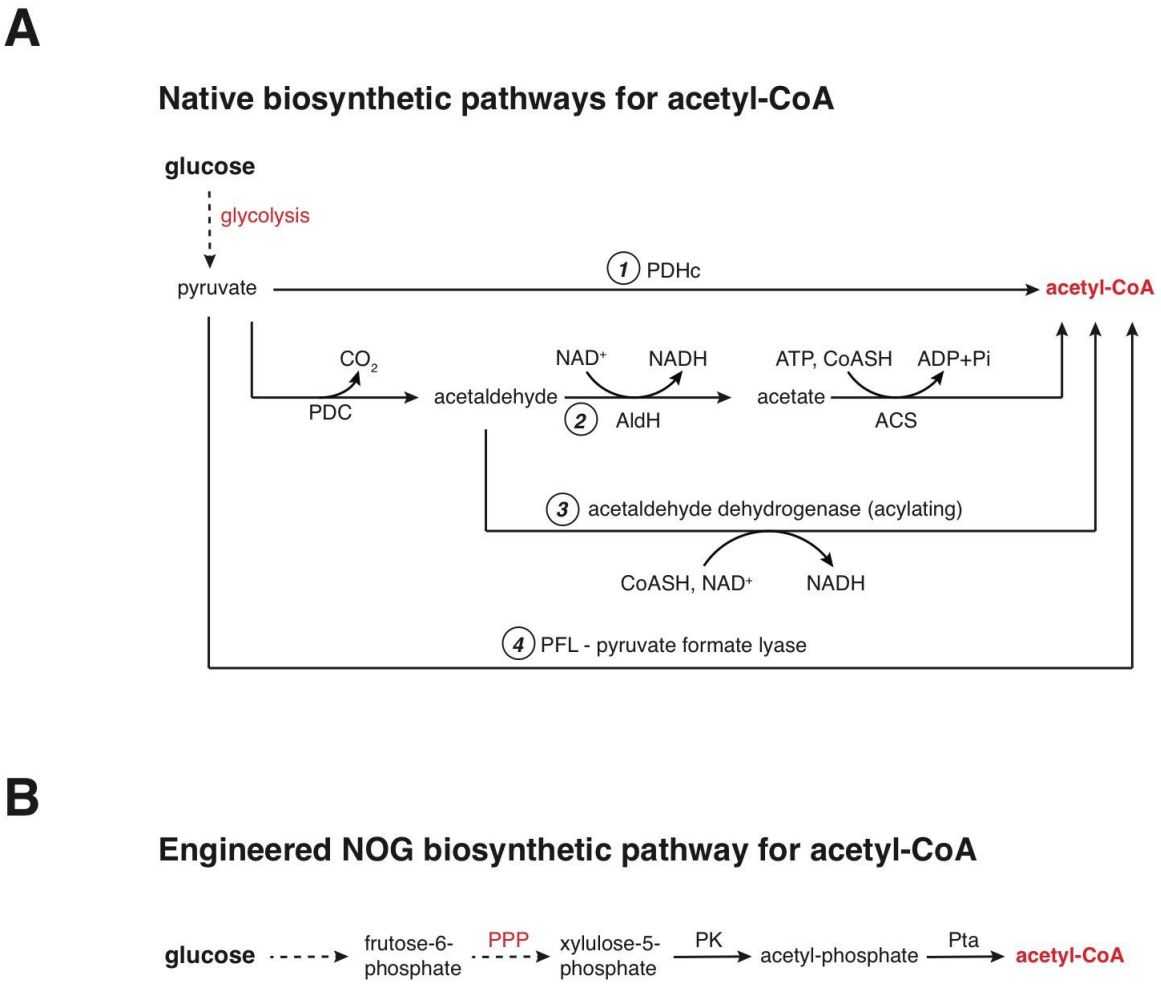


Figure 1.2. Routes for cytosolic acetyl-CoA biosynthesis. (A) 1. Acetyl-CoA can be made from pyruvate by the PDHc, which is a three-subunit complex and requires four cofactors (thiamine pyrophosphate, lipoic acid, FAD and NAD^+). 2. Pyruvate can be decarboxylated to acetaldehyde by PDC, which is subsequently oxidized to acetate. Finally, acetate is activated to produce acetyl-CoA by ACS using ATP. 3. Acetyl-CoA can also be produced directly from acetaldehyde and CoA by an acylating acetaldehyde dehydrogenase. 4. Under anaerobic conditions, acetyl-CoA can be produced from pyruvate directly by PFL via a radical-dependent mechanism. PDHc: pyruvate dehydrogenase complex; PDC: pyruvate decarboxylase; AldH: aldehyde dehydrogenase; ACS: acetyl-CoA synthase; PFL: pyruvate formate lyase; b) Acetyl-CoA can also be produced by activating acetyl-phosphate by Pta. Acetyl-phosphate come from the intermediate from the PPP, xylulose-5-phosphate, which is catalyzed by the PK. PPP: pentose phosphate pathway; Pta: phosphotransacetylase; PK: phosphoketolase. NOG: non-oxidative glycolysis. Dotted lines represent multiple steps.

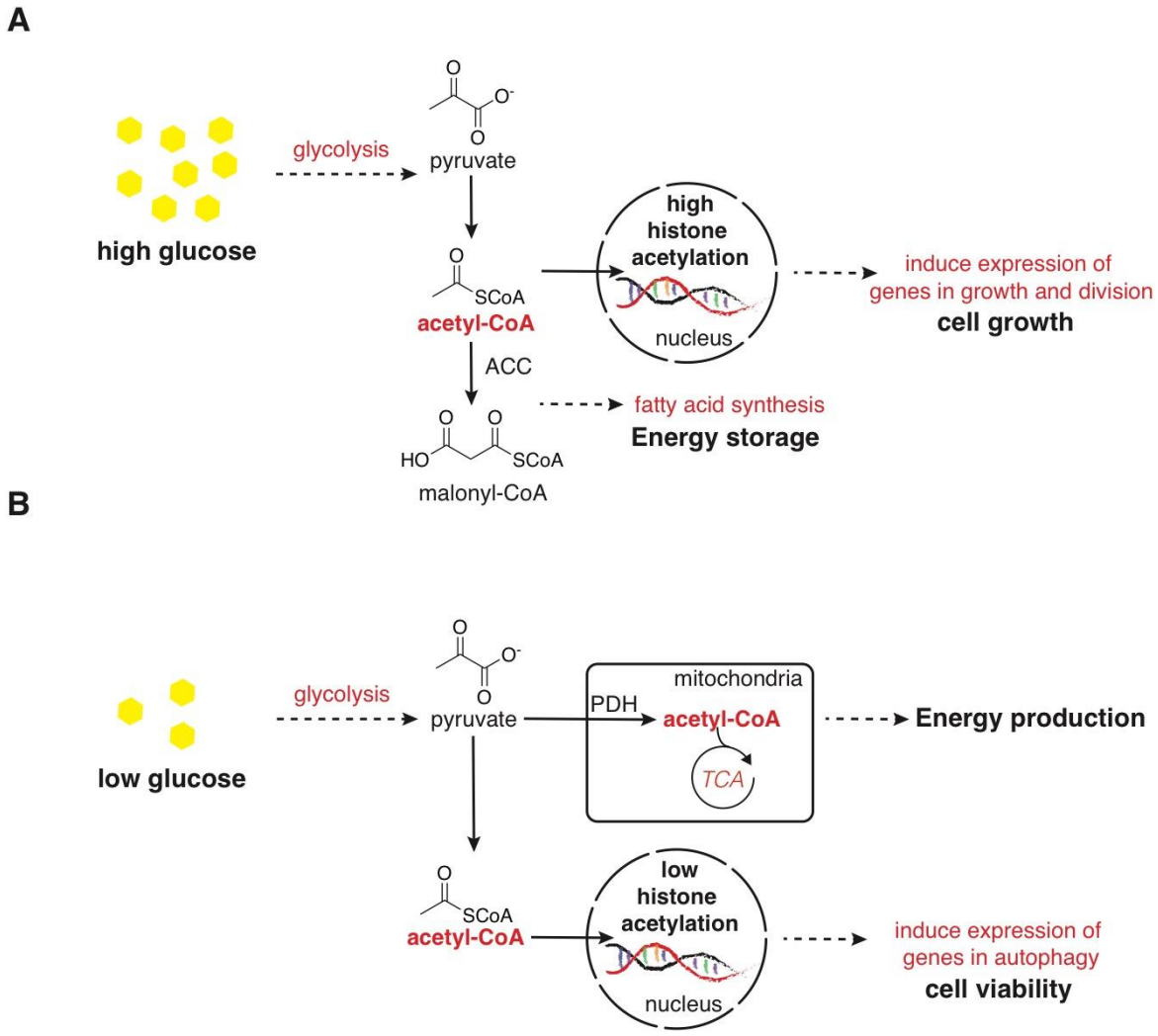


Figure 1.3. Regulation of acetyl-Co-A under high and low glucose availability. When carbon availability is high, both cytosolic and nucleocytosolic acetyl-CoA concentrations are high. Carbon in the form of acetyl-CoA is converted to malonyl-CoA and stored in the form of fatty acids. High nucleocytosolic acetyl-CoA concentration results high global histone acetylation in the nucleus, which induces the expression of genes involved in cell growth. Under low carbon availability, cells direct carbon in the form of pyruvate to the mitochondria, which is subsequently converted to acetyl-CoA by the pyruvate dehydrogenase complex (PDHc) and enters the TCA cycle for energy production. Low cytosolic acetyl-CoA leads to low nucleocytosolic acetyl-CoA and low global histone acetylation, which induce expression of genes in autophagy. Dotted lines represent multiple steps. PDHc, pyruvate dehydrogenase complex, TCA, tricarboxylic acid; ACC: acetyl-CoA carboxylase.

understanding lipid accumulation in oleaginous yeasts in the hope that it can translate to improved titers or design principles for genetic engineering of non-oleaginous yeasts. In this arena, *Yarrowia lipolytica* has served as a major model system for study as it naturally produces up to 40% lipid by dry cell weight (DCW) from a wide range of carbon sources. From comparative genomic studies of oleaginous and non-oleaginous yeasts, it has been found that the ATP citrate lyase (ACL), mitochondrial β-oxidation pathways, as well as leucine- and lysine-metabolism may all contribute to supporting acetyl-CoA availability [*17*]. Indeed, it has been shown that in *Y. lipolytica* and other oleaginous yeasts that acetyl-CoA used for lipid synthesis is derived mainly from the transport of mitochondrial citrate from the TCA cycle to the cytosol, which is split by ACL to form acetyl-CoA and oxaloacetate (*Figure 1.4*). The other challenge is producing sufficient NADPH to power lipid synthesis. In many oleaginous yeasts, the reducing power is provide by a cytosolic malic enzyme (ME), which operates in a transhydrogenase cycle with PDC to achieve the overall conversion of NADH to NADPH with the input of ATP [*18*]. However, *Y. lipolytica* does not possess a cytosolic $NADP^+$ to NAPDH conversion, which appears to come from the pentose phosphate pathway (PPP) instead [*19*, *20*]. As the first committed step of lipid synthesis is the carboxylation of acetyl-CoA to form the malonyl-CoA extender unit, the acetyl-CoA carboxylase (ACC) from oleaginous organisms may also have different regulatory properties compared to conventional organisms [*21*]. These ACCs have been overexpressed in other organisms, giving rise to an increase in lipid content, but it is unclear whether this increase can be attributed solely to their biochemical properties [*22*].

While natural accumulation of lipid by oleaginous yeasts like *Y. lipolytica* enable the design and optimization of fermentation processes to upgrade glucose and other carbon sources to lipids, the native process has several drawbacks related to the tight control typically exerted on fatty acid biosynthesis. Fatty acid biosynthesis is highly resource intensive, so it is not surprising that optimization of lipid accumulation requires several factors. It has long been known that fatty acid biosynthesis can be amplified by nitrogen restriction [*23*], and this is no different in *Y. lipolytica*. However, nitrogen is an essential element for cell growth and these fermentations therefore yield low growth rates and require a prolonged cultivation time [*24*]. Therefore, many engineering efforts have focused on enhancing the natural productivity of hosts like *Y. lipolytica*. Metabolic engineering studies of *Y. lipolytica* have shown large gains in lipid productivity [*25*]. For example, overexpression of ACL or ME have led to gains of up to ~25% and 9% lipid accumulation respectively, measured by oil content [*26*, *27*]. In addition, the five different routes for cytosolic acetyl-CoA generation were tested and compared directly, including the PDHc acetate bypass, PDHc acetaldehyde bypass, PFL, and a non-oxidative PPP pathways (*Figure 1.4*) [*26*]. The authors also tested the standard eukaryotic acetyl-CoA transport pathway, which is the carnitine acyltransferase (Cat2) to shuttle them from both the mitochondria and peroxisome to the cytosol (*Figure. 1.4*). Of these, overexpression of the carnitine shuttling pathway achieved the best productivity of lipogenesis. Overexpressing Cat2 under optimized carbon:nitrogen ratios, a 3-fold improvement (0.565 g L^{-1} H^{-1}) on lipid productivity was achieved compared to the unengineered strain [*26*]. As a non-model organism, metabolic engineering of *Y. lipolytica* can be challenging [*28*], so methods to improve heterologous protein expression and pathway methods [*29*] as well as the development of advanced genome editing tools [30] should support advances in this area.

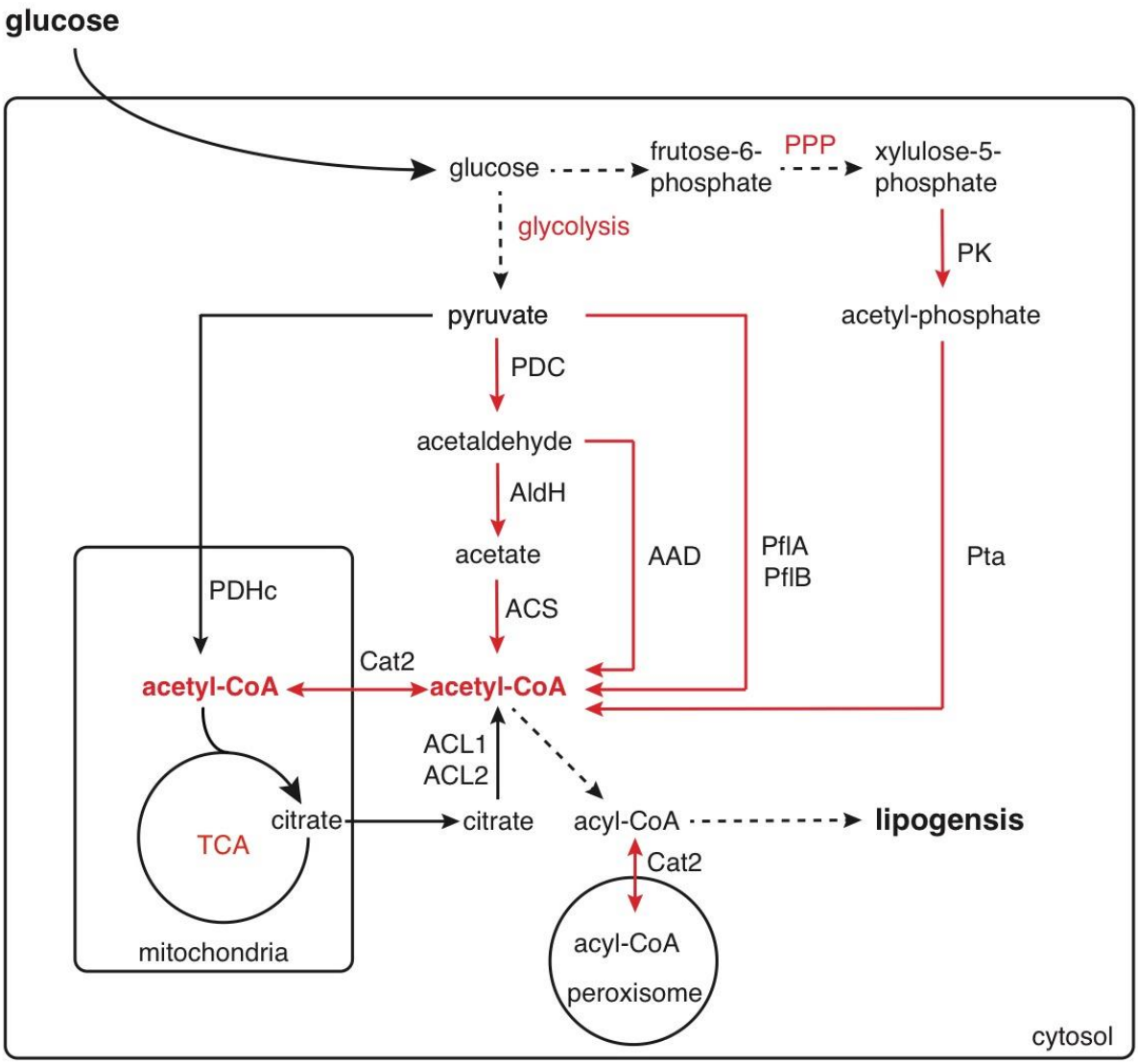


Figure 1.4. Acetyl-CoA pools in *Y. lipolytica*. Acetyl-CoA is predominantly produced from citrate by the ATP citrate lyase (ACL) under nitrogen-limited conditions in the oleaginous yeast, *Y. lipolytica*. Pathways for cytosolic acetyl-CoA production were expressed in *Y. lipolytica* to improve lipid yield. PDHc acetate bypass pathway (PDC, AldH, ACS), PDHc acetaldehyde bypass pathway (PDC, AldH, ACS), PFL, non-oxidative PPP pathway (PK, Pta), carnitine shuttle (Cat2). Dotted lines represent multiple steps. Red lines represent co-expressed acetyl-CoA pathways. ACL, ATP citrate lyase; PDHc, pyruvate dehydrogenase complex; PDC, pyruvate decarboxylase; AldH acetaldehyde dehydrogenase; ACS, acetyl-CoA synthase; AAD, acylating acetaldehyde dehydrogenase; Pfl, pyruvate formate lyase; PK, phosphoketolase; Pta, phosphotransacetylase; Cat2, peroxisomal/mitochondrial carnitine acyltransferase.

1.4. Bioinspired engineering of acetyl-CoA availability

Another approach is to use the design of native systems that naturally generate acetyl-CoA flux as a template for rewiring model and industrial hosts such as *Saccharomyces cerevisiae* (Baker's Yeast). *S. cerevisiae* is an important industrial host as well as a genetic model organism that has extensively characterized genetics and metabolism. In addition, many genetic tools and advanced genome editing technologies have already been developed, allowing rapid modification of *S. cerevisiae* at the DNA level. Therefore, *S. cerevisiae* has been a longstanding target for metabolic engineering studies to produce a broad range of molecules from natural products to biofuels [*31*] [*32, 33*]. One major challenge in this area is the relatively low availability of carbon flux for biosynthetic pathways in *S. cerevisiae*. Since it has long been selected as an ethanol-producing host, *S. cerevisiae* relies mainly on anaerobic fermentation for ATP generation, resulting in efficient and quantitative conversion of glucose to ethanol [*34, 35*]. As a result, glucose is converted to pyruvate through glycolysis and is directly converted to ethanol, with only a low level of flux to acetyl-CoA for cellular maintenance. Therefore, studies in this area have focused on developing pathways to route flux to cytosolic acetyl-CoA that is available for biosynthesis.

One approach involves the expression of cytosolic pathways to produce acetyl-CoA in *S. cerevisiae*, such as optimization of the existing PDHc acetate bypass pathway or expression of a cytosolic PDHc. The rate-limiting step for this pathway is the activation of acetate to acetyl-CoA catalyzed by the ACS. Engineering the AcsL641P mutant from *Salmonella enterica*, which resulted in an increased yield from two acetyl-CoA dependent pathways, for isoprenoid (amorphadiene : from 0.356 ± 0.001 to 0.435 ± 0.009 mM; mevalonate: from 1.78 ± 0.006 to 2.52 ± 0.017 mM) [*36*] and *n*-butanol (reached 20 mg L^{-1}) [*3*], respectively. The native mitochondrial PDHc can also be re-localized upon deletion of its signal sequence to the cytosol, which is able to further increase *n*-butanol titers 3-fold (~30 mg L^{-1}) [*3*]. Bacterial PDHc enzymes have been expressed successfully in *S. cerevisae* as well. Heterologous expression of the PDHc from *Enterococcus faecalis* in *S. cerevisiae* can complement knockout of the PDHc bypass pathway, as shown by similar growth rates between a Δ*acs* strain expression the bacterial PDHc and wild-type yeast [*37*].

In addition to the introduction of a cytosolic pathway for acetyl-CoA generation, the mitochondrial acetyl-CoA pool can also be tapped as it accounts for 30% of the total cellular acetyl-CoA [*2, 38, 39*]. Indeed, introduction of ACL1 and ACL2 from *Y. lipolytica* can improve *n*-butanol titers by 2 fold [*3, 40*]. Since ACL is used mainly in oleaginous yeast and is not a typical pathway in other yeasts, import of acetyl-CoA into the cytosol has also been engineered using the canonical carnitine- mediated translocation system, which is a unidirectional system that transfers acetyl-CoA from the mitochondria to the cytosol for fatty acid synthesis [*41*]. In *S. cerevisiae*, transcription of genes involved in the carnitine shuttle are strongly repressed under glucose-rich media, which means a significant portion of acetyl-CoA is unavailable for cytosolic biosynthesis. In order to identify targets for engineering, strains designed for constitutive expression of the carnitine shuttle and conditional shutdown of cytosolic acetyl-CoA synthesis were subjected to adaptive evolution [*42*]. Genome sequencing of the evolved strains revealed mutations in genes involved in fatty acid synthesis (MCT1), nuclear-mitochondrial communication (RTG2), and a carnitine acetyltransferase (YAT2). Introducing these mutations in the parent strain showed L-carnitine-dependent growth on glucose [*42*]. These findings suggest that the transport of

mitochondrial acetyl-CoA pool can be engineered for use in downstream biosynthetic cytosols pathways.

By combining many advances in rewiring central carbon metabolism, highly efficient production of isoprenoids has been achieved to produce β-farnesene at cost-effective yields at the industrial scale. β-farnesene is sequiterpene with versatile industrial applications for polymers and biofuel [*43, 44*]. In this system, the authors showed that four non-native metabolic reactions were needed to rewire central carbon metabolism in *S. cerevisiae*. In their first generation system, just the PDHc acetate bypass was used in order to create cytosolic acetyl-CoA as a precursor for isoprenoid production [*45*]. However, this pathway has high carbon loss and ATP usage as each acetyl-CoA requires both decarboxylation of pyruvate and activation of acetate. To address this problem, carbon was rewired through the non-oxidative PPP pathway [*46*] and the PDHc acetate bypass was replaced with the PDHc acetaldehyde pathway, resulting in reduced use of ATP and O_2 as well improved carbon yield and redox balance (*Figure 1.5*). Specifically, xPK and PTA were overexpressed to allow acetyl-CoA synthesis from acetyl-phosphate derived from the PPP at zero net ATP and reducing power usage [*46*]. Second, the native PDHc-bypass is energy expensive, thus a prokaryotic acylating acetaldehyde dehydrogenase was overexpressed to generate acetyl-CoA directly from acetaldehyde without ATP input (*Figure 1.3A*, *Figure 1.5*). Finally, they addressed the redox balance challenge by replacing the native NADPH-dependent 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) of the mevalonate pathway with a NADH-specific version. With all of these changes together, this strain showed a large improvement in all metrics as compared to the previous generation strain, which had already been highly optimized (21% improvement in yield to 0.173 g/g glucose, 77% improvement in volumetric productivity to 2.24 g /L h, 25% drop in glucose usage, and 75% drop in O_2 usage) [*45, 47*]. Taken together, these studies show that working with native pathways can lead to intrinsic metabolic and energetic limitations that can be addressed by rewiring the metabolic network with new pathways.

1.5. Exploring new pathways for improving theoretical yields

As discussed above, the central building block acetyl-CoA is mainly generated in heterotrophs by decarboxylation of pyruvate, automatically reducing the theoretical carbon yield by 33%. Thus, yields are already lowered greatly in acetyl-CoA pathways even with an efficient downstream pathway to produce target compounds. To address this challenge, a non-oxidative cyclic pathway termed non-oxidative glycolysis (NOG) was designed that enables the production of stoichiometric amounts of C_2 metabolites from hexose, pentose, and triose phosphate sugars without this carbon loss (*Figure 1.6*) [*46*]. The design of the NOG pathway starts with one input fructose-6-phosphate (F6P) molecule and two equivalents of F6P derived from the cycle. These three F6P are broken down into three acetyl phosphate (AcP) and three erythrose-4-phosphate (E4P) equivalents by phosphoketolase (PK) in an irreversible step serves as the driving force to the NOG pathway. The acetyl phosphate is converted to acetyl-CoA by the phosphotransacetylase (Pta), while E4P is returned to the cycle and rearranged to regenerate F6P for this cycle. The design of this NOG pathway was validated *in vitro* using purified enzymes as well as *in vivo*, where it was shown that acetate can be produced from xylose in *E. coli* (2.2 acetate per xylose) at near theoretical carbon yield limit (2.5 acetate per xylose) and at greater yield than produced using conventional metabolism (1.67 acetate per xylose). This NOG design has been widely adapted to improve cytosolic acetyl-CoA pool as mentioned previously in sections 1.2 and 1.3 [*26, 47*].

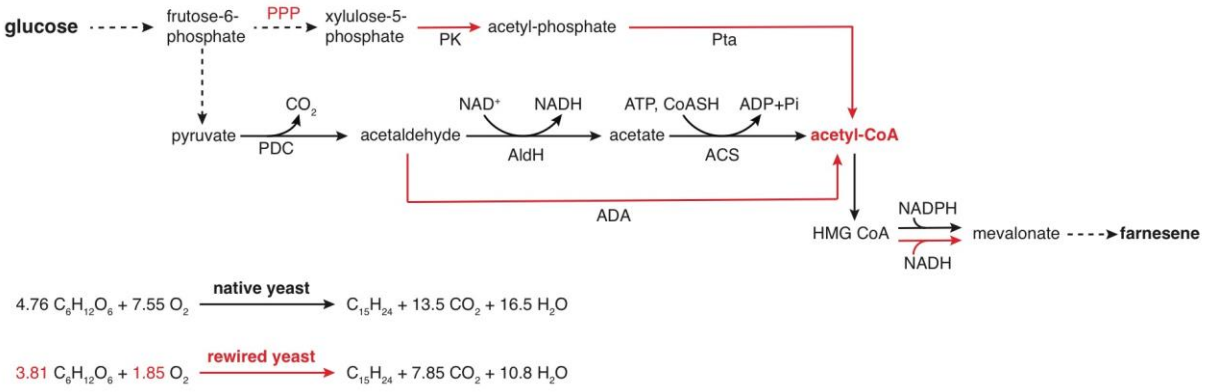


Figure 1.5. Rewiring acetyl-CoA metabolism for farnesene production. Combining the endogenous PPP and overexpressing PK and Pta allows acetyl-CoA synthesis from acetyl-phosphate at zero net carbon loss. Introducing ADA enables direct conversion of acetaldehyde to acetyl-CoA without the cost of ATP. Replacing the NADPH dependent HMG-CoA reductase with a NADH-dependent homolog in the biosynthesis of farnesene improves redox balance. Red lines represent heterologous expressed pathways. PPP: pentose phosphate pathway; PK: phosphoketolase; Pta: phosphtransacetylase; PDC: pyruvate decarboxylase; AldH: aldehyde dehydrogenase; ACS: acetyl-CoA synthetase; ADA: acetaldehyde dehydrogenase acylating.

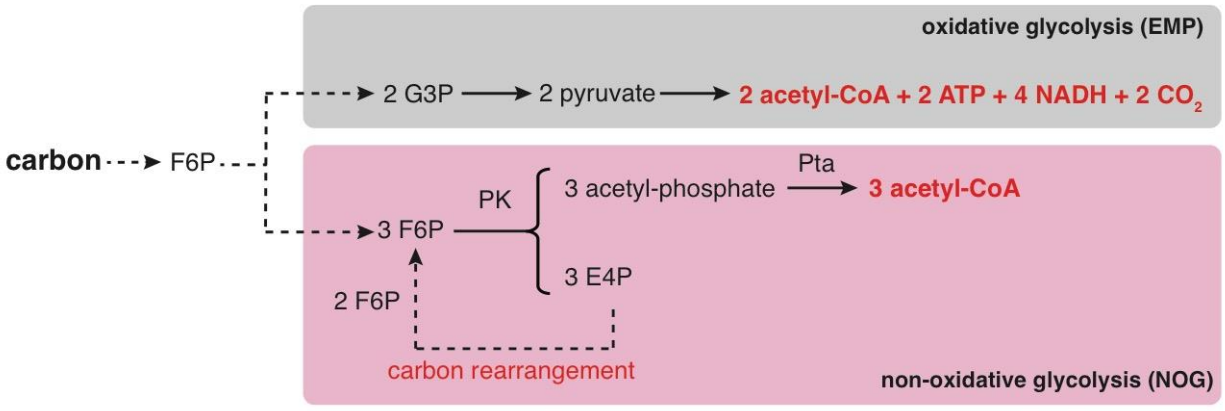


Figure 1.6. Biosynthesis of acetyl-CoA from oxidative glycolysis (EMP) vs. non-oxidative glycolysis (NOG). From canonical glycolysis (EMP), two acetyl-CoA molecules are produced per glucose, along with ATP, NADH, and CO_2. Thus, the theoretical carbon yield from the EMP pathway is 66% due to the loss of carbon in the form of CO_2. Three acetyl-CoA molecules are produced per glucose via the non-oxidative glycolysis (NOG) pathway, reaching the stoichiometric amount production of product, at zero net production of ATP and reducing power. Dotted lines represent multiple steps.

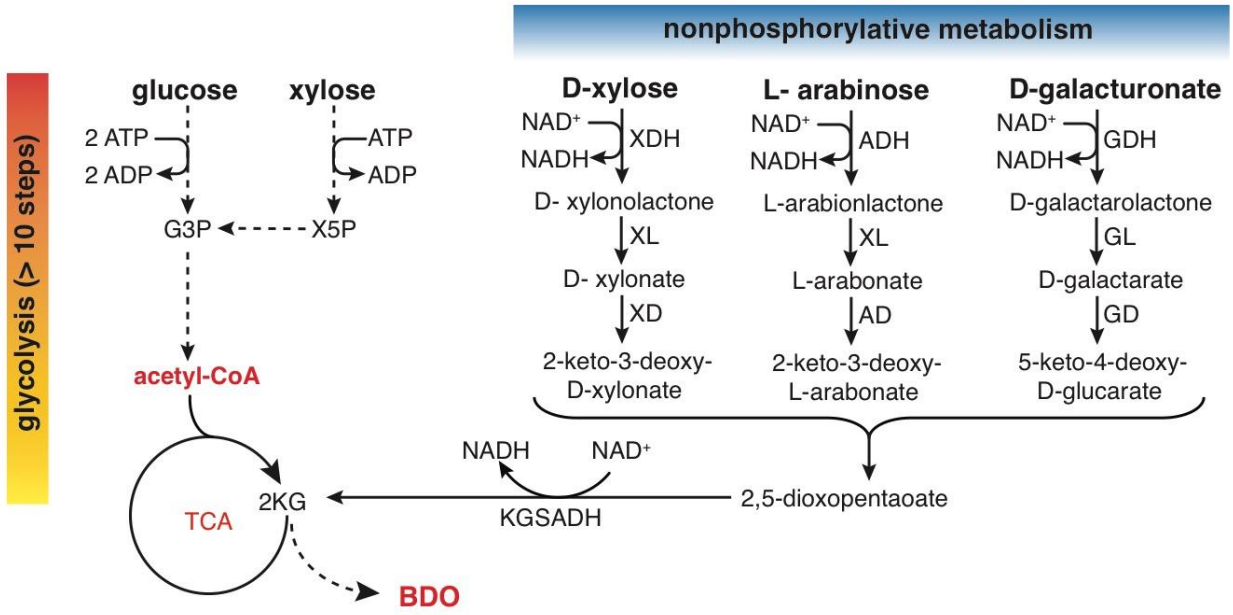


Figure 1.7. Production of BDO from lignocellulosic sugars through nonphosphorylative metabolism. The key TCA building block, 2-ketoglutarate (2KG) was produced from the non-phosphorylative pathway from C_5 sugars at a reduced number of metabolic steps compared to glycolysis. 2KG-dependent butanediol (BDO) pathway was introduced into the engineered host with the nonsphorylative pathway. High yield of BDO was achieved from three different sugars, xylose, arabinose, and galacturonate. The pathway for D-xylose metabolism consists of D-xylose dehydrogenase (XDH), D-xylonolactonase (XL), D-xylonate dehydratase (XD) and 2-keto-3-deoxy-D-xylonate dehydratase (KdxD). The L-arabinose assimilation pathway is composed of L-arabinose dehydrogenase (ADH), L-arabinolactonase (AL), L-arabonate dehydratase (AD) and 2-keto-3-deoxy-L-arabonate dehydratase (KdaD). The pathway for D-galacturonate metabolism was designed by using uronate dehydrogenase (UDH), D-galactarate dehydratase (GD) and 5-keto-4-deoxy-D-glucarate dehydratase (KdgD). DOP produced from these feedstocks is then converted into 2KG by 2-ketoglutarate semialdehyde dehydrogenase (KGSADH), which is a key intermediate of the TCA cycle. PPP, pentose phosphate pathway.

In another example where routing carbon through new pathways can overcome theoretical yield barriers found in conventional metabolism, a nonphosphorylative pathway was designed to produce useful targets in a greatly reduced number of steps and increased yield. Carbon typically enters metabolism through either glycolysis or the PPP, taking multiple steps (>10) before entering the TCA cycle. All three pathways serve as hubs to provide precursor supplies for biosynthesis, but the large number of steps lead to inefficiencies from carbon leakage while also amplifying the complexity of cellular regulation. To address this issue, nonphosphorylative pathways to produce a key TCA cycle building block, 2-ketoglutarate (2KG), from C_5 sugars in six fewer steps that conventional metabolism was identified in *Caulobacter crescentus* and *Pseudomonas fragi* (*Figure. 1.7*). In these, D-xylose and L-arabinose are oxidized and then converted in two steps to 2,5-dioxopentanoate (DOP), which is further oxidized to 2-KG and can feed into the TCA cycle. In this work, the production of 2-KG from uronic acids such as D-galacturonate was demonstrated, increasing the theoretical yield to 100% from 83% through the PPP. Furthermore, a key bioproduct to produce synthetic rubber, butanediol (BDO), could be produced at high yield from all three of these different sugars (D-xylose; 12 g L^{-1}; L-arabinose; 16.5 g L^{-1}; D-galacturonate, 16.5 g L^{-1}) [*48*].

1.6. Examining redox regeneration

Besides carbon yield, it is also important to consider the energetics of redox balance. In heterotrophs, $NAD(P)^+$ and NAD(P)H serve as the key carriers for redox chemistry and provide the reducing power for the cell. These carriers are involved in ~800 biochemical reactions and interact with ~ 400 enzymes in microbial systems [*49*]. From a physiological function perspective, these redox carriers also regulate energy metabolism, intracellular redox state, carbon flux, and cell cycle and imbalances in their homeostasis lead to energy and carbon loss as well as metabolic arrest and cell death [*49*]. As such, there are many systems in place to maintain redox homeostasis, which cannot be perturbed by a biosynthetic pathway if maximal carbon flux is to be achieved [*49*]. For example, in *S. cerevisiae*, there are multiple routes to achieve a neutral redox state, allowing near quantitative conversion of sugar to ethanol [*50, 51*] (*Figure 1.8*). Multiple approaches have been taken to re-balance cellular redox state after the introduction of synthetic pathways, which typically consume reducing power. They include tuning the expression level of cofactor-dependent proteins, engineering proteins to change the specificity of co-factors, and constructing cofactor regeneration systems [*52*] (*Figure 1.9*).

One major example of a key biosynthetic pathway that generates a redox imbalance when run at high flux is fatty acid biosynthesis, which utilizes two NADPH per chain extension cycle. Even in the oleaginous yeast, *Y. lipolytica*, lipid accumulation is limited by the supply of NADPH [*20*]. Since this observation may be a result of the lack of $NADP^+$ ME, multiple approaches were implemented to increase the NADPH pool. First, two $NADP^+$-dependent glycerol-3-phosphate dehydrogenases (GPD; GapC from *Clostridium acetobutylicum* and GPD1 from *Kluyveromyces lactis*) were introduced to switch the cofactor preference from the native NAD^+ GPD, resulting in a ~20% improvement in lipid yield. Similarly, the endogenous ME (ylMAE) is NADH specific and overexpressing a cytosolic $NADP^+$-dependent ME (MCE2 from *Mucor circinelloides*) showed another ~20% improvement on lipid yield to 0.21 g/g glucose [*27*] (*Figure 1.10*). Examples of

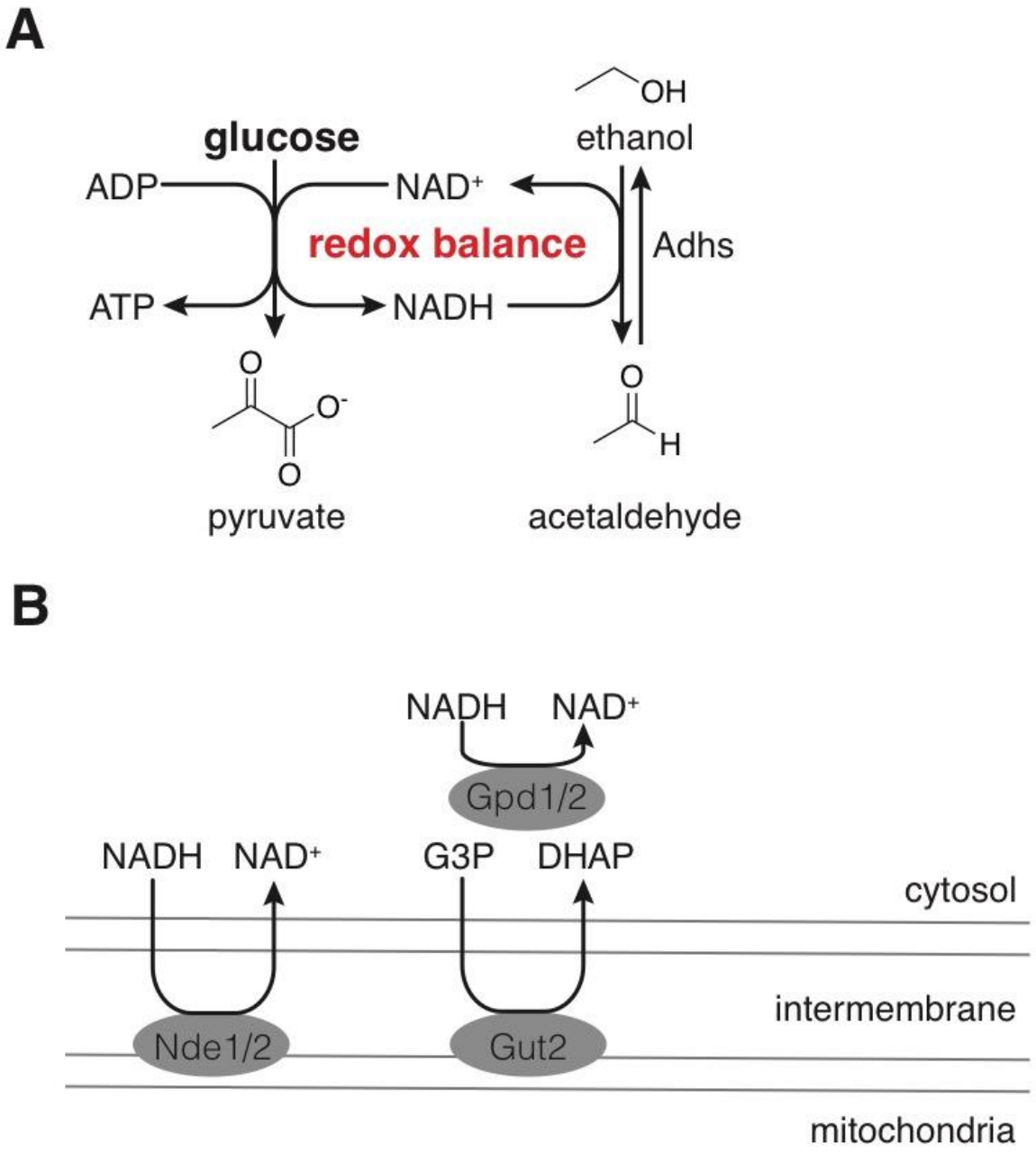


Figure 1.8. Self-redox balancing system in *S. cerevisiae*. (A) Representative example of the redox balanced high flux ethanol fermentation pathway under anaerobic condition. Cytosolic NADH generated from glycolysis can be oxidized by alcohol dehydrogenases to allow glycolysis and ATP production to continue under anaerobic condition. (B) Under aerobic conditions, cytosolic NADH can be oxidized by the external mitochondrial NADH dehydrogenases or the through the respiratory chain via the glycerol-3-phosphate dehydrogenase shuttle. Nde1/2: NADH dehydrogenase; Gut2: membrane-bound glycerol-3-phosphate: ubiquinone oxidoreductase; GPD1/3: cytosolic NADH-linked glycerol-3-phosphate dehydrogenase; G3P: glycerol-3-phosphate; DHAP: dihydroxyacetone phosphate.

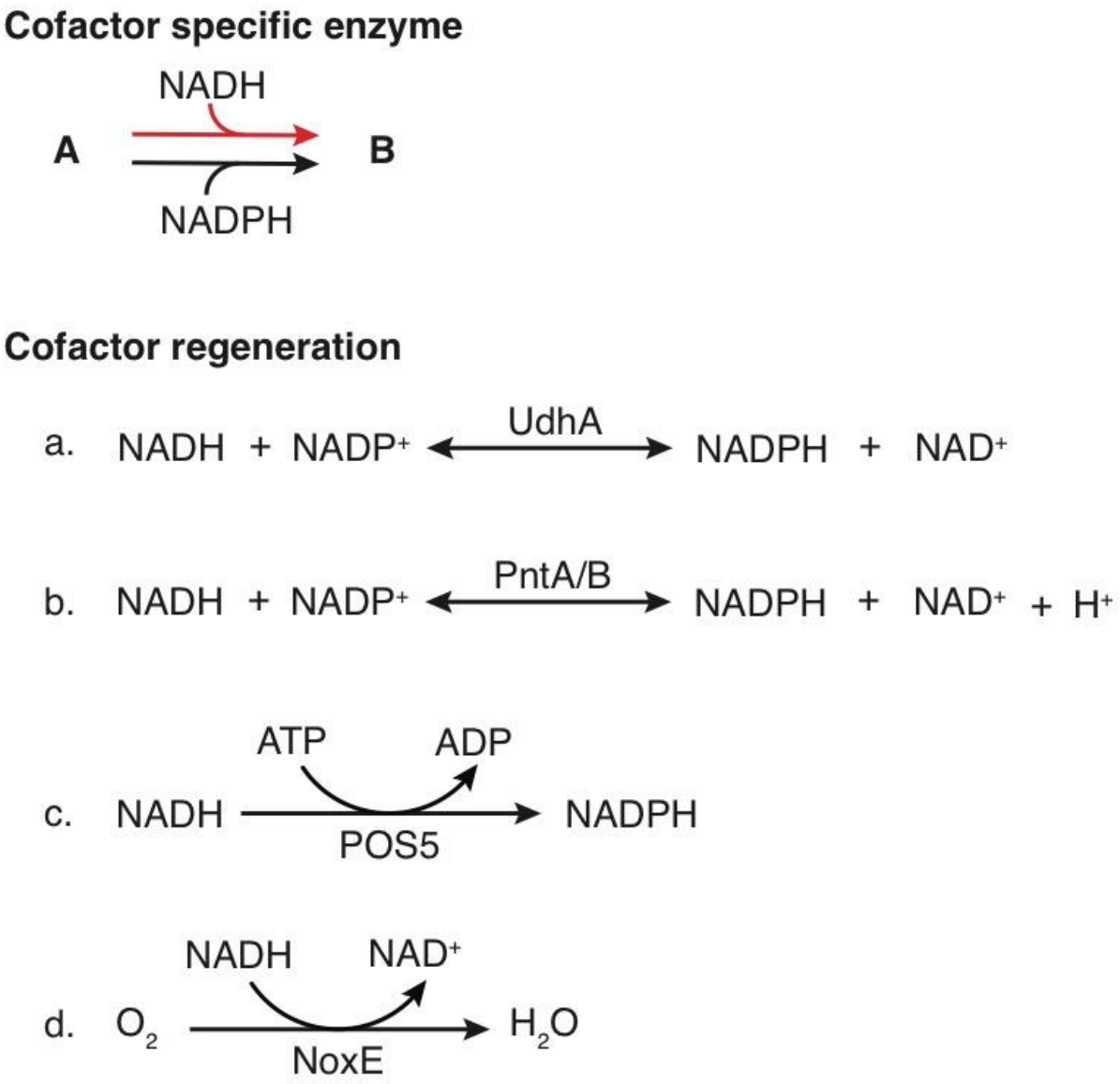


Figure 1.9. Programming redox pools. Redox pools can be balanced by introducing enzymes with specific cofactor preferences. Reactant A can be converted to product B by either a NADH- or NADPH-dependent enzyme. Intracellular cofactor pools can also be manipulated via regeneration reactions. NADH and NADPH can be interconverted by the transhydrogenases, UdhA and PntA/PntB. NADH can also be converted biosynthetically to NADPH by the NADH kinase, POS5, from *S. cerevisiae*. The intracellular NADH:NAD^+ ratio can be changed by overexpressing the water-forming NADH oxidase, NoxE, which catalyzes the reduction of O_2 with NADH.

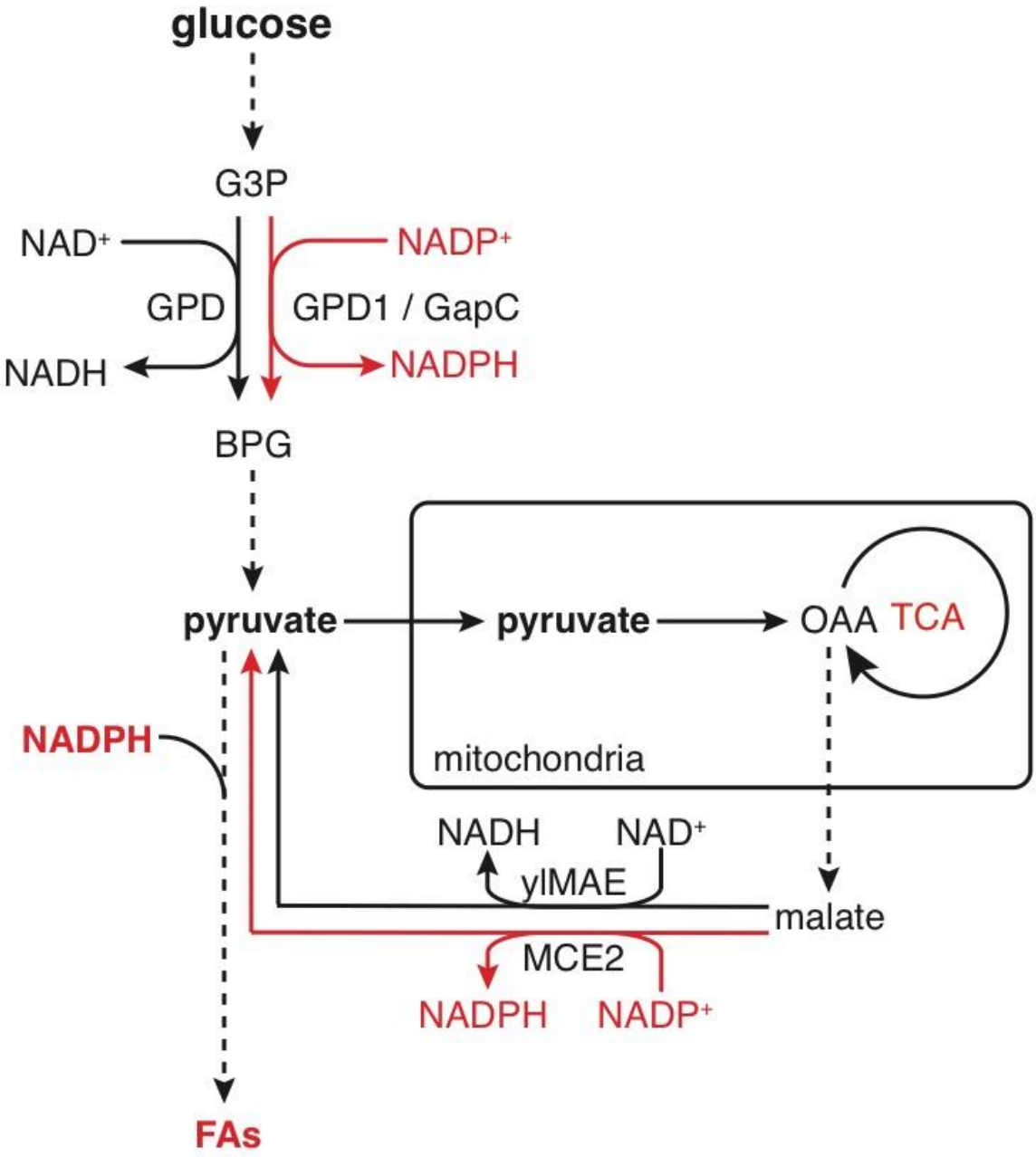


Figure 1.10. Improving lipogenesis by overexpressing NADPH dependent enzymes. Two heterologous $NADP^+$-dependent glyceraldehyde-3-phosphate dehydrogenases (GPDs) were introduced to *Y. lipolytica* to replace the endogenous NAD^+-dependent GPD to increase cellular NADPH for the production of fatty acids. The $NADP^+$-dependent malic enzyme (MCE2) was introduced to replace the native NAD^+-dependent malic enzyme, ylMAE for the oxidative decarboxylation from malate to pyruvate. $NADP^+$-dependent GPDs:, GapC from *Clostridium acetobutylicum* and GPD1 from *Kluyveromyces lactis;* ylMAE, endogenous malic enzyme; MCE2, $NADP^+$-dependent malic enzyme from *Mucor circinelloides*.

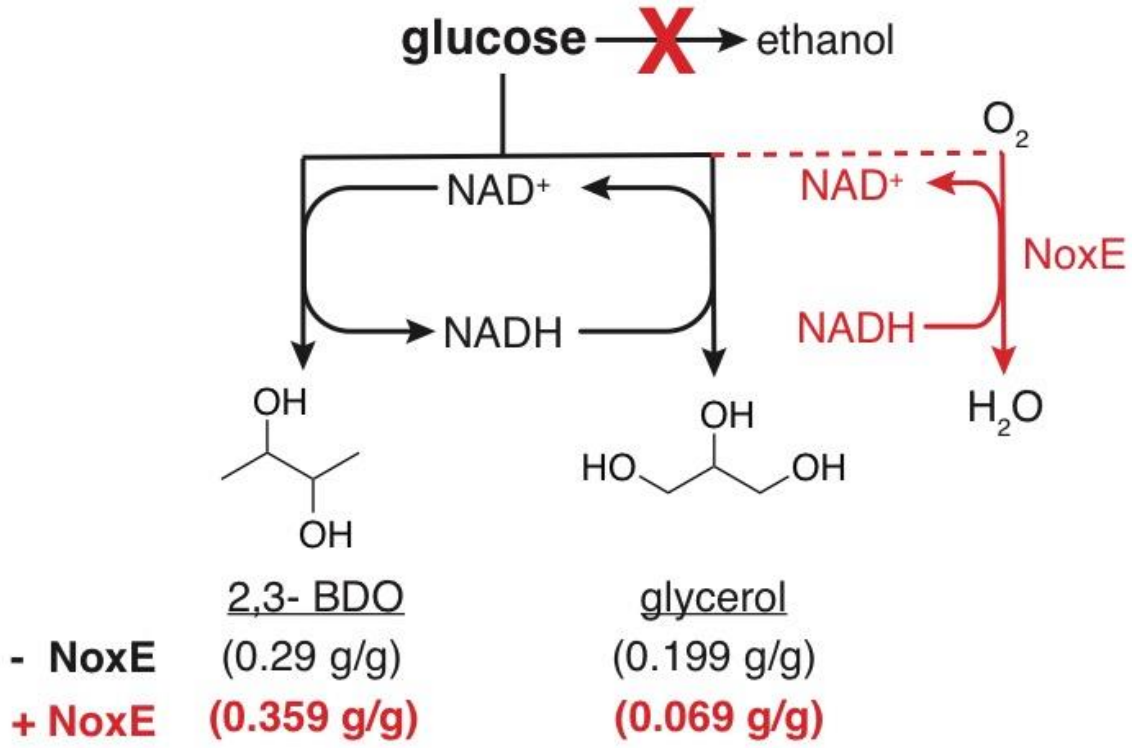


Figure 1.11. Improving 2,3-BDO production by introducing a NAD^+ generation system. Pyruvate decarboxylase (Pdc) is deleted in *S. cerevisiae* as a strategy to increase production titer to 2,3-butanediol (2,3-BDO) by minimizing carbon flux to ethanol production. Excess NADH in the Pdc-deficient strain can be balanced by overexpressing the water-forming NADH oxidase, NoxE. Reduction of the NADH pool from by NoxE results in increased production of 2,3-BDO and reduces the production of glycerol as a side product.

increasing productivity in engineered systems are wide-ranging [*35, 53*], utilizing systems such as transhydrogenases to balance NAD(H) and NADP(H) as well as the use of an NADH kinase to convert NADH to NADPH [*54*].

The synthetic 2,3-butanediol (2,3-BDO) pathway has been used to demonstrated the importance of redox pool on overall performance of the synthetic pathway. The bacterial synthetic 2,3-BDO pathway starts with the condensation of two molecules of pyruvate to produce α-acetolactate, which can be decarboxylated to produce acetoin. Upon reduction, 2,3-BDO can be produced as the biological precursor to butadiene. To eliminate the production of byproducts, a PDC-deficient strain was used as a production host but resulted in increased glycerol production as a redox sink for the excess cytosolic NAD^+ generated. Introduction of NADH oxidase could successfully diverted carbon flux from glycerol to 2,3-BDO. The yield of 2,3-BDO increased from 0.29 g/g glc to 0.359 g/g glc, while the byproduct production of glycerol decreased from 0.199 g/g glc to 0.069 g/g glc [55] (*Figure 1.11*).

1.7. Engineering other cellular processes

In addition to the engineering of metabolic reactions, efficient pathways can be developed using other approaches. Many of these designs are inspired by natural processes, such as feedback regulation, pathway compartmentalization, and metabolism in microbial consortia.

Designing synthetic regulation. Regulation is a key attribute of naturally-occurring metabolic pathways that allows the host to manage and organize resources as well as to minimize system-wide perturbation. With the introduction of a synthetic pathway that feeds upon the natural metabolic network, perturbation occurs at many levels and can limit productivity. In order to address this problem, systems have been design to achieve dynamic regulation that rely on the use of intracellular sensors to balance pathway flux [*56, 57*]. These sensors can be native transcription factors that bind the metabolite of interest and respond to generate a downstream signal or can also be designed from other sensor classes of proteins such as G-coupled protein receptors (GCPRs) [*58*]. For example, a malonyl-CoA sensor, FapR, was used as a metabolic switch to allow dynamic regulation of fatty acids biosynthesis in *E. coli* [*59*]. It has been reported that FapR is a putative transcription repressor for fatty acids biosynthesis genes, responding to malonyl-CoA. Taking the advantage of this natural transcription regulator, two regulatory elements were designed to downregulate a hybrid regulatory unit including the *fap* operator sequence (fap*O*) and either the T7 or native GAP promoters (*Figure 1.12*). Interestingly, these two hybrids behaved differently, leading to malonyl-CoA-dependent upregulation (GAP) as well as downregulation (T7). Thus, a switch could be designed where FapR activates gene expression from the GAP promoter while repressing the T7 promoter when malonyl-CoA levels are low. This malonyl-CoA metabolic switch was used to control fatty acid production with the ACC placed under control of the GAP promoter and the fatty acid synthase, which respectively produce and consume malonyl-CoA. The introduction of this dynamic regulation significantly increased titers of fatty acids from 1.25 g L^{-1} to 3.9 g L^{-1}.

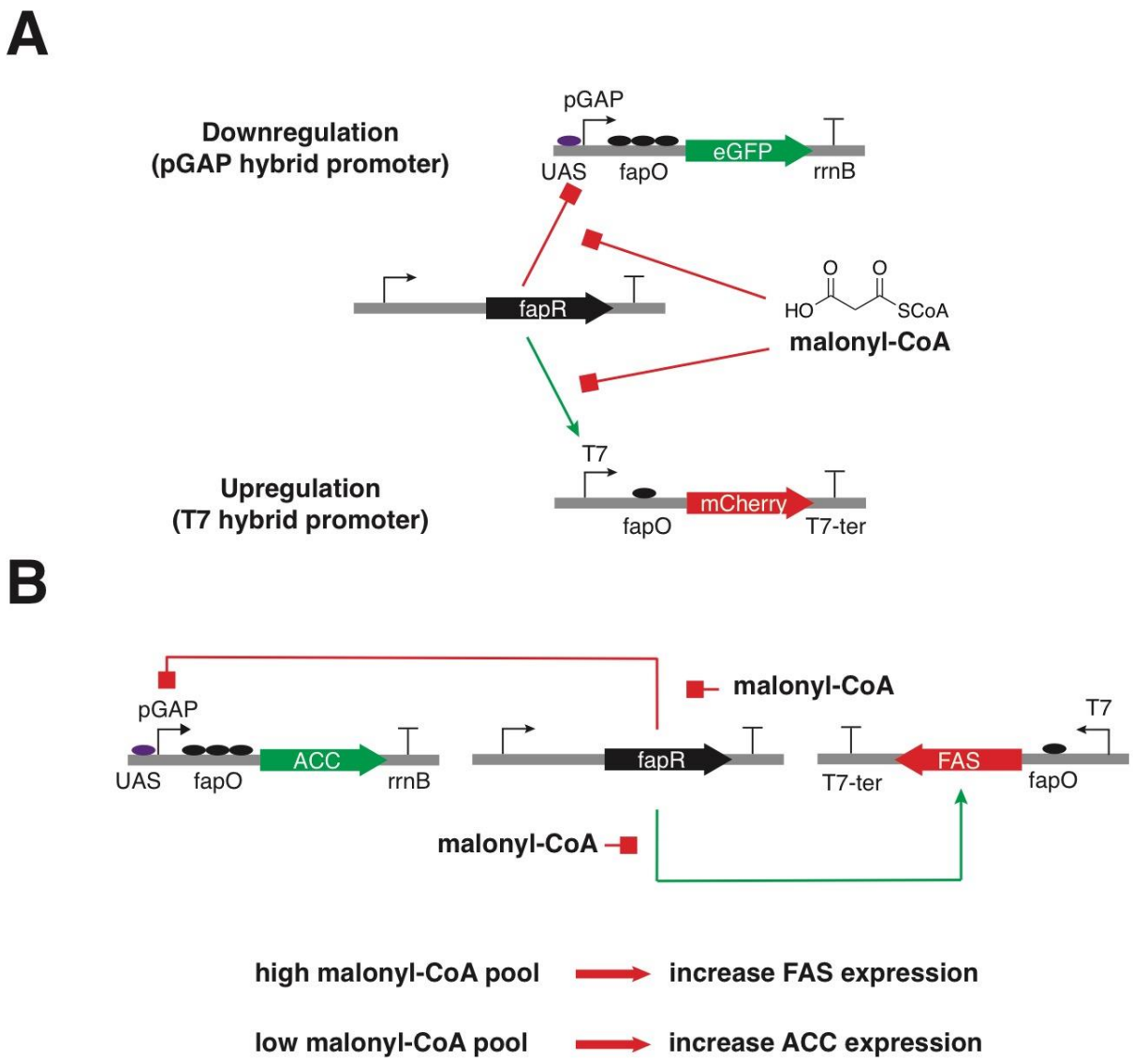


Figure 1.12. Malonyl-CoA regulating hybrid promoters. (A) Two hybrid promoters were designed based on the malonyl-CoA-respsonsive transcription factor, FapR. Coupling the *fap* operator sequence with the native GAP promoter or the T7 promoter resulted in hybrid promoters that respond to malonyl-CoA by upregulation and downregulation of the gene of interest, respectively. (B) These malonyl-CoA regulated hybrid promoters were implemented to produce a malony-CoA switch for fatty acid production. High malonyl-CoA concentrations would lead to upregulation of the fatty acid synthase (FAS) driven by the T7 hybrid promoter and consume malonyl-CoA. Low malonyl-CoA concentrations would release repression of the hybrid GAP promoter, thereby increasing expression of the acetyl-CoA carboxylase (ACC) to increase the production of malonyl-CoA.

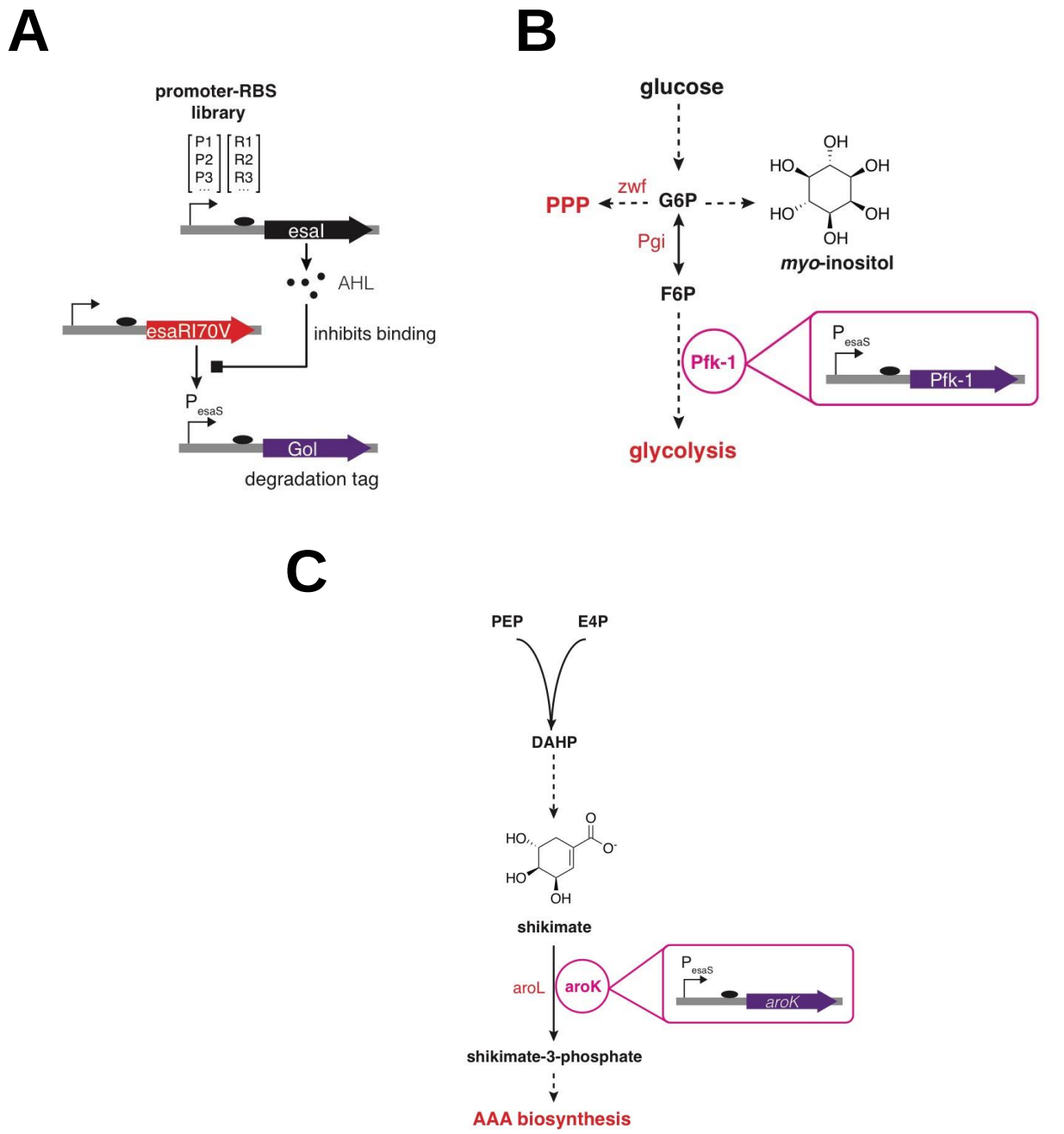


Figure 1.13. Quorum sensing circuit. (A) The level of the transcriptional regulator, EsaRI70V, is controlled by the concentration of AHL produced by EsaI and ultimately regulates protein expression from the P_{esaS} promoter. The construction of a promoter-RBS library to regulate the expression level of EsaI allows dynamic regulation of protein expression driven by the P_{esaS} promoter. (B) The quorum sensing circuit was implemented to control the expression of Pfk-1 and funnel carbon flux to the production of myo-inositol. (C) Introducing the quorum sensing circuit to regulate the expression level of AroK improves production of shikimate without supplementing with aromatic amino acids. AHL, 3-oxohexanoylhomoserine lactone; Pfk-1, phosphofructokinase-A; AroK, shikimate kinase; PEP, phosphoenoylpyruvate; E4P, erythrose-4-phosphate.

Building on this concept of automatous dynamic regulation, more general circuit designs can also be achieved to self-regulate and direct carbon fluxes. One interesting example is the use of quorum sensing pathways to dynamically balance and optimize flux between endogenous and heterologous pathways [*60*] (*Figure 1.13*). Quorum sensing relies on the accumulation of specific small molecules, such as 3-oxohexanoylhomoserine lactone (AHL), in cell populations. In this system, the transcriptional regulator EsaRI70V binds the P_{esaS} promoter in the absence of AHL, whose production is controlled by expression level of the AHL synthase, EsaI. In the presence of AHL, binding is disrupted, activating expression from the P_{esaS} promoter. To develop this system for designing metabolic control valves, a library of promoter and RBS (ribosome binding site) sequences was screened for their response to AHL This circuit was implemented in two different systems to control the relative expression of endogenous and engineered pathways. In the first system, circuit system was implemented to improve production of myo-inositol (MI), which could be converted to glucaric acid, a precursor for biopolymers. To achieve high yield, heterologous expressed pathway must be able to compete with endogenous high flux pathway such as glycolysis. In order to dynamically control glycolytic flux, the key controller for upper glycolysis, phosphofructokinase-A (*pfk-1*) was placed under control of the engineered P_{esaS} promoter, allowing balance to be achieved between growth (high Pfk-1 level and glycolytic flux) and production (low Pfk-1 level and glycolytic flux) phases. This tuning of glycolytic flux yielded up to a 5.5-fold increase in titer of MI up to 1.8 g L^{-1}. The production of a semisynthetic precursor for Tamiflu, shikimate, could also be optimized by targeting a different metabolic branch point. Shikimate is a precursor for aromatic amino acids, which are essential for cell growth. Thus shikimate is usually produced by knockout out the kinases that divert flux to aromatic amino acid pathways while supplemented growth with these amino acids. Using the quorum sensing to dynamically regulate the aroK kinase, shikimate could be produced in minimal media.

Engineering pathway compartmentalization. Another approach that cells use to coordinate metabolic processes is co-localization or compartmentalization. Compartmentalized space in the form of organelles can optimize metabolic activity by controlling and isolating the environment. The mitochondria is a prime example that both sequesters dedicated metabolism and also offers a unique environment compared to the cytosol. For instance, the pH is higher and oxygen concentration is lower in the mitochondria. In addition, a more reducing environment is maintained and the confined space allows for higher local concentrations of metabolites and enzymes. All of these factors play into the optimization for a wide range of redox enzymes, including iron-sulfur cluster-containing enzymes. In one example, a pathway encoding the biofuel, isobutanol, was delivered to the mitochondria via the insertion of a N-terminal mitochondrial localization tag from subunit IV of the yeast cytochrome *c* oxidase (*Figure 1.14*) [*61*]. The fully compartmentalized pathway was found in this case to produce higher titers (~ 500 mg L^{-1}) compared to a partially compartmentalized version (~ 150 mg L^{-1}). It was hypothesized that the higher titers were related to increased local concentration of a pathway enzyme. Indeed, titration of this enzyme in the partially compartmentalized pathway led to concomitant increased in product titer and measurement of subcellular enzyme concentration showed that mitochondrially-targeted enzymes showed as much as a 4 fold increase as compare to cytosolic enzyme.

Exerting cell population quality control. Cell-to-cell variation is often extreme. For example, it has been reported that within the same *E. coli* culture, protein concentration could reach a 10-fold difference between individual cells [*62–65*], highlighting the existence of high- and

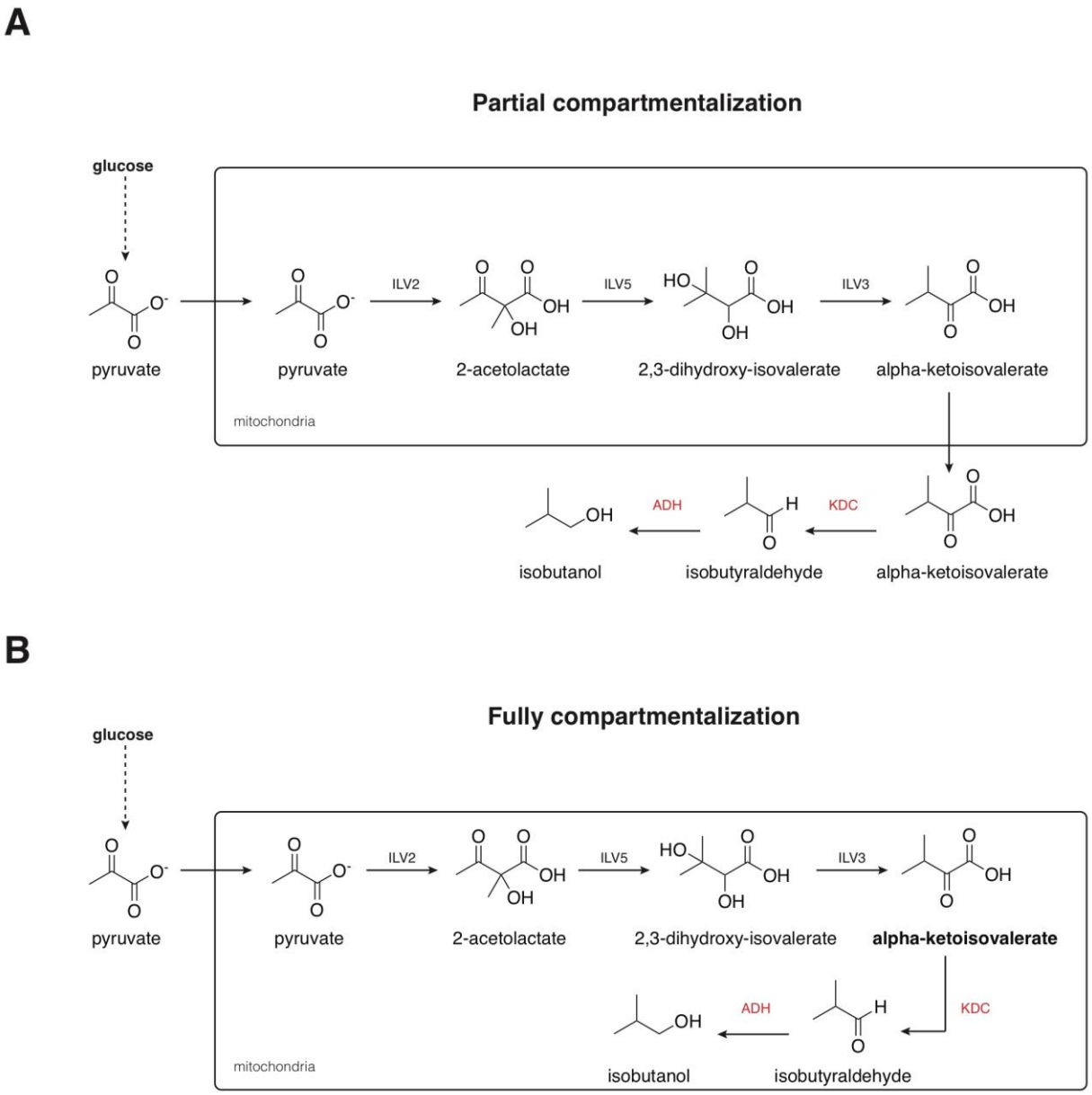


Figure 1.14. Compartmentalization of the isobutanol pathway in *S. cerevisiae*. The isobutanol pathway was used to examine the effect of synthetic pathway comparmentalization. (A) The isobutanol pathway is partially compartmentalized by targeting the first 3 steps of the isobutanol pathway to the mitochondria. The downstream pathway, which is catalyzed by KDC and ADH, is expressed in the cytoplasm. (B) The same isobutanol pathway can also be fully compartmentalized in the mitochondria. KDC, α-keto acid decarboxylase; ADH, alcohol dehydrogenase.

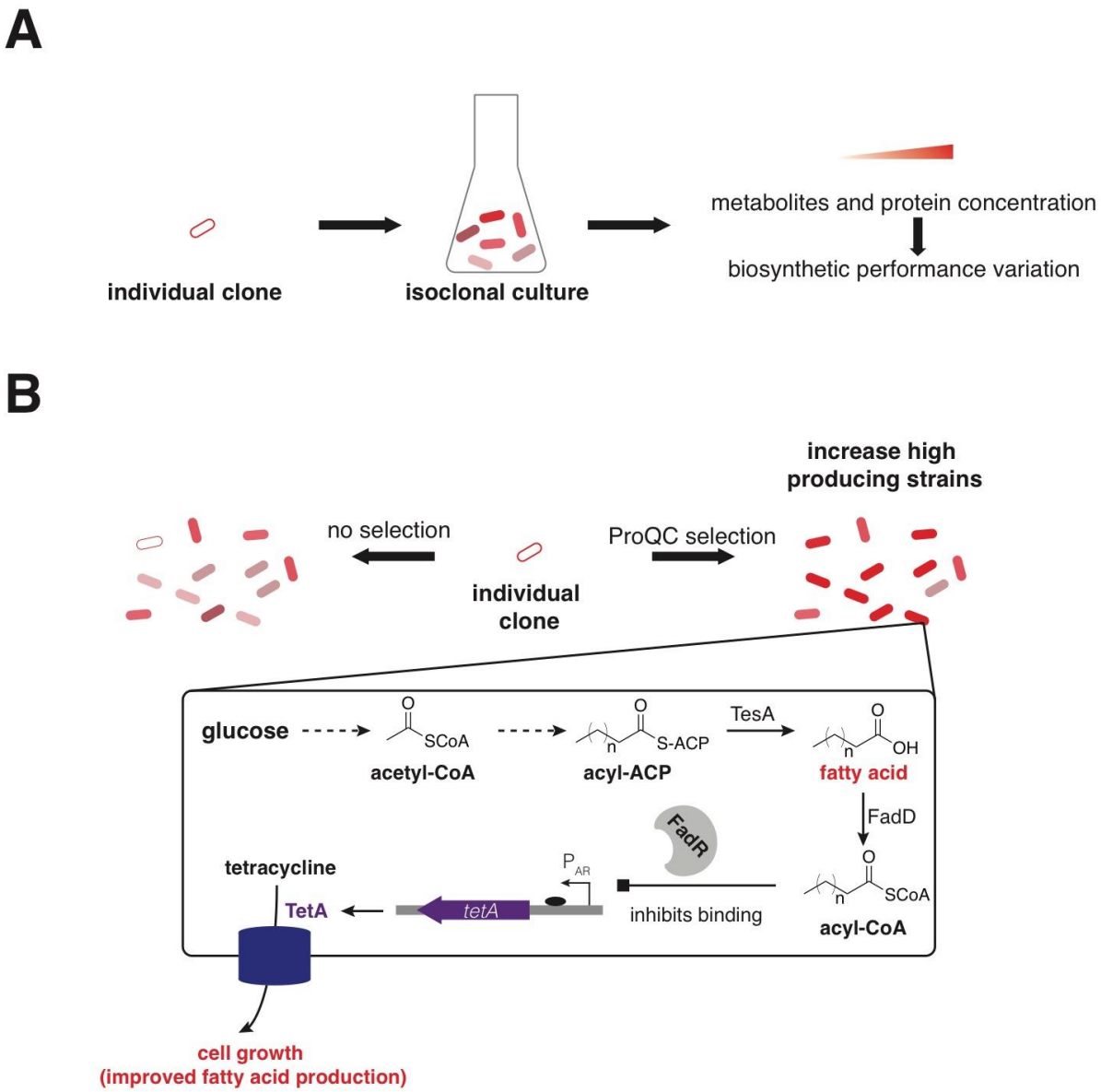


Figure 1.15. Controling non-genetic cell to cell variation. (A) Isoclonal cultures result in cells with very different metabolite and protein concentrations (up to 10-fold difference). This non-genetic cell to cell variation can greatly influence biosynthetic performance. (B) An example of the *in vivo* population quality control (PopQC) design to continuously select high performance within the isoconal culture using the fatty acid synthesis pathway. Cell survival in the presence of tetracycline is coupled to the level of acyl-CoA produced by placing expression of the survival gene (TetA) under the control of the P_{AR} promoter, which is regulated by the acyl-CoA concentration. FadR, acyl CoA-binding protein and transcription factor.

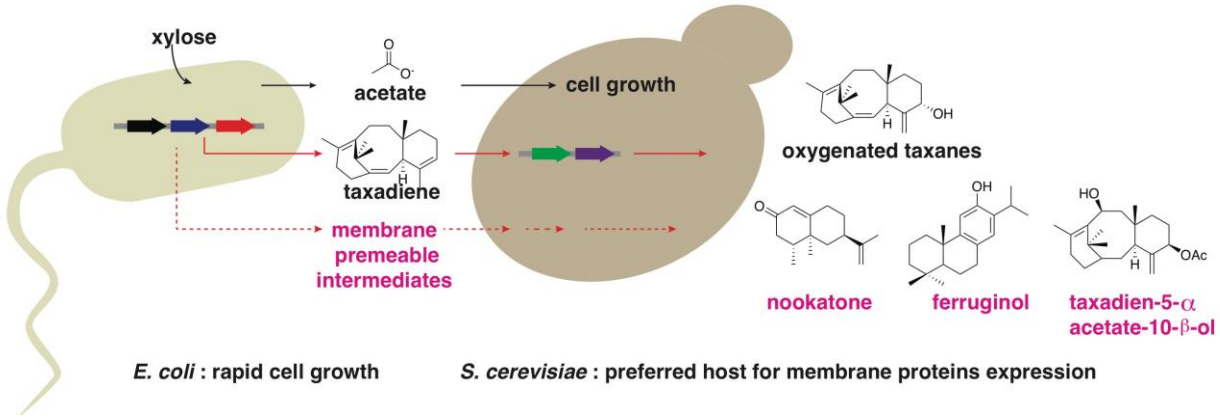


Figure 1.16. Production of complex molecules by microbial partnership. Heterologous expression of long synthetic pathways for complex molecules can be accomplished by leveraging unique characteristics of different hosts. The production of oxygenated taxanes was conducted via two different hosts, *E. coli* and *S. cerevisiae*. The upstream pathway was expressed in the rapid growing *E. coli* host, resulting the production of the taxadiene intermediate from xylose. *S. cerevisiae*, which is typically a better host for the expression of membrane-bound plant enzymes, was used to express the downstream steps for the production of oxygenated taxanes. The similar design could be used for other targets such as nookatone, ferruginol, and taxadien-5-α-acetate-10-beta-ol.

low-performers within a single culture. Applying this approach to fatty acid biosynthesis, cells were treated with a fluorescent fatty acid stain, allowing fluorescence-activated cell sorting (FACS) approach to bin cells based on their differential FFA titer, which was found to range by 9-fold. To address this problem, an *in vivo* population quality control (PopQC) was implemented to continuously select for high-performing non-genetic variants. The design of the technology uses a product- responsive biosensor for fatty acids (FadR) that continuously monitors product abundance and correspondingly regulates expression of a survival gene (TetA for tetracycline resistance) in each cell (*Figure 1.15*). Applying this technology, nongenetic high performers with three-fold increased free fatty acid were selected and with the PopQC in a fed-batch FFA production, (21.5 g L^{-1}) [*66*].

Controlling cell morphology. Accumulation of certain classes of products, such as polymers, can be affected by physical properties of the cell. For example, the storage of polyhydroxyalkanoate (PHA) polymers, which are a family of biodegradable and biocompatible thermal bioplastics, has been reported to be impacted by cell division and morphology. It has been hypothesized that changing the cell division pattern to be non-binary or resulting in two unequal daughter cells could result in a change PHB storage capacity. The deletion of cell fission-related gene, *minC* and *minD*, along with overexpression genes involved in division process (*ftsQ, ftsL, ftsW, ftsN* and *ftsZ*) as well as the cell shape control gene *mreB* resulted in an elongated *E. coli* host. Remarkably, this new morphology resulted better cell growth and an 80% increase in PHB accumulation as compared to the native binary fission cells [*67*].

Production of complex molecules in a microbial consortium. There are many complex metabolic processes that are carried out by microbial consortia [*68, 69*], allowing different biochemical roles to be assigned for each members. In this way, pathways can be optimized in an individual host based on their unique traits and then balanced in the overall metabolic process in partnership other hosts who contribute different chemical abilities to the consortium. Like compartmentalization, this design allows for incompatible metabolic pathways to controlled and coordinated. In engineered systems, one example is the production of complex natural products that come from plant sources where high-flux precursor pathways can be more easily in prokaryotes whereas downstream tailoring enzymes can be more easily expressed in a eukaryotic host. This design was implemented for the production of taxane intermediates in the production of the anticancer drug Taxol [*70*]. Specifically, the biosynthetic pathway for taxane production is expressed in two different hosts – *E. coli* and *S. cerevisiae*, by exploiting the unique traits of each host. The upstream pathway that carries the methylerythritol phosphate- (MEP) pathway, geranylgeranyl diphosphate synthase, and taxadiene synthase and geranyl geranyldiphosphate synthase) to produce the membrane-permeable unfunctionalized taxadiene was expressed in *E. coli*. This strain could be co-cultured with *S. cerevisiae* expressing a cytochrome P450 tailoring enzyme, taxadiene 5-α–hydroxylase (5-α-CYP) and its partner reductase[*71*] to selectively modify taxadiene. The carbon source was selected to optimize co-culture survival, utilizing xylose for *E. coli*, which is not used by *S. cerevisiae*, to prevent ethanol production that would be toxic for *E. coli*. The xylose would then be converted to acetate, which would provide the sole carbon source for *S. cerevisiae*, With this design, along with optimizing the expression of 5-α-CYP and CPR by promoter screening, oxygenated taxanes were successfully produced (33 mg L^{-1}) (*Figure. 1.16*).

1.8. Conclusion and thesis organization

Cells provide enormous potential for synthetic biology, where we could build tools to create innovative solutions to address our current challenges, including health care, energy, and the environment. Many challenges remain in understanding how to efficiently control and rewire carbon flux and metabolism. This thesis describes the design of adaptive evolution strategies to explore regulation of central carbon networks in *E. coli* (*Chapter 2*) and studies to elucidate the underlying mechanisms that control flux through these pathways (*Chapter 3*). This strategy was also implemented in *S. cerevisiae* to enable the study of eukaryotic regulation and metabolic compartmentalization (*Chapter 4*).

1.7. References

1. Zhang, K.; Zheng, S.; Yang, J. S.; Chen, Y.; Cheng, Z. Comprehensive Profiling of Protein Lysine Acetylation in *Escherichia coli*. *J Proteome Res* **2013**, *12*, 844–851.
2. Galdieri, L.; Zhang, T.; Rogerson, D.; Lleshi, R.; Vancura, A. Protein Acetylation and Acetyl Coenzyme a Metabolism in Budding Yeast. *Eukaryot Cell* **2014**, *13*, 1472–1483.
3. Lian, J.; Si, T.; Nair, N. U.; Zhao, H. Design and Construction of Acetyl-CoA Overproducing *Saccharomyces Cerevisiae* Strains. *Metab Eng* **2014**, *24*, 139–149.
4. Zhang, Y.; Dai, Z.; Krivoruchko, A.; Chen, Y.; Siewers, V.; Nielsen, J. i. *FEMS Yeast Res* **2015**, *15*, fov024.
5. Kozak, B. U.; van Rossum, H. M.; Benjamin, K. R.; Wu, L.; Daran, J. M.; Pronk, J. T.; van Maris, A. J. Replacement of the *Saccharomyces Cerevisiae* Acetyl-CoA Synthetases by Alternative Pathways for Cytosolic Acetyl-CoA Synthesis. *Metab Eng* **2014**, *21*, 46–59.
6. Wang, Q.; Zhang, Y.; Yang, C.; Xiong, H.; Lin, Y.; Yao, J.; Li, H.; Xie, L.; Zhao, W.; Yao, Y.; et al. Acetylation of Metabolic Enzymes Coordinates Carbon Source Utilization and Metabolic Flux. *Science (80-.).* **2010**, *327*, 1004–1007.
7. Friis, R. M. N.; Wu, B. P.; Reinke, S. N.; Hockman, D. J.; Sykes, B. D.; Schultz, M. C. A Glycolytic Burst Drives Glucose Induction of Global Histone Acetylation by PicNuA4 and SAGA. *Nucleic Acids Res* **2009**, *37*, 3969–3980.
8. Cai, L.; Sutter, B. M.; Li, B.; Tu, B. P. Acetyl-CoA Induces Cell Growth and Proliferation by Promoting the Acetylation of Histones at Growth Genes. *Mol Cell* **2011**, *42*, 426–437.
9. Kim, E. Y.; Kim, W. K.; Kang, H. J.; Kim, J. H.; Chung, S. J.; Seo, Y. S.; Park, S. G.; Lee, S. C.; Bae, K. H. Acetylation of Malate Dehydrogenase 1 Promotes Adipogenic Differentiation via Activating Its Enzymatic Activity. *J Lipid Res* **2012**, *53*, 1864–1876.
10. Shi, L.; Tu, B. P. Acetyl-CoA and the Regulation of Metabolism: Mechanisms and Consequences. *Curr Opin Cell Biol* **2015**, *33*, 125–131.
11. Sangkharak, K.; Prasertsan, P. Screening and Identification of Polyhydroxyalkanoates Producing Bacteria and Biochemical Characterization of Their Possible Application. *J. Gen. Appl Microbiol* **2012**, *58*, 173–182.
12. Verlinden, R. A. J.; Hill, D. J.; Kenward, M. A.; Williams, C. D.; Radecka, I. Bacterial Synthesis of Biodegradable Polyhydroxyalkanoates. *J Appl Microbiol* **2007**, 1437–1449.
13. Shi, S.; Zhao, H. Metabolic Engineering of Oleaginous Yeasts for Production of Fuels and Chemicals. *Frontiers Microbiol* **2017**.
14. Fan, J.; Cui, Y.; Wan, M.; Wang, W.; Li, Y. Lipid Accumulation and Biosynthesis Genes Response of the Oleaginous Chlorella Pyrenoidosa under Three Nutrition Stressors. *Biotechnol Biofuels* **2014**, *7*.
15. Minhas, A. K.; Hodgson, P.; Barrow, C. J.; Adholeya, A. A Review on the Assessment of Stress Conditions for Simultaneous Production of Microalgal Lipids and Carotenoids. *Front Microbiol*. **2016**.
16. Keasling, J. D. Manufacturing Molecules through Metabolic Engineering. *Science* **2010**, 1355–1358.
17. Vorapreeda, T.; Thammarongtham, C.; Cheevadhanarak, S.; Laoteng, K. Alternative Routes of Acetyl-CoA Synthesis Identified by Comparative Genomic Analysis:

Involvement in the Lipid Production of Oleaginous Yeast and Fungi. *Microbiology* **2012**, *158*, 217–228.

18. Spaans, S. K.; Weusthuis, R. A.; van der Oost, J.; Kengen, S. W. M. NADPH-Generating Systems in Bacteria and Archaea. *Front Microbiol* **2015**.
19. Hao, G.; Chen, H.; Du, K.; Huang, X.; Song, Y.; Gu, Z.; Wang, L.; Zhang, H.; Chen, W.; Chen, Y. Q. Increased Fatty Acid Unsaturation and Production of Arachidonic Acid by Homologous Over-Expression of the Mitochondrial Malic Enzyme in *Mortierella alpina*. *Biotechnol Lett* **2014**, *36*, 1827–1834.
20. Wasylenko, T. M.; Ahn, W. S.; Stephanopoulos, G. The Oxidative Pentose Phosphate Pathway Is the Primary Source of NADPH for Lipid Overproduction from Glucose in Yarrowia Lipolytica. *Metab Eng* **2015**, *30*, 27–39.
21. Li, S. J.; Cronan, J. E. Growth Rate Regulation of *Escherichia coli* Acetyl Coenzyme A Carboxylase, Which Catalyzes the First Committed Step of Lipid Biosynthesis. *J Bacteriol* **1993**, *175*, 332–340.
22. Davis, M. S.; Solbiati, J.; Cronan, J. E. Overproduction of Acetyl-CoA Carboxylase Activity Increases the Rate of Fatty Acid Biosynthesis in *Escherichia coli*. *J Biol Chem* **2000**, *275*, 28593–28598.
23. Shpilka, T.; Welter, E.; Borovsky, N.; Amar, N.; Shimron, F.; Peleg, Y.; Elazar, Z. Fatty Acid Synthase Is Preferentially Degraded by Autophagy upon Nitrogen Starvation in Yeast. *Proc Natl Acad Sci U S A* **2015**, *112*, 1434–1439.
24. Blazeck, J.; Hill, A.; Liu, L.; Knight, R.; Miller, J.; Pan, A.; Otoupal, P.; Alper, H. S. Harnessing *Yarrowia Lipolytica* Lipogenesis to Create a Platform for Lipid and Biofuel Production. *Nat Commun* **2014**, *5*.
25. Sitepu, I. R.; Garay, L. A.; Sestric, R.; Levin, D.; Block, D. E.; German, J. B.; Boundy-Mills, K. L. Oleaginous Yeasts for Biodiesel: Current and Future Trends in Biology and Production. *Biotechnol Adv* **2014**, 1336–1360.
26. Xu, P.; Qiao, K.; Ahn, W. S.; Stephanopoulos, G. Engineering *Yarrowia Lipolytica* as a Platform for Synthesis of Drop-in Transportation Fuels and Oleochemicals. *Proc Natl Acad Sci U S A* **2016**, *113*, 10848–10853.
27. Qiao, K.; Wasylenko, T. M.; Zhou, K.; Xu, P.; Stephanopoulos, G. Lipid Production in *Yarrowia Lipolytica* Is Maximized by Engineering Cytosolic Redox Metabolism. *Nat Biotechnol* **2017**, *35*, 173–177.
28. Chen, A. A. S. *Yarrowia Lipolytica* as an Oleaginous Cell Factory Platform for Production of Fatty Acid-Based Biofuel and Bioproducts. *Front Energy Res* **2014**, *2*.
29. Madzak, C. *Yarrowia Lipolytica*: Recent Achievements in Heterologous Protein Expression and Pathway Engineering. *Appl Microbiol Biotechnol* **2015**, *99*, 4559–4577.
30. Schwartz, C.; Shabbir-Hussain, M.; Frogue, K.; Blenner, M.; Wheeldon, I. Standardized Markerless Gene Integration for Pathway Engineering in *Yarrowia lipolytica*. *ACS Synth Biol* **2017**, *6*, 402–409.
31. Nielsen, J.; Keasling, J. D. Engineering Cellular Metabolism. *Cell* **2016**, *164*, 1185–1197.
32. Galanie, S.; Thodey, K.; Trenchard, I. J.; Filsinger Interrante, M.; Smolke, C. D. Complete Biosynthesis of Opioids in Yeast. *Science (80-.).* **2015**, *349*, 1095–1100.
33. Paddon, C. J.; Keasling, J. D. Semi-Synthetic Artemisinin: A Model for the Use of Synthetic

Biology in Pharmaceutical Development. *Nat Rev Microbiol* **2014**, 355–367.

34. Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*, Sixth.; W. H. Freeman: New York, NY, 2012.
35. Chubukov, V.; Mukhopadhyay, A.; Petzold, C. J.; Keasling, J. D.; Martín, H. G. Synthetic and Systems Biology for Microbial Production of Commodity Chemicals. *NPJ Syst Biol Appl* **2016**, *2*, 16009.
36. Shiba, Y.; Paradise, E. M.; Kirby, J.; Ro, D. K.; Keasling, J. D. Engineering of the Pyruvate Dehydrogenase Bypass in *Saccharomyces cerevisiae* for High-Level Production of Isoprenoids. *Metab Eng* **2007**, *9*, 160–168.
37. Kozak, B. U.; van Rossum, H. M.; Luttik, M. A.; Akeroyd, M.; Benjamin, K. R.; Wu, L.; de Vries, S.; Daran, J. M.; Pronk, J. T.; van Maris, A. J. Engineering Acetyl Coenzyme A Supply: Functional Expression of a Bacterial Pyruvate Dehydrogenase Complex in the Cytosol of *Saccharomyces cerevisiae*. *MBio* **2014** *5*, e01696-14.
38. Pietrocola, F.; Galluzzi, L.; Bravo-San Pedro, J. M.; Madeo, F.; Kroemer, G. Acetyl Coenzyme A: A Central Metabolite and Second Messenger. *Cell Metabolism*. 2015, pp 805–821.
39. Uchida, M.; Sun, Y.; McDermott, G.; Knoechel, C.; Le Gros, M. A.; Parkinson, D.; Drubin, D. G.; Larabell, C. A. Quantitative Analysis of Yeast Internal Architecture Using Soft X-Ray Tomography. *Yeast* **2011**, *28*, 227–236.
40. Rodriguez, S.; Denby, C. M.; Van Vu, T.; Baidoo, E. E.; Wang, G.; Keasling, J. D. ATP Citrate Lyase Mediated Cytosolic Acetyl-CoA Biosynthesis Increases Mevalonate Production in *Saccharomyces cerevisiae*. *Microb Cell Fact* **2016**, *15*, 48.
41. van Maris, A. J.; Luttik, M. A.; Winkler, A. A.; van Dijken, J. P.; Pronk, J. T. Overproduction of Threonine Aldolase Circumvents the Biosynthetic Role of Pyruvate Decarboxylase in Glucose-Limited Chemostat Cultures of *Saccharomyces cerevisiae*. *Appl Env Microbiol* **2003**, *69*, 2094–2099.
42. van Rossum, H. M.; Kozak, B. U.; Niemeijer, M. S.; Dykstra, J. C.; Luttik, M. A.; Daran, J. M.; van Maris, A. J.; Pronk, J. T. Requirements for Carnitine Shuttle-Mediated Translocation of Mitochondrial Acetyl Moieties to the Yeast Cytosol. *MBio* **2016**, *7*.
43. George, K. W.; Alonso-Gutierrez, J.; Keasling, J. D.; Lee, T. S. Isoprenoid Drugs, Biofuels, and Chemicals--Artemisinin, Farnesene, and Beyond. *Adv Biochem Eng Biotechnol* **2015**, *148*, 355–389.
44. McPhee, D. J. Compositions Comprising a Farnesene Interpolymer .
45. Paddon, C. J.; Westfall, P. J.; Pitera, D. J.; Benjamin, K.; Fisher, K.; McPhee, D.; Leavell, M. D.; Tai, A.; Main, A.; Eng, D.; et al. High-Level Semi-Synthetic Production of the Potent Antimalarial Artemisinin. *Nature* **2013**, *496*, 528–532.
46. Bogorad, I. W.; Lin, T. S.; Liao, J. C. Synthetic Non-Oxidative Glycolysis Enables Complete Carbon Conservation. *Nature* **2013**, *502*, 693–697.
47. Meadows, A. L.; Hawkins, K. M.; Tsegaye, Y.; Antipov, E.; Kim, Y.; Raetz, L.; Dahl, R. H.; Tai, A.; Mahatdejkul-Meadows, T.; Xu, L.; et al. Rewriting Yeast Central Carbon Metabolism for Industrial Isoprenoid Production. *Nature* **2016**, *537*, 694–697.
48. Tai, Y. S.; Xiong, M.; Jambunathan, P.; Wang, J.; Wang, J.; Stapleton, C.; Zhang, K. Engineering Nonphosphorylative Metabolism to Generate Lignocellulose-Derived

Products. *Nat Chem Biol* **2016**, *12*, 247–253.

49. Chen, X.; Li, S.; Liu, L. Engineering Redox Balance through Cofactor Systems. *Trends Biotechnol* **2014**, *32*, 337–343.
50. Murray, D. B.; Haynes, K.; Tomita, M. Redox Regulation in Respiring *Saccharomyces cerevisiae*. *Biochim Biophys Acta* **2011**, *1810*, 945–958.
51. Herrero, E.; Ros, J.; Belli, G.; Cabiscol, E. Redox Control and Oxidative Stress in Yeast Cells. *Biochim Biophys Acta* **2008**, *1780*, 1217–1235.
52. Sauer, U.; Canonaco, F.; Heri, S.; Perrenoud, A.; Fischer, E. The Soluble and Membrane-Bound Transhydrogenases UdhA and PntAB Have Divergent Functions in NADPH Metabolism of *Escherichia Coli*. *J Biol Chem* **2004**, *279*, 6613–6619.
53. Uppada, V.; Satpute, K.; Noronha, S. B. Redesigning Cofactor Availability: An Essential Requirement for Metabolic Engineering. In *Current Developments in Biotechnology and Bioengineering: Functional Genomics and Metabolic Engineering*; 2016; pp 223–242.
54. Lee, W. H.; Kim, J. W.; Park, E. H.; Han, N. S.; Kim, M. D.; Seo, J. H. Effects of NADH Kinase on NADPH-Dependent Biotransformation Processes in *Escherichia coli*. *Appl Microbiol Biotechnol* **2013**, *97*, 1561–1569.
55. Kim, J. W.; Seo, S. O.; Zhang, G. C.; Jin, Y. S.; Seo, J. H. Expression of Lactococcus Lactis NADH Oxidase Increases 2,3-Butanediol Production in Pdc-Deficient *Saccharomyces Cerevisiae*. *Bioresour Technol* **2015**, *191*, 512–519.
56. Nandagopal, N.; Elowitz, M. B. Synthetic Biology: Integrated Gene Circuits. *Science*. **2011**, 1244–1248.
57. Bulter, T.; Lee, S.-G.; Wong, W. W.; Fung, E.; Connor, M. R.; Liao, J. C. Design of Artificial Cell–Cell Communication Using Gene and Metabolic Networks. *Proc Natl Acad Sci. U S A* **2004**, *101*, 2299–2304.
58. Ehrenworth, A. M.; Claiborne, T.; Peralta-Yahya, P. Medium-Throughput Screen of Microbially Produced Serotonin via a G-Protein-Coupled Receptor-Based Sensor. *Biochemistry* **2017**, *56*, 5471–5475.
59. Xu, P.; Li, L.; Zhang, F.; Stephanopoulos, G.; Koffas, M. Improving Fatty Acids Production by Engineering Dynamic Pathway Regulation and Metabolic Control. *Proc Natl Acad Sci U S A* **2014**, *111*, 11299–11304.
60. Gupta, A.; Reizman, I. M.; Reisch, C. R.; Prather, K. L. Dynamic Regulation of Metabolic Flux in Engineered Bacteria Using a Pathway-Independent Quorum-Sensing Circuit. *Nat Biotechnol* **2017**, *35*, 273–279.
61. Avalos, J. L.; Fink, G. R.; Stephanopoulos, G. Compartmentalization of Metabolic Pathways in Yeast Mitochondria Improves the Production of Branched-Chain Alcohols. *Nat Biotechnol* **2013**, *31*, 335–341.
62. Taniguchi, Y.; Choi, P. J.; Li, G. W.; Chen, H.; Babu, M.; Hearn, J.; Emili, A.; Sunney Xie, X. Quantifying *E. Coli* Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* **2010**, *329*, 533–538.
63. Li, G.-W.; Xie, X. S. Central Dogma at the Single-Molecule Level in Living Cells. *Nature* **2011**, *475*, 308–315.
64. Guimaraes, J. C.; Rocha, M.; Arkin, A. P. Transcript Level and Sequence Determinants of Protein Abundance and Noise in *Escherichia coli*. *Nucleic Acids Res* **2014**, *42*, 4791–4799.

65. Zenobi, R. Single-Cell Metabolomics: Analytical and Biological Perspectives. *Science*. **2013**.
66. Xiao, Y.; Bowen, C. H.; Liu, D.; Zhang, F. Exploiting Nongenetic Cell-to-Cell Variation for Enhanced Biosynthesis. *Nat Chem Biol* **2016**, *12*, 339–344.
67. Wu, H.; Fan, Z.; Jiang, X.; Chen, J.; Chen, G. Q. Enhanced Production of Polyhydroxybutyrate by Multiple Dividing *E. coli*. *Microb Cell Fact* **2016**, *15*, 128.
68. Fredrickson, A. G.; Stephanopoulos, G. Microbial Competition. *Science* **1981**, *213*, 972–979.
69. Ghoul, M.; Mitri, S. The Ecology and Evolution of Microbial Competition. *Trends Microbiol* **2016**, 833–845.
70. Zhou, K.; Qiao, K.; Edgar, S.; Stephanopoulos, G. Distributing a Metabolic Pathway among a Microbial Consortium Enhances Production of Natural Products. *Nat Biotechnol* **2015**, *33*, 377–383.
71. Ajikumar, P. K.; Xiao, W. H.; Tyo, K. E.; Wang, Y.; Simeon, F.; Leonard, E.; Mucha, O.; Phon, T. H.; Pfeifer, B.; Stephanopoulos, G. Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in *Escherichia coli*. *Science.* **2010**, *330*, 70–74.

Chapter 2. *Evolution of cellular chemistry using synthetic pathways for production of C_4 monomers*

This work was performed in collaboration with Dr. Matthew A. Davis.

2.1. Introduction

The ability of living systems to carry out the tasks needed to support life relies on the existence of a dynamic and complex network of chemical reactions within each cell. Indeed, it is the cell's capacity for chemistry that allows it to intake simple carbon sources and transform them into the thousands of molecules needed to drive and coordinate the fundamental processes that are the hallmarks of life, such as response to the environment, homeostasis, growth and maturation, as well as self-reproduction. As such, cells possess an enormous synthetic potential that can be engineered for targeted chemical synthesis, enabling the reduction of multi-stage traditional synthetic routes into a single fermentation step that can be carried out in water and under ambient temperature and pressure [*1–5*].

However, one major challenge in the development of cell-based chemical synthesis is that the living reaction network used to produce target compounds is also needed to carry out basic cell functions. These reactions are thus subject to many levels of local- and systems-level regulation in order to maintain the necessary coordination between parts of the metabolic network [*6–8*]. In particular, key hubs of the metabolic map, such as the central carbon pathways of glycolysis and the tricarboxylic acid cycle (TCA), form many connections with the rest of the network and are difficult to manipulate as their behavior is affected by multiple inputs and outputs [*2*]. As a result, the construction of high-yielding pathways can be difficult to achieve as evolution drives the cell to direct carbon flux to cell growth and biomass in competition with engineered biosynthesis.

Since these central carbon pathways are closely tied to cell state, they are correspondingly subject to homeostatic mechanisms to ensure robustness to change. Therefore, many simultaneous alterations are needed to rationally engineer carbon flow to insufficiently active nodes [*9–11*]. Another possibility is to use evolution as a non-targeted tool to remodel the metabolic network if product titers can be tied to cell growth [*12, 13*]. In this work, we demonstrate the design of a synthetic pathways to selectively produce three industrially-relevant C_4 monomers, 2-hydroxybutanone, 1,3-butanediol, and *n*-butanol, as bioproduct precursors to methyl vinyl ketone [*14*], 1,3-butadiene [*15*], and 1-butene [*16*] (*Figure 1A*). Using a genetic selection, these pathways could be evolved from theoretical yields of 7-20% to near quantitative yield. Genome sequencing of the evolved strains showed that two gene loci, *pcnB* and *rpoBC*, were found mutated in the most successful daughter cells. Subsequent characterization demonstrates that mutations at these two loci are sufficient to capture the majority of the evolved phenotype and likely operate by large-scale shifts in the transcriptome. Taken together, these results highlight the possibility of synthetic pathways to be used not only for scalable chemical production but also as a platform for discovery and study of cellular function.

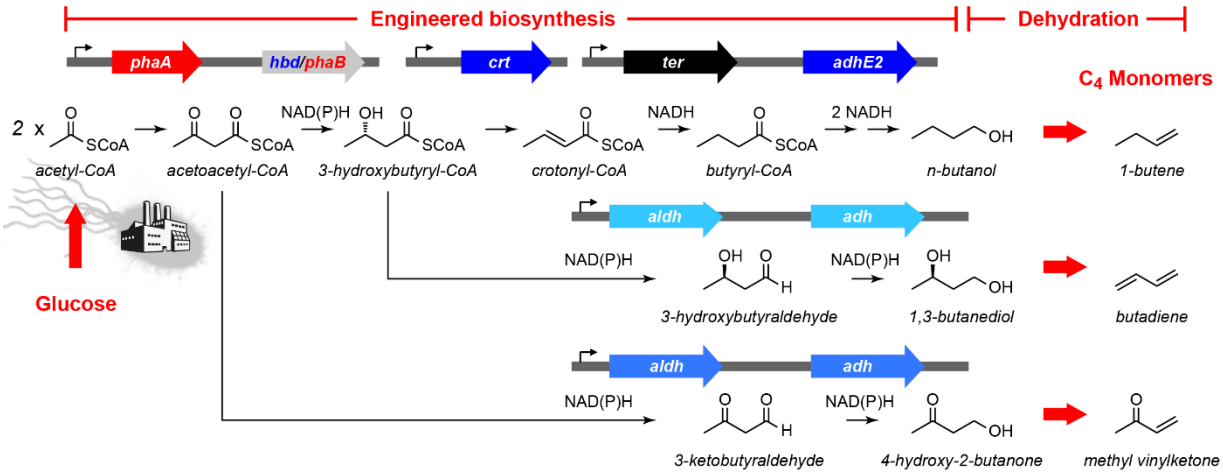


Figure 2.1. Synthetic pathways for production of C_4 monomers. (A) Design of a platform for production of C_4 monomers based on *n*-butanol formation. Identification of selective aldehyde and alcohol dehydrogenases enables the formation of three different C4 products from glucose, n-butanol, 1,3-butanediol, and 4-hydroxy-2-butanone via engineered microbes. Chemical dehydration of these compounds produces the industrially-relevant C_4 monomers, 1-butene, butadiene, and methyl vinylketone, respectively. (*phaA*, acetoacetyl-CoA synthase; *phaB*, *R*-specific NADPH-dependent acetoacetyl-CoA dehydrogenase; *hbd*, *S*-specific NADH-dependent acetoacetyl-CoA dehydrogenase; *crt*, crotonase; *ter*, *trans*-enoyl-CoA reductase; *adhE2*, bifunctional aldehyde/alcohol dehydrogenase; *aldh*, aldehyde dehydrogenase; *adh*, alcohol dehydrogenase. Genes derived from the poly(hydroxyl)alkanote pathway of *Ralstonia eutrophus* are labeled in red. Genes derived from the acetone-butanol-ethanol pathway of *Clostridium acetobutylicum* are labeled in royal blue. Gene from *Treponema denticola* is labeled in black. Light blue *aldh* and *adh* genes denote their general function.)

2.2. Materials and methods

Commercial materials. Terrific Broth (TB), LB Broth Miller (LB), LB Agar Miller, and glycerol, and methylsulfoxide (DMSO) were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), Kanamycin (Km), chloramphenicol (Cm), isopropyl-ß-D-thiogalactopyranoside (IPTG), phenylmethanesulfonyl fluoride (PMSF), tris (hydroxymethyl) aminomethane hydrochloride (Tris-HCl), sodium chloride, dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), and magnesium chloride hexahydrate were purchased from Fisher Scientific (Pittsburgh, PA). Tris-(2-carboxyethyl)phosphine hydrochloride (TCEP) was purchased from Biosynth, Inc. (Itasca, IL). Imidazole was purchased from Acros Organics (Morris Plains, NJ). Sodium hydroxide was purchased from Avantor Performance Materials (Center Valley, PA). A sodium salt hydrate (CoA), acetyl-CoA, butyryl-CoA, acetoacetyl-CoA, β-nicotinamide adenine dinucleotide reduced dipotassium salt (NADH), β-nicotinamide adenine dinucleotide hydrate (NAD^+), formic acid, trichloroacetic acid (TCA), β-mercaptoethanol (BME), lysozyme from chicken egg white, and bovine serum albumin (BSA) were purchased from Sigma-Aldrich (St. Louis, MO). 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, Q5 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. Amicon Ultra 10,000 centrifugal concentrators were purchased from EMD Millipore (Billerica, MA). cOmplete EDTA-free protease inhibitor were purchased from Roche Applied Science (Penzberg, Germany). TEV protease was purchased from the QB3 MacroLab at UC Berkeley. Amicon Ultra spin concentrators and $MultiScreen_{HTS}$ 0.22µm filter plates were purchased from Merck Millipore (Cork, Ireland). D-(+)-glucose was purchased from MP Biochemicals (Santa Ana, CA). 2,4-pentanediol, 1,3-butanediol, and 4-hydroxy-2-butanone were purchased from Sigma-Aldrich (St. Louis, MO). DNA purification kits, Ni-NTA agarose, genomic DNA isolation, and RNeasy RNA isolation kit were purchased from Qiagen (Valencia, CA). Genome library prep Kapa Biosystem Hyper Plus Kit was purchased from Kapa Biosystem (Wilmington, MA). Illumina TruSeq RNA Sample Prep Kit was purchased from Illumina (Hayward, CA).

Bacterial strains. *E. coli* DH10B was used for DNA construction. *E. coli* DH1 (ATCC 39936), DH1Δ5, BW25113Δ5-T1R, DH1Δ5_2406_pcnB(R149L), DH1Δ5_2406_rpoC(M466L), DH1Δ5_2406_pcnB(R149L)_rpoC(M466L) were used for production and evolution experiments.

Gene and plasmid construction. Plasmid construction was carried out using standard molecular biology techniques using the Gibson protocol [*17*]. PCR amplifications were carried out with Q5 DNA polymerase or Phusion DNA polymerase, following manufacturer instructions. Primer

sequences are listed in *Table 2.1.A*. Constructs were verified by sequencing (Quintara Biosciences; Berkeley, CA).

Constructs for genome mutation. The pCRISPR-Gibson1 plasmids were constructed to clone constructs with specific guide sequence to target *E. coli* genome for introduction of point mutants. The parent plasmid, pCRISPR-Gibson1 (#2786), was generated from pCRISPR (Addgene 42875) to introduce cut sites between sgRNA promoter and the sgRNA to facilitate the use of Gibson assembly to introduce guide sequences for the target DNA. All guide sequences were generated using the Benchling CRISPR tool (*Appendix 2.3* for guide sequences).

pCRISPR-PcnB2409 (#2784) was constructed by insertion of the annealed oligonucleotides, P1155 and P1156, and inserted into the XbaI-HindIII site of pCRISPR-Gibson1 using the Gibson protocol.

pCRISPR-RpoC2406 (#2794) was constructed by insertion of the annealed oligonucleotides, P1232 and P1233, and inserted into the XbaI-HindIII site of pCRISPR-Gibson1 using the Gibson protocol.

Production of C_4 compounds in shake flasks. 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 (Kimble Glass; Chicago, IL) or a 250 mL-baffled anaerobic flask with GL45 threaded top (Chemglass). 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. The growth temperature was then 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 min 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. For cultures grown with an oleyl alcohol layer, cultures (40 mL) were grown at 37°C for 3 h before induction with IPTG (1.0 mM). Oleyl alcohol (10 mL) was the added. Cultures were sealed and the headspace was sparged with argon for 3 min. At this time, the growth temperature was reduced to 30 °C. Cultures were grown for 5 d before harvesting. Both the aqueous and organic layers for quantification by GC-FID.

Quantification of *n*-butanol titers. Samples (2 mL) were removed from cell culture and cleared of biomass by centrifugation at 20,817*g* 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 hexanol internal standard (10 g 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} H_2 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). Standard curve was normalized for injection volume using the internal standard.

Quantification of 1,3-butanediol (BDO) and 4-hydroxy-2-butanone (HB) titers. Samples (2 mL) were removed from cell culture and cleared of biomass by centrifugation at 20,817*g* for 2 min using an Eppendorf 5417R centrifuge. The cleared medium samples, or standards prepared in TB medium, were diluted 1:1000 into water and filtered through a 0.22 μm filter (EMD Millipore MSGVN2210). Supernatants were diluted 1- to 1,000-fold fold with water containing 2,4-pentanediol (10 μM) added as internal standard and analyzed on an Agilent 1290 HPLC using a Rezex ROA-Organic Acid H^+ (8%) column (150 × 4.6 mm, Phenomenex) with isocratic elution (0.5% *v/v* formic acid, 0.6 mL min^{-1}, 55 °C). Samples were detected with an Agilent 6460C triple quadrupole MS with Jet Stream ESI source, operating in positive MRM mode (*m/z* 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 0.3125, 0.625, 1.25, 2.5, 5, 10 g L^{-1} 1,3-butanediol and 4-hydroxy-2-butanone.

Anaerobic growth competition and enrichment validation. DH1Δ*5* 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. The qPCR reactions were performed using Bio-Rad Sybr-Green master mix according to the manufacturer protocol, and OD-normalized boiled cell pellet was used as template. Insert primers

Adaptive evolution. Host strains were transformed with appropriate synthetic pathways and plated on LB agar plate with appropriate antibiotics over night at 37 °C. Colonies were picked and grew in 5 ml TB media with 2.5% (*w/v*) glucose replaced with the standard glycerol carbon source overnight at 37 OC with 200 RPM. Overnight cultures were then inoculated to fresh 30 ml TB media with 2.5% glucose with initial OD600 of 0.05 and grew at 37 °C and 200 RPM. Once cultures reached OD_{600}~0.3 – 0.4, induced cultures with 1 mM IPTG and sparged cultures with argon for 3 minutes. Growth temperature was then lowered to 30 °C. Cultures were then serially transferred to fresh media every 24-72 hours with initial OD of 0.05, to approximate continuous growth with limited time spent in stationary phase. The growth time of 24-72 hours was chosen such that the cultures would be in late-log or early-stationary phase. Growth media was TB with 2.5% glucose, 1 mM IPTG, and appropriate antibiotics. Culture OD600 was monitored daily and cultures were transferred when majority of cultures were in late log-phase growth, usually OD600 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 min, and growth was continued at 30 °C in a rotary shaker (200 rpm). Selections were continued until (from three weeks or three months) until no improved strains were isolated from the culture. 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, hydroxybutanone, and butanol production relative to wild type strains.

Genome sequencing. Cells were grown on 10 ml LB media with 2.5% (*w/v*) glucose with appropriate antibiotics overnight at 37 °C. Cell were then spun down at 8000*g* at the Beckman centrifuge. Cell pellets were then processed using the Qiagen Genomic DNA Isolation Kit according to manufacturer specifications. Genomic libraries were the prepared for sequencing using the Kapa Biosystem Hyper Plus Kit with no modification to the standard protocol. For each library, 1 µg of genomic DNA was used with 3 µl of adapter (40 µM) per ligation. A double-sided selection to obtain 600 bp fragments was then performed using 0.55 vol of right and 0.6 vol of left Ampure XP beads (Beckman Coulter). No PCR amplification was carried out after the size selection. Libraries were sequenced at the UC Davis DNA Core Facility with PE300 sequencing using an Illumina MiSeq. Sequencing results were mapped against the the *E. coli* genome (DH1-Accession ID: NC_017625, BW25113 - Accession ID: NZ_CP009273) and compared against reads obtained from our DH1Δ*5* or BW25113Δ*5* parent strain using Breseq v. 0.25d [*18*].

Cell lysate enzyme assays. Biomass was harvested at the end of production and stored at -80 °C. Frozen cell pellets (from 2 ml culture) were thawed and resuspended in 500 µL of 100 mM Tris-HCl pH 7.5 containing DTT (5 mM) and PMSF (0.5 mM).

PhaA. Thiolysis activity was measured by monitoring the enolate form of acetoacetyl CoA as previously described[19]. Assays were performed at 30 °C in a 96 well plate in a total volume of 100 µL containing 100 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 10 µM CoA, and 20 µM acetoacetyl CoA.

Hbd, Ter, Aldh, and Adh activities were assayed as described[20]. Briefly, all assays were perfomed at 30 °C in a 96 well plate in a total volume of 100 µL. The mixture for the *hbd* assays contained 100 mM Tris-HCl, pH 7.5, 100 µM acetoacetyl CoA, 100 µM NADH. The *hbd* activity was monitored by the oxidation of NADH at 340 nm. The mixture for the *Ter* assays contained 100 mM Tris-HCl, pH 7.5, 100 µM NADH, and 50 µM crotonyl CoA. The *Ter* activity was monitored by the oxidation of NADH at 340 nm. The mixture to assay the aldehyde domain of AdhE2 assays contained 100 mM Tris-HCl, pH 7.5, 0.5 mM DTT, 400 µM NAD^+, 400 µM CoA, and 10 mM butyraldehyde. The acitivity of the aldehyde domain was monitored by the reduction of NAD^+ at 340 nm. The mixture to assay the alcohol domain of AdhE2 contained 100 mM Tris-HCl, pH 7.5, 0.5 mM DTT, 400 µM NADH, and 10 mM butyraldehyde. The activity of the alcohol domain was monitored by the oxidation of NADH at 340 nm.

RNA sequencing and analysis. Cells with synthetic pathways were harvested after 24 hours post induction with IPTG for RNA isolation. RNA was isolated using the RNeasy RNA isolation kit (Qiagen). In house rRNA removal method was used to remove rRNA before sequencing. 5 µg of total RNA was treated with 4.5 µL of TURBO DNaseI (ThermoFischer) in a 50 µL reaction including 5 µL of 10X buffer to remove genomic DNA. The reaction was incubated at 37 ^oC for 30 minutes. The reaction was diluted with 100 µL of Bufffer RLT and 200 µL of 70% ethanol and transferred to an RNeasy column (Qiagen) for RNA cleanup following the manufacture instructions. 1 µg of DNase treated RNA was combined with 1 µL of 0.5 µM DNA probes (*Table 2.1.B*) with Hybridization buffer (200 mM NaCl, 100 mM Tris-HCl[7.5]) up to 20 µL.

Hybridization of oligos occurred by holding at 95 °C for 2 minutes, followed by a gradient to 45 °C at -0.1 C/s. 5U of RNase H (Epicentre) in 2.5 μL of 10 X Digestion buffer (0.5M Tris-HCl [7.5], 1M NaCl, 200 mM MgCl2) were added, and the resulting mixture was incubated at 45 °C for 30 minutes. Following cleanup with the Qiagen RNeasy Kit, the sample was treated with 3 U of TURBO DNaseI. Finally, the Qiagen RNeasy Kit was used to clean up samples one last time before RNA-Seq library prep. RNA-Seq libraries were prepared using the Illumina TruSeq RNA Sample Prep Kit. Samples were sequenced with Illumina HiSeq4000 at UC Davis DNA core. Reads were mapped using the Kallisto [*21*] and Sleuth [*22*]. Functional enrichment analysis of differentially expressed genes is based on clusters of orthologous groups (COG) categories provided by the IMG-ER annotation [*23*].

Generation of chromosomal point mutations. Point mutations were made using the CRISPR Cas9 system[24] [25]. Briefly, cell was transformed with the pKD46-Cas9-RecA-Cure which allows the expression of the Cas9 protein for double stranded DNA break and the RecA protein to assist homology recombination. Single transformant was picked and inoculate in liquid culture to make electro-competent cells. Then the cells carried the pKD46-Cas9-RecA-Cure plasmid was transformed with both the pCRISPR plasmid with specific guide and the double stranded DNA repair fragment that carry the desire sequence. The repair fragment also carries a silent mutation to remove the PAM site and a phosphatioate modification at both the 5' and 3' end. Transformations were recovered and plated on plate with appropriate selection markers. Colonies were validated by Sanger sequencing.

DH1Δ*5* _2406_*pcnB*(R149L) - C<u>G</u>C → C<u>T</u>C mutation at position 446 that corresponds to the *pcnB*(R149L) mutation was made in the strain DH1Δ5 using the CRISPR Cas9 system. DH1Δ*5* was transformed with pKD46-Cas9-RecA-Cure and plated in appropriate antibiotic resistant LB agar plate and incubated at 30^{o}C overnight. Single colony was picked and inoculated in 10 ml LB liquid media with appropriate antibiotic overnight at 30 °C. Overnight culture was then diluted in fresh LB media with 0.2% of arabinose (to induce RecA) added to OD_{600} ~ 0.01. Once culture reached an OD_{600} of 0.4 and cells were harvested to make electro-competent cells. DH1_ pKD46-Cas9-RecA-Cure electro-competent cells were then transformed with pCRISPR_gibson_1guide_2409pcnB (#2784) plasmid and repair fragments (P1227_2406_pcnB RF_R and P1226_2406_pcnB RF_F). Cells were recovered at 30 °C for 1.5 hrs and plated on appropriate antibiotic selection LB agar plate. Plate was incubated at 30 °C incubator overnight. Colonies were picked and validated desire sequence by Sanger sequencing (Quintara Biosciences). Once sequences were confirmed, colony was inoculated in 10 ml LB media with 0.05 mM IPTG to induce the guide to target and cure the pCRISPR_gibson_1guide_2409pcnB (#2784) plasmid. Once the pCRISPR_gibson_1guide_2409pcnB (#2784) plasmid is cured, cells were grown at 37°C to cure the pKD46-Cas9-RecA-Cure plasmid, which contains the temperature sensitive origin of replication.

DH1Δ*5* _2406_*rpoC*(M466L) – <u>A</u>TG →<u>C</u>TG mutation at position 1396 that corresponds to the *rpoC*(M466L) mutation was made in the strain DH1Δ5 using the CRISPR Cas9 system as described above. The pCRISPR_gibson_1guide_2406_rpoC (#2794) plasmid and repair fragments (P1231_2406_rpoC_RF_R and P1230_2406_rpoC_RF_F) were used.

DH1Δ*5* _2406_*pcnB*(R149L)_*rpoC*(M466L) – the double mutant was made using the CRISPR Cas9 system as described above in a sequencial manner. Once the *pcnB*(R149L) mutation was confirmed and the pCRISPR_gibson_1guide_2409pcnB (#2784) plasmid was cured, cells were grown up to make electro-compotent cells. Cells were then transformed with the pCRISPR_gibson_1guide_2406_rpoC (#2794) construct and repair fragments (P1231_2406_rpoC_RF_R and P1230_2406_rpoC_RF_F). Once the desired mutations were confirmed with sequencing, cells were growing in IPTG containing media to cure the pCRISPR_gibson_1guide_2406_rpoC (#2794) plasmid. Finally, cells were growing at 37 ºC to cure the pKD46-Cas9-RecA-Cure plasmid, which contains the temperature sensitive origin of replication.

2.3. Results and Discussion

Design of a genetic selection for C_4 production. A large number of naturally-occurring pathways that are capable of quantitative transformation of a primary carbon source to product participate in anaerobic fermentation. Under anaerobic conditions, carbon assimilation pathways like glycolysis serve as the primary route for cellular ATP synthesis, since aerobic respiration is unavailable or shut down due to the lack of oxygen as a terminal electron acceptor [*12, 26*] (*Figure 2.2*). Fermentation pathways then convert the metabolic intermediate of carbon assimilation to product in a reaction that allows for the stoichiometric recycling of redox cofactors, thereby maintaining ATP synthesis and cell maintenance. High pathway flux is thus driven by cell survival as well as the low ATP yield of fermentation compared to oxidative phosphorylation, which provides an added advantage of minimal loss of carbon to competing biomass accumulation [*27*]. As such, anaerobic production is often preferred for industrial fermentations for the advantages provided by high theoretical yields as well as eliminating the challenge of culture oxygenation on large-scale [*28*]. Lactate and ethanol production provide the paradigms for this process, resulting in rapid and near-quantitative yield from sugar via reduction of pyruvate or pyruvate-derived acetaldehyde, respectively (*Figure 2.2*).

Like ethanol and lactate, the C_4 alcohol, *n*-butanol, can serve to balance glucose fermentation because its biosynthesis recycles the four NADH produced per glucose. However, a major challenge is that the production of *n*-butanol, as well as a broad range of other target compounds, typically depends on the use of the acetyl-CoA building block, which is highly regulated at many levels [*29, 30*]. Acetyl-CoA serves as a central point of many metabolic decision points and its synthesis and usage are thus tightly controlled. In particular, flux to acetyl-CoA drops drastically under anaerobic conditions as both biosynthesis and cell growth are greatly reduced during fermentative growth (*Figure 2.2*). Indeed, *n*-butanol titers drop drastically when our first-generation *Escherichia coli* production strain [20] is cultured anaerobically. In order to reduce carbon flow to competing native pathways, the major fermentation pathways [*31*] were knocked out of *E. coli* DH1 to generate a selection strain (DH1 Δ*ldhA* Δ*adhE* Δ*frdBC* Δ*poxB* Δ*ackA-pta* strain (DH1Δ*5*)) that would require the production of *n*-butanol under anaerobic conditions (*Figure 2.3 and 2.4A*).

Developing a platform for the production of C_4 commodity chemicals. We wanted to explore the possibility of producing other important C_4 commodity chemicals from our *n*-butanol pathway

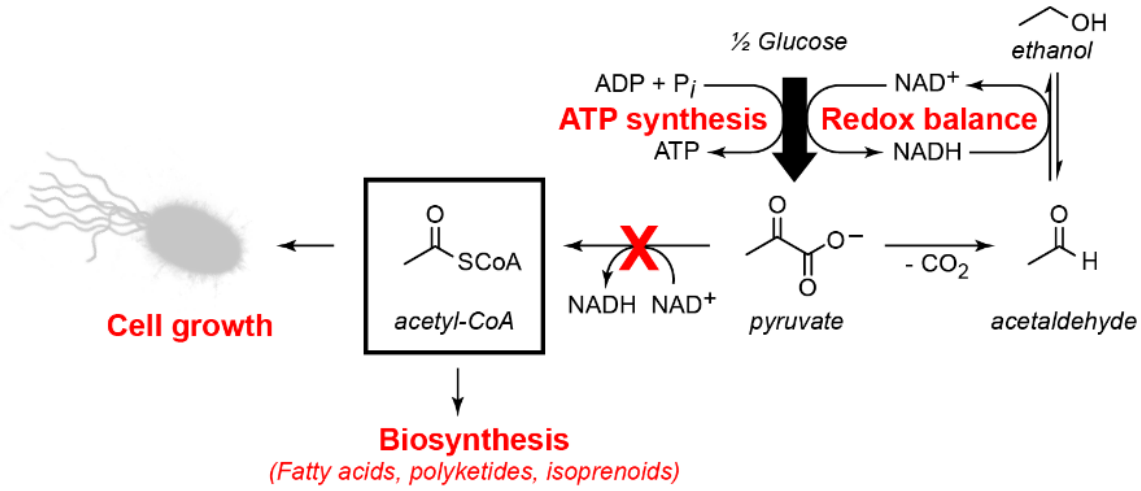


Figure. 2.2. Anaerobic fermentation pathways can operate at near quantitative yields in the absence of O_2. Under these conditions, substrate-level phosphorylation pathways such as glycolysis serve as the only route to ATP synthesis but require the use of NAD^+. In Baker's yeast (*Saccharomyces cerevisiae*), decarboxylation of pyruvate and subsequent reduction to ethanol allows for the stoichiometric regeneration of NAD^+ and is required for cell survival. Because of the low ATP yield under anaerobic growth, cell growth is greatly reduced as well as flux to anabolic pathways utilizing the key building block, acetyl-CoA. As a result, acetyl-CoA is not readily available for the downstream biosynthesis of a broad range of target compounds during anaerobic growth.

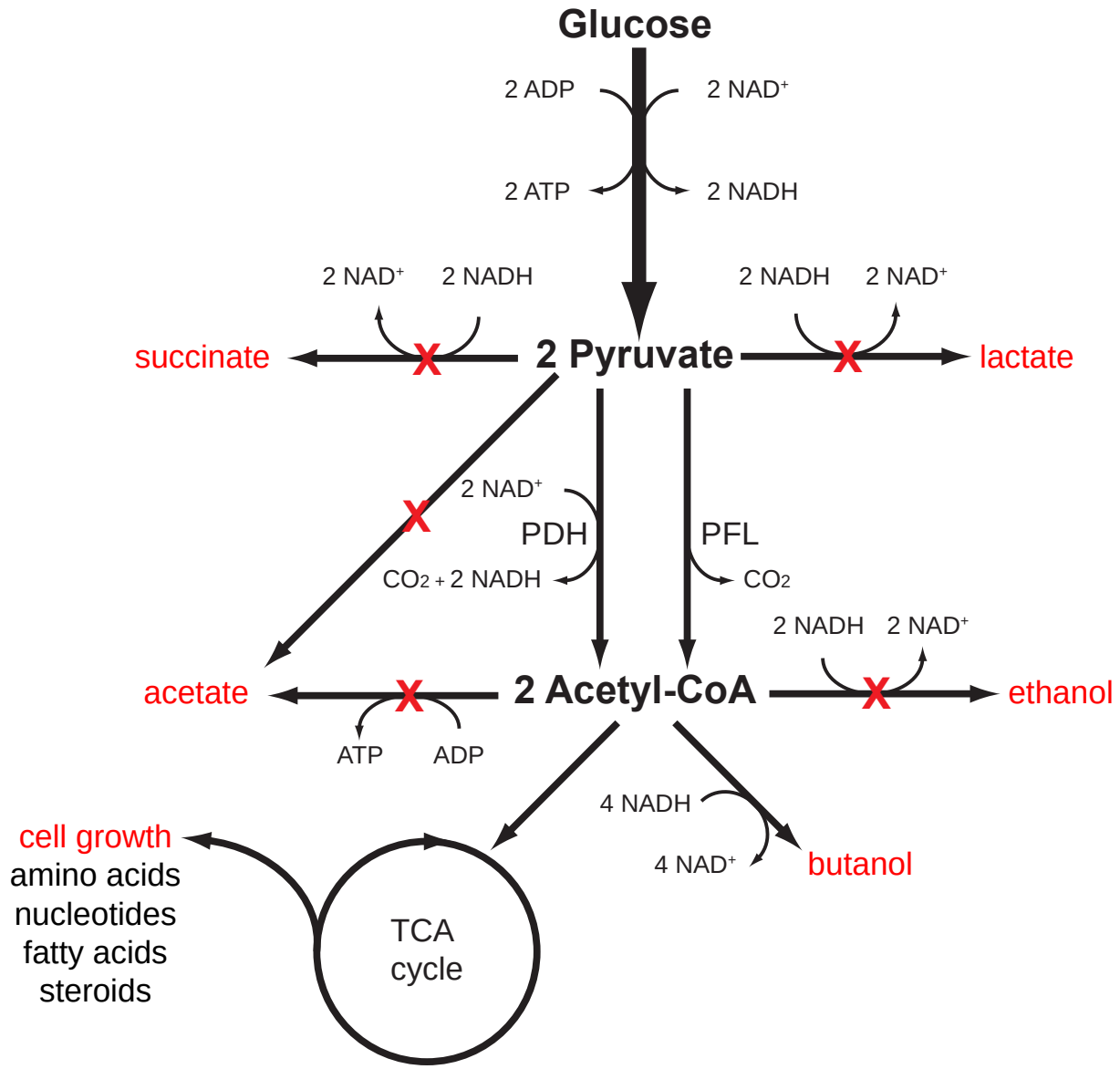


Figure. 2.3. Fermentation pathways of *E. coli* and gene knockouts. Major fermentation pathways of *E. coli* and the five gene loci deleted in the DH1Δ5 strain (Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC*).

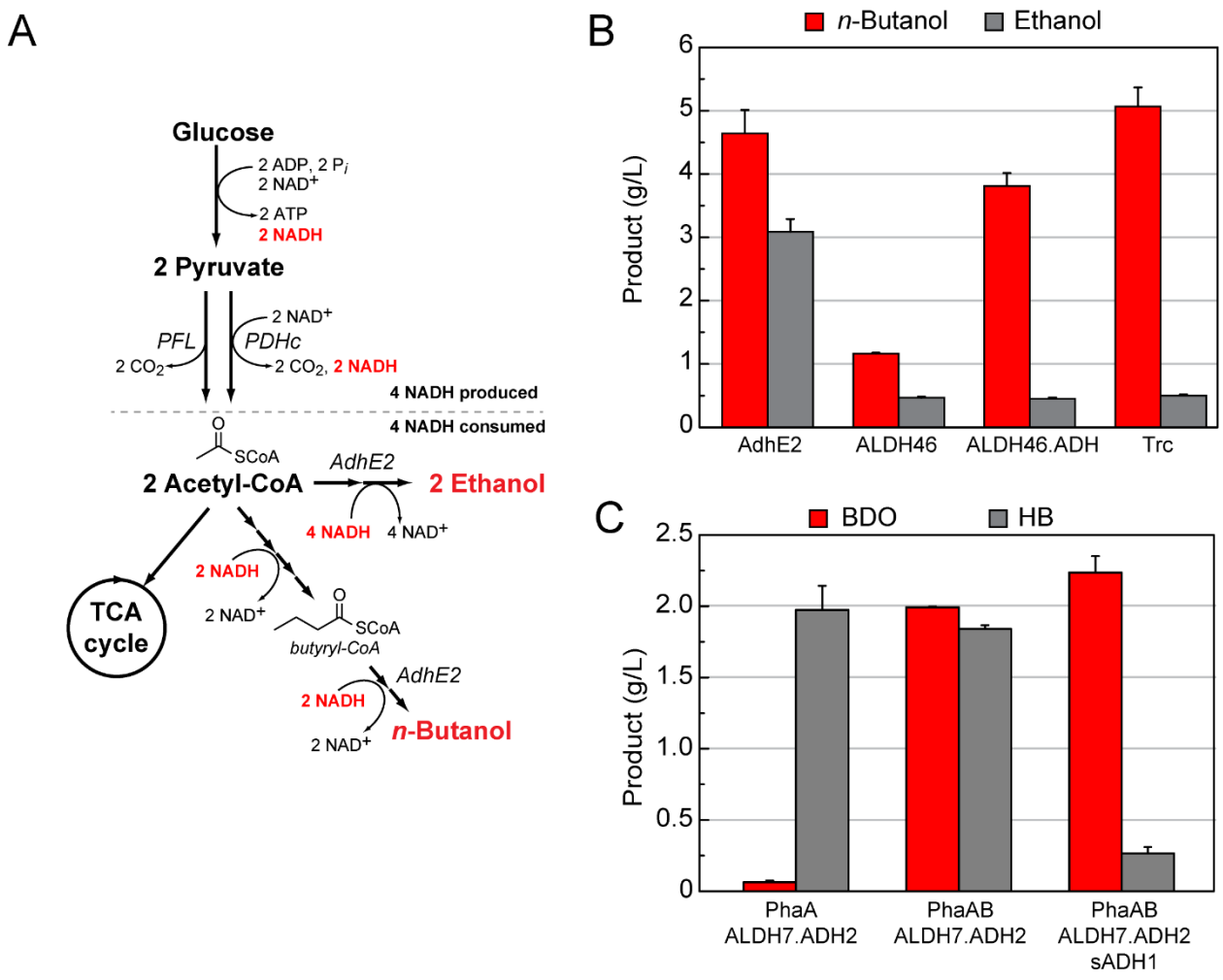


Figure 2.4. Production of C_4 monomer precursors in engineered *E. coli*. (A) Design of a host for the anaerobic production of target compounds from acetyl-CoA. Deletion of the major fermentation pathways of *E. coli* in DH1Δ*5* allows for the synthetic *n*-butanol pathway as the major mechanism for balanced NAD^+ regeneration production via the acetyl-CoA intermediate. However, the promiscuity of AdhE2 towards acetyl-CoA and butyryl-CoA leads to ethanol fermentation as a pathway short circuit that also maintains stoichiometric redox balance. (B) Screening of AdhE, ALDH, and ADH candidates in *E. coli* DH1Δ*5* pBBR1-AceEF.Lpd pBT33-Bu1 pCWori.TdTer-Trc.ALDH.ADH yields a C_4-selective fermentation pathway under anaerobic conditions. When AdhE2 is included, high levels of ethanol produced along with the target *n*-butanol product. Replacement with ALDH46 reduces ethanol production to background levels but concomitantly drops *n*-butanol titers. Addition of the ADH domain from AdhE2 and tuning the promoter for expression allows for high *n*-butanol yields with very little ethanol being formed. All strains were grown in TB with 2.5% (*w/v)* glucose media for 3 d post induction. (C) Screening of ALDH, ADH, and sADH candidates in *E. coli* DH1Δ*5* pBBR1-AceEF.Lpd pBT33-PhaA/PhaAB pCWori-Trc.ALDH.ADH lead to identification of the ALDH7.ADH2 pair for production of HB and BDO under anaerobic conditions. In the absence of PhaB, HB is selectively produced. Addition of PhaB leads to a 1:1 ratio of both products formed. The inclusion of an sADH then allows for HB to be converted to BDO. All strains were grown in TB with 2.5% (*w/v)* glucose media for 3 d post induction.

leveraging the family of C_4-selective monofunctional ALDHs and ADHs. A library of ALDHs and ADHs were identified to improve the substrate specificity of the bifunctional aldehyde and alcohol hydrogenase AdhE2 for the first generation pathway design for the *n*-butanol pathway [*20, 32*]. We were interested in the reduction of the 3-hydroxybutyryl-CoA intermediate on this pathway yields 1,3-butanediol (BDO) as a product (*Figure 1A*). Upon chemical dehydration, BDOs can be used to produce butadiene for synthetic rubber production, which is currently produced from fossil fuel sources at the level of >10 million metric tonnes per year [*33, 34*]. We therefore set out to screen our ALDH and ADH library for potential candidate enzymes to construct a BDO pathway from the reduction of 3-hydroxybutyryl-CoA. In this screen, we achieved a titer of 2.2 g L^{-1} (*Figure 2.4C*) [*32*].

During this analysis, we identified 4-hydroxy-2-butanone (HB) as a side-product that appears to arise from the reduction of an earlier pathway intermediate, acetoacetyl-CoA (*Figure 1A*). HB is also an interesting product as its dehydration produces methyl vinyl ketone, a reagent used in the production of fine chemicals [*35*], as well as a potential monomer unit for polymers [*36*]. We therefore set out to characterize the selectivity of ALDH-ADH pairs by examining the pathway partition between BDO and HB. This screen indicated HB production is highly specific to the ALDH7-ADH2 pair, providing an even distribution of products at high titer (3.4 ± 0.1 g L^{-1}). On the other end, the ALDH3-ADH22 pair was found to capture a large fraction of the C_4 product pool as BDO (81%), producing 2.9 ± 0.1 g L^{-1} of total products under screening conditions.

A selective pathway for production of HB was engineered by simply removing the PhaB ketoreductase from the pathway to eliminate production of 3-hydroxybutyryl-CoA. With this change, the PhaA-ALDH7-ADH2 pathway generates 2.0 ± 0.2 g L^{-1} HB (*Figure 2.4C*). However, the engineering of a selective BDO pathway is more challenging, as acetoacetyl-CoA is a precursor that is required for its production. We then took the approach to utilize a secondary alcohol dehydrogenase (sADH) that would be capable of reducing the acetoacetaldehyde generated from promiscuous activity of the ALDH on acetoacetyl-CoA directly to BDO (*Figure 2.5*). The net result of this pathway would ultimately be BDO, channeled from reduction of either acetoacetaldehyde or 3-hydroxybutyryl-CoA. A number of secondary alcohol dehydrogenases (SADHs) have been reported to reduce 4-hydroxy-2-butanone or similar substrates [*37*]. Several of these SADHs were co-expressed with the ALDH7-ADH2 pair, which consistently produced an even mixture of butanediol and hydroxybutanone. Several of the SADHs enabled a shift in the product profile, producing high levels of BDO (>2 g L^{-1}) within minimal production of HB (<250 mg L^{-1}). With these SADHs in hand, we can control the product profile between HB, BDO, or mixture of the two (*Figure 2.4C*).

Adaptive evolution of C_4 pathways. With a highly specific pathway for *n*-butanol in place, we next set out to develop a genetic selection for increasing titers under anaerobic conditions with the long-term goal of gaining new insight into the manipulation of central carbon homeostasis. In contrast to our results with the promiscuous *n*-butanol pathway containing the ethanol short-circuit (*Figure 2.6*), growth of the fermentation-deficient strain, DH1Δ*5*, depends solely on *n*-butanol production. Using a set of control plasmids with low, medium, and high *n*-butanol productivity, we observe that the ability of the synthetic pathway to rescue of anaerobic growth of DH1Δ*5* correlates directly with product titer and thus its capacity to recycle NADH. Indeed, strains

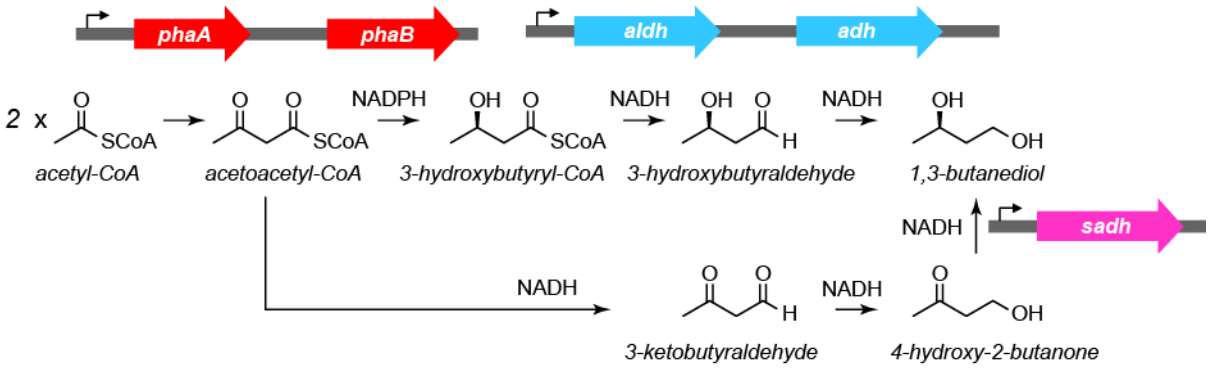


Figure 2.5. Introduction of a sADH to increase BDO selectivity. A strategy for increasing the selectivity of BDO production is to use an sADH to reduce HB to BDO, allowing the products of unselective reduction of acetoacetyl-CoA to be rescued.

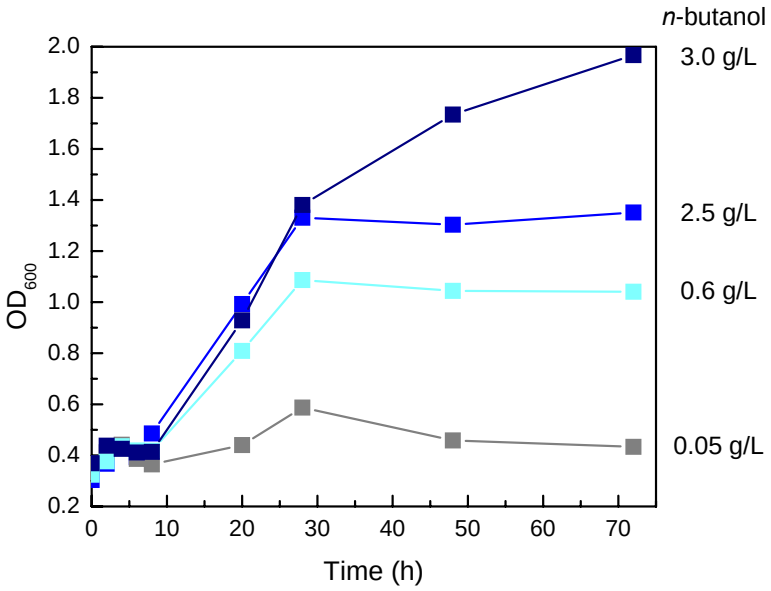
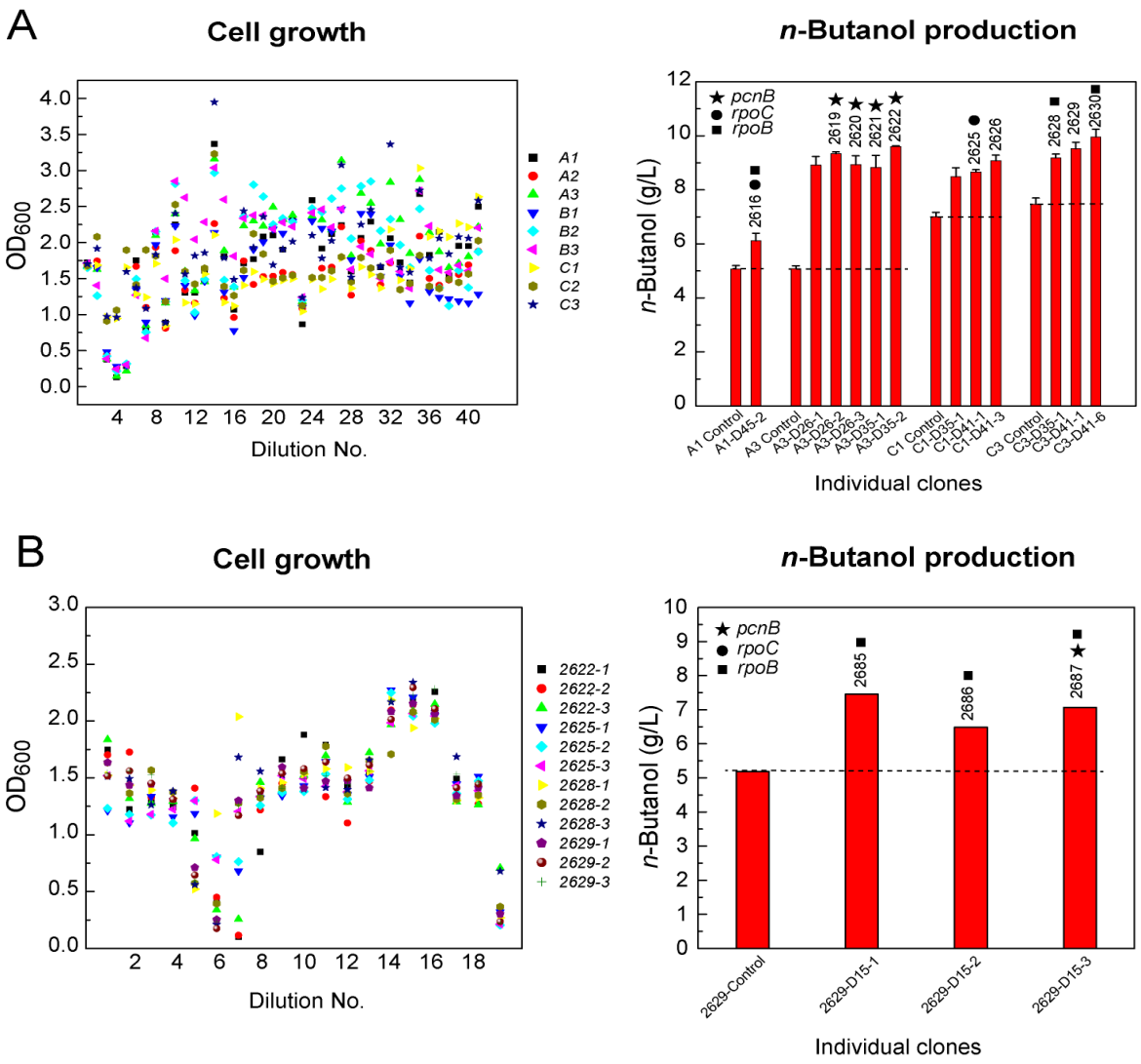
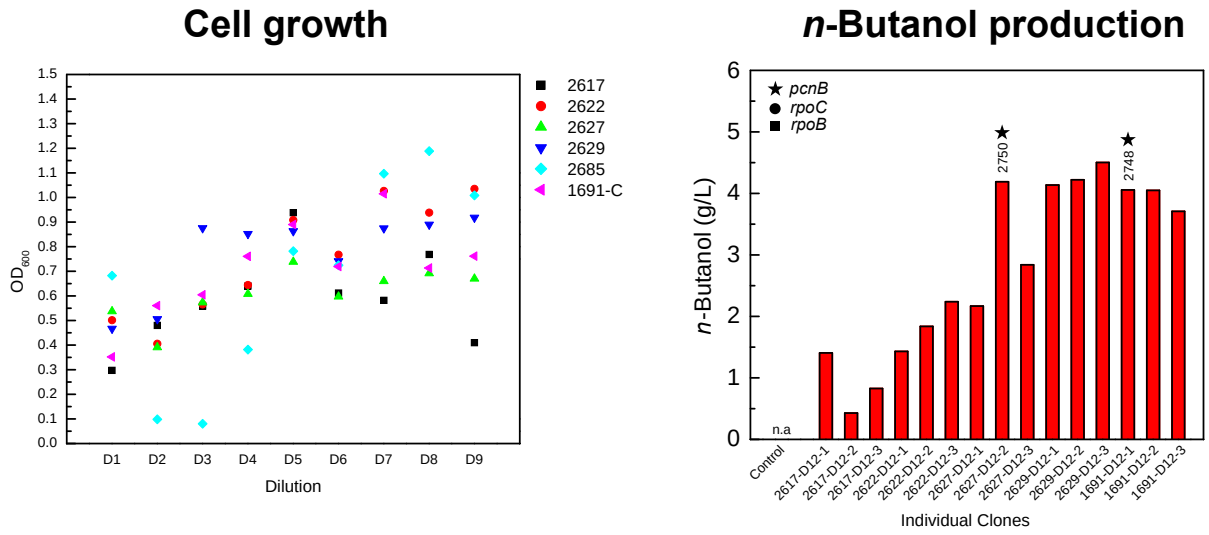


Figure. 2.6. Development of a genetic selection for *n*-butanol production. The *n*-butanol pathway complements the deletion of the native fermentation pathways of E. coli under anaerobic conditions. *n*-Butanol pathway variants displaying a range of yields were transformed into DH1Δ*5* and cultured anaerobically. Growth was monitored by OD_{600} and *n*-butanol production was quantified at the end of the experiment. All strains were grown in TB with 2.5% (*w/v)* glucose media for 3 d post induction.


A
Cell growth
OD600
Dilution No.
A1
A2
A3
B1
B2
B3
C1
C2
C3
n-Butanol production
n-Butanol (g/L)
pcnB
rpoC
rpoB
2616
2619
2620
2621
2622
2625
2626
2628
2629
2630
A1 Control
A1-D45-2
A3 Control
A3-D26-1
A3-D26-2
A3-D26-3
A3-D35-1
A3-D35-2
C1 Control
C1-D35-1
C1-D41-1
C1-D41-3
C3 Control
C3-D35-1
C3-D41-1
C3-D41-6
Individual clones
B
Cell growth
OD600
Dilution No.
2622-1
2622-2
2622-3
2625-1
2625-2
2625-3
2628-1
2628-2
2628-3
2629-1
2629-2
2629-3
n-Butanol production
n-Butanol (g/L)
pcnB
rpoC
rpoB
2685
2686
2687
2629-Control
2629-D15-1
2629-D15-2
2629-D15-3
Individual clones
C
Cell growth
OD600
Dilution No.
A1
A2
A3
B1
B2
B3
C1
C2
C3
n-Butanol production
n-Butanol (g/L)
pcnB
rpoC
rpoB
2726
2727
2728
2729
2730
2731
Individual clones

D



Parent strain	Description	Label
DH1Δ5	Pathway A	2617
DH1Δ5.2622	Isolated Pathway A clone from LB selection (*Figure 2.7A*: A3-D35-2)	2622
DH1Δ5	Pathway C	2627
DH1Δ5.2629	Isolated Pathway C clone from LB selection (*Figure 2.7A*: C3-D41-1)	2629
DH1Δ*5*.2685	Isolated Pathway C clone from M9/LB selection (*Figure 2.7B*: 2629-D15-1)	2685
BW25113Δ*5*	Pathway C	1691-C

Figure 2.7. Characterization of adaptive evolution of *n*-butanol strains under anaerobic conditions. All selections were performed in triplicate with cultures supplemented with 2.5% (*w/v*) glucose. OD_{600} for each flask was measured before every dilution. Production titers were validated in the selection media and controls represent *E. coli* parent strains freshly transformed with the appropriate plasmids. Strain labels indicate plasmids/flask-dilution number-clone number. Numbers above bars correspond to a unique identifier number for the sequenced strain with a shape indicating specific genetic loci mutated. (A) Adaptive evolution with *E. coli* DH1Δ*5* as the host in LB media with three different *n*-butanol pathways. All strains contained the pBBR1-AceEF.Lpd and pT5T33-Bu2 plasmids with different downstream plasmids (A, pCWori.TdTer-trc.ALDH46.ADH2; B, pCWori.TdTer-trc.ALDH46.ADH8; C, pCWori.TdTer-trc.ALDH21.ADH2). (B) Adaptive evolution with *E. coli* DH1Δ*5* as the host in M9 media supplemented with 10% LB (*v/v*). The parent strains for this evolution were derived from the selection in LB media: A35-D35-2 (2622), C1-D41-1 (2625), C3-D35-1 (2628), and C3-D41-1 (2629). (C) Adaptive evolution with *E. coli* BW21153Δ*5* as the host in M9 media supplemented with 10% LB (*v/v*). All strains contained the pBBR1-AceEF.Lpd and pT5T33-Bu2 plasmids with different downstream plasmids (A, pCWori.TdTer-trc.ALDH46.ADH2; B, pCWori.TdTer-trc.ALDH46.ADH8; C, pCWori.TdTer-trc.ALDH21.ADH2). (D) Adaptive evolution with *E. coli* DH1Δ*5* and BW21153Δ*5* as the hosts in M9 media. Star, circle, and square shape above the bar represents mutation in *pcnB*, *rpoC*, and *rpoB* gene respectively.

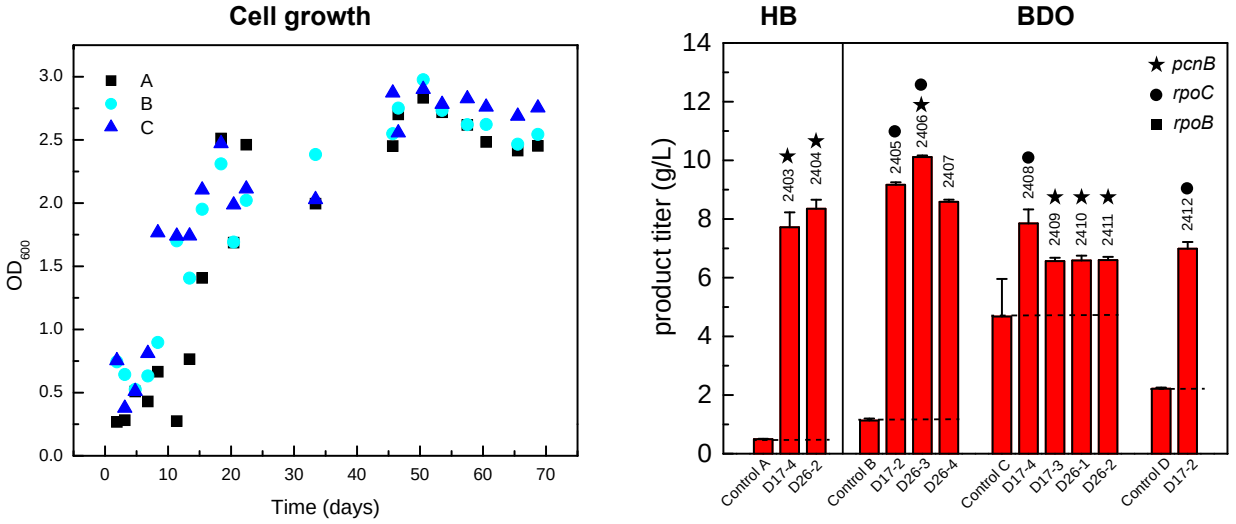


Figure 2.8. Characterization of adaptive evolution of BDO and HB strains under anaerobic conditions. All selections were performed in triplicate with TB cultures supplemented with 2.5% (*w/v*) glucose. OD_{600} for each flask was measured before every dilution. Production was validated in the selection media and controls represent *E. coli* parent strains freshly transformed with the appropriate plasmids. Strain labels indicate plasmids/flask-dilution number-clone number. Numbers above bars correspond to a unique identifier number for the sequenced strain with a shape indicating specific genetic loci mutated. (A) Growth curves of adaptive BDO evolution with (a) DH1Δ*5* pT533-phaA pCWO.trc-ter-aldh7.adh2 pBBR2-PDHc, (b) DH1Δ*5* pT533-phaA.phaB pCWO.trc-ter-aldh7.adh2 pBBR2-PDHc, (c) DH1Δ*5* pT533-phaA.phaB pCWO.trc-ter-sadh1.aldh7.adh2 pBBR2-PDHc. Cultures were grown (B) Control BDO and HB production with plasmids extracted from evolved strains and transformed into a clean parental *E. coli* DH1Δ*5* host. The similar production compared to fresh plasmids indicates that mutations responsible for increasing product titer are likely found on the chromosome. Strain numbers for evolved strains are indicated above each bar in the figure. Star, circle, and square shape above the bar represents mutation in *pcnB*, *rpoC*, and *rpoB* gene respectively.

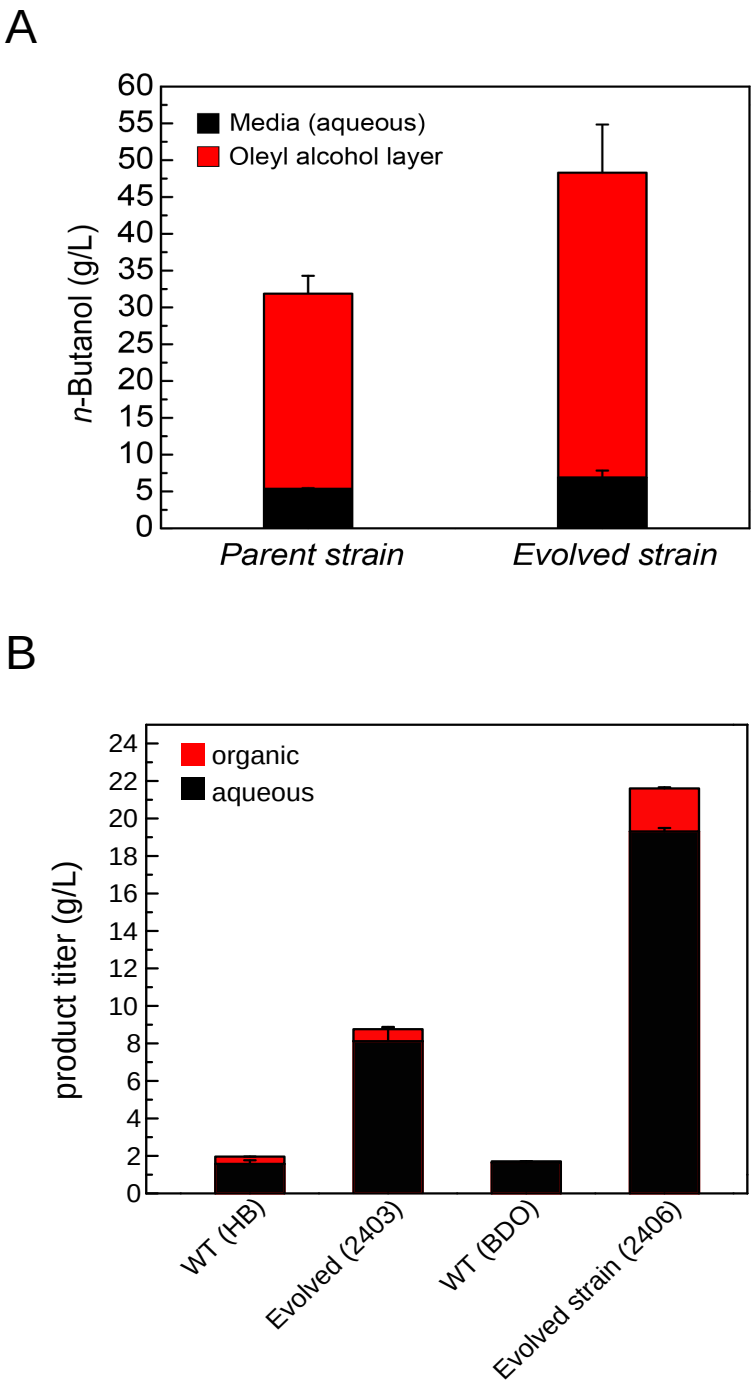


Figure 2.9. Production titers of C_4 monomers compared to parent strains with high glucose loading. Cells were cultured in TB media supplemented with 8% (*w/v*) glucose with a 20% (*v/v*) oleyl alcohol overlay. The theoretical yield for this experiment is 37.6 g L^{-1}. (A) *n*-Butanol production in the parent DH1Δ*5* strain compared to the evolved DH1Δ*5*.2622 strain both bearing the pBBR1-AceEF.Lpd pT5T33-Bu2 pCWOri.TdTer-trc.ALDH46.ADH2 plasmids. Titer indicated that the evolved strain can utilize carbon sources other than glucose to produce *n*-butanol. (B) HB and BDO production in the parent DH1Δ*5* strain compared to the evolved strains (HB – 2403, BDO – 2406).

Product	Parent	Plasmids	Media	Identifier	No.
n-Butanol	DH1Δ*5*	339-499-1866	LB/2.5% Glc	A1-D45-2	2616
	DH1Δ*5*	339-499-1866	LB/2.5% Glc	A3-D26-2	2619
	DH1Δ*5*	339-499-1866	LB/2.5% Glc	A3-D26-3	2620
	DH1Δ*5*	339-499-1866	LB/2.5% Glc	A3-D35-1	2621
	DH1Δ*5*	339-499-1866	LB/2.5% Glc	A3-D35-2	2622
	DH1Δ*5*	339-499-2456	LB/2.5% Glc	C1-D41-1	2625
	DH1Δ*5*	339-499-2456	LB/2.5% Glc	C1-D41-3	2626
	DH1Δ*5*	339-499-2456	LB/2.5% Glc	C3-D35-1	2628
	DH1Δ*5*	339-499-2456	LB/2.5% Glc	C3-D41-1	2629
	DH1Δ*5*	339-499-2456	LB/2.5% Glc	C3-D41-6	2630
	DH1Δ*5*	339-499-2456	M9/10% LB/2.5% Glc	D15-12-1	2685
	DH1Δ*5*	339-499-2456	M9/10% LB/2.5% Glc	D15-12-2	2686
	DH1Δ*5*	339-499-2456	M9/10% LB/2.5% Glc	D15-12-3	2687
	BW25113Δ*5*	339-499-2456	M9/10% LB/2.5% Glc	C1-D4-3	2726
	BW25113Δ*5*	339-499-2456	M9/10% LB/2.5% Glc	D4-C3-3	2727
	BW25113Δ*5*	339-499-1866	M9/10% LB/2.5% Glc	D17-A3-1	2728
	BW25113Δ*5*	339-499-1867	M9/10% LB/2.5% Glc	D17-B3-1	2729
	BW25113Δ*5*	339-499-2456	M9/10% LB/2.5% Glc	D17-C1-3	2730
	BW25113Δ*5*	339-499-2456	M9/10% LB/2.5% Glc	D17-C3-2	2731
	BW25113Δ*5*	339-499-2456	M9/2.5% Glc	C1-D9-1	2748
	DH1Δ5	339-499-2456	M9/2.5% Glc	C1-D11-2	2750
HB	DH1Δ*5*	339-2080-2076	TB/2.5% Glc	A3-D17-4	2403
	DH1Δ*5*	339-2080-2076	TB/2.5% Glc	A3-D26-2	2404
BDO	DH1Δ*5*	339-1319-2076	TB/2.5% Glc	B1-D17-2	2405
	DH1Δ*5*	339-1319-2076	TB/2.5% Glc	B1-D26-3	2406
	DH1Δ*5*	339-1319-2076	TB/2.5% Glc	B3-D26-4	2407
	DH1Δ*5*	339-1319-2430	TB/2.5% Glc	C1-D17-4	2408
	DH1Δ*5*	339-1319-2430	TB/2.5% Glc	C3-D17-3	2409
	DH1Δ*5*	339-1319-2430	TB/2.5% Glc	C2-D26-1	2410
	DH1Δ*5*	339-1319-2430	TB/2.5% Glc	C3-D26-2	2411
	DH1Δ*5*	339-1319-2468	TB/2.5% Glc	D3-D17-2	2412

Table 2.1. Strains isolated from evolutions. All strains contained the pBBR1-AceEF.Lpd plasmid (#339) for overexpression of the pyruvate dehydrogenase complex. Parent strains for n-butanol production contained the pT5T33-Bu2 plasmid (#499) and one of the following three plasmids: pCWori.TdTer-trc.ALDH46.ADH2 (#1866), pCWori.TdTer-trc.ALDH46.ADH8 (#1867), or pCWori.TdTer-trc.ALDH21.ADH2 (#2456). The parent strain for HB production contained the pT533-PhaA (#2080) and pCWori-trc.ALDH7.ADH2 (#2076) plasmids. The parent strain for BDO production contained the pT533-PhaAB (#1319) plasmid and one of the following three plasmids: pCWori-trc.ALDH7.ADH2 (#2076), pCWori-trc.ALDH3.ADH22 (#2468), or pCWori.sADH1-trc.ALDH7.ADH2 (#2430). An identifier number was used during isolation of individual clones from an evolution experiment consisting of plasmid combination (A, B, C, D), flask number-dilution number-clone.

complemented with a very low-flux pathway variants do not grow significantly, if at all, while strains complemented with robust pathway variants are indistinguishable from wild-type [*32*].

In order to select for variants with improved *n*-butanol productivity under anaerobic conditions, we turned to adaptive evolution. In this approach, the natural mutation frequency is utilized, which requires longer evolution times but selects for more advantageous mutations and minimizes the occurrence of neutral mutations[*38, 39*]. Since every evolutionary trajectory has the potential to yield different results, we evolved two different host strains, DH1Δ*5* and BW25113Δ*5*, using media ranging in richness from M9, 10% (*v*/*v*) LB in M9, and LB, by diluting the culture every 24 h from 4 days to 70 days (*Figure 2.7*, *Table 2.1*). Using this approach, we were able to evolve strains six-fold from 11% to 66% carbon conversion as well as from 43% to >95% yield under these various conditions (*Figure 2.7*). Although the redox balance is not stoichiometric as it is with *n*-butanol, we were also able to evolve BDO and HB production in DH1Δ*5* from 20% to ~95% theoretical carbon conversion in TB (*Figure 2.8*). Furthermore, scaled-up growth of these strains in shake flasks yielded high titers (31 ± 2 to 47 ± 6 g L^{-1}) and yields (>95%) of all three products (*Figure 2.9*). Taken together, the evolved strains demonstrate large shifts in central carbon metabolism, allowing for the robust production of a range of C_4 products from acetyl-CoA under anaerobic conditions.

Identifying two key players in transcriptional re-programming. We took a genome scale approach to explore key factors responsible for the evolution of this large shift in central carbon flow. A total of 31 isolated strains from three independent selections for *n*-butanol (21 strains), BDO (8 strains), and HB (2 strains) production carried out under different growth conditions were sequenced to identify the changes between the genomes of the parental strains and evolved strains. Interestingly, we found mutations only in a handful of genes, which consistently appeared regardless of selection conditions (*Table 2.2 and Appendix 2.8*). In addition, a few mutations mapped to the non-coding portions of the genome (0-1 mutations per strain with a total number of six distinct mutations from all 31 strains that were sequenced) along with rearrangements that appeared to be mostly associated with mobile elements. Of the mutations in coding regions, the most striking is the finding that polynucleotide adenyltransferase (*pcnB*) and/or the RNA polymerase ββ' subunits (*rpoBC*) were found to be mutated in nearly all of the most successful evolved strains. These two gene loci are involved in regulating the transcriptional landscape of the cell by forming part of the transcription complex (*rpoBC*) [*40, 41*] as well as by controlling the lifetime of mRNAs by polyadenylation (*pcnB*) [*42*]. Mutations in ribonuclease E (*rne*) also occurred frequently (12%) in the evolved *n*-butanol hosts.

The discovery that genes involved in RNA metabolism appear to drive metabolic network evolution led us to the hypothesis that the phenotypic changes were being controlled in large part by alterations in the global transcriptional program. This model is consistent with measurements of pathway enzyme activity that showed no significant different between a parent and evolved strain, suggesting that yield increases were not derived from simple overexpression of heterologous pathway genes measured by enzymatic assays (*Figure 2.11*). To further characterize this phenomenon, we performed an RNA-Seq experiment on the evolved BDO strain with the largest improvement in production titer (DH1Δ*5*.2406) containing point mutations in *pcnB* and

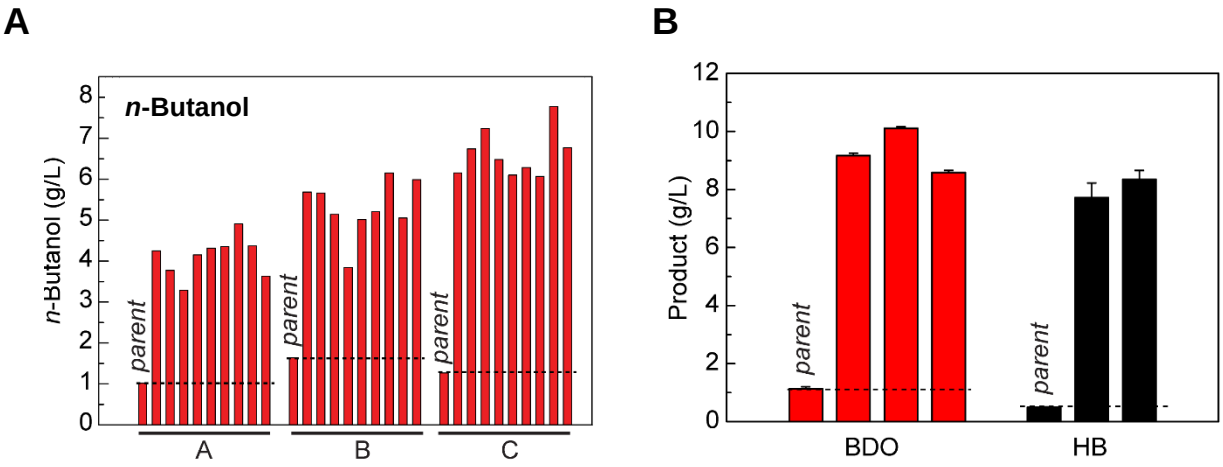
A

Product	Gene	Codon change	Amino acid change	Strain #
4-hydroxy-2-butanone	*pcnB*	GGC → GCC	G141A	2403, 2404
1,3-butanediol	*pcnB*	CGC → CTC	R149L	2406
		CGC → CAC	R149H	2409
		CCT → ACT	P78T	2410
		TTG → TGG	L208W	2411
	rpoC	ATG → CTG	M466L	2405, 2406, 2408
		Δ ACCAAGCGTAAAAAGCTG (634 - 651 nt)	Δ TKRKKL (212 - 217)	2412
	rsmB	CAA → AAA	Q314K	2409
	pyrG	GAT → GAA	D42E	2411
	pspE	TCA → CCA	S14P	
	dcuA	CAG → CCG	Q64P	
	pnp	Δ GGCGATATCTCTGAGTTCGCACCGCGT (1636-1662 nt)	Δ GDISEFAPR (546- 554)	2407
n-butanol	*pcnB*	GAT → GAG	D194E	2619, 2620, 2621, 2622
		GCT → ACT	A98T	2687
		CGC → CCC	R149P	2750
		GAA → GCA	E108A	2748
		AAC → CAC	N138H	2726
		Δ G (1176 nt)	Frame shift after D391	2728
	rpoC	GGT → CGT	G 1161 R	2616
		AAA → GAA	K1192E	2625
	rpoB	GAC → GCC	D199A	2616
		GGC → GTC	G467V	2628, 2630, 2685, 2686, 2687
	rne	CGT → AGT	R373S	2626
		AAA → AAC	K255N	2685
		CGC → CTC	R109L	2730
		CGC → CAC	R488H	2731
	lysP	GTT → GCT	V276A	2685, 2686, 2687
	pnp	ATC → AAC	I541N	2686
	gluQ	add ACG (887 nt)	add S298	2727
	cadB	TGA → AGA	stop 41 R (pseudogene)	2616, 2630, 2685, 2686, 2687

B

Product	Gene	Description	Mutation	Annotation	Position	Strain #
4-hydroxy-2-butanone	ECDH1_10830 (bottom) / ECDH1_RS10835 (t	hypothetical protein /nad(p) transhydrogenase subunit alpha	Added GGT	intergenic (-38 / -486)	2,200,089	2403, 2404
1,3-butanediol	ECDH1_RS07795 (bottom) / ECDH1_RS0780	hypothetical protein / nucleoid-associated protein	(T)8 to 9	intergenic (-59 / -123)	1,592,789	2410
n-butanol	ECDH1_RS10460 (top) / ECDH1_RS10465 (tc	hypthetical proteion / 4Fe-4S ferredoxin	C to T	intergenic (+146 / -309)	2,123,692	2625
	ECDH1_RS21465 (bottom) / rrf	UDP-N-acetylenolpyruvoyglucosamine reductase / 5S ribomsal RNA	C to T	intergenic (-276 / +27)	4,342,689	2625
	rrf (bottom) / ECDH1_RS21275 (bottom)	5S ribosomal RNA / 23S ribosomal RNA	delta 1 bp	intergenic (-70 / +8)	4,301,498	2630
	BW25113_RS00715 (bottom) /BW25113_RS0	polynucleotide adenyltransferase pcnB / tRNA glutamyl-Q(34) synthetase GluQRS	C to T	intergenic (-43 / +50)	155,623	2729

Table 2.2. Genome sequencing of evolved strains. 31 different evolved strains were sequenced along with the DH1Δ5 or BW25115Δ5 parent strain. Reads were mapped to the reference genome of DH1 or BW25113 and analyzed for changes including SNPs as well as rearrangements using Breseq. (A) Predicted point mutations in the coding region. (B) Predicted point mutations in intergenic regions.



C

Gene	Annotation	Pathways
pcnB	Poly(A) polymerase	*n*-Butanol, BDO, HB
rpoB	RNA polymerase β subunit	*n*-Butanol
rpoC	RNA polymerase β´ subunit	*n*-Butanol, BDO
rne	Ribonuclease E	*n*-Butanol
pnp	Polyribonucleotide nucleotidyltransferase	*n*-Butanol, BDO
cadB	Cadaverine/Lys antiporter	*n*-Butanol

Figure 2.10. High C4 monomer producing strains were isolated from adaptive evolution. (A) A representative adaptive evolution for *n*-butanol production. *E. coli* BW25113Δ*5* pBBR1-AceEF.Lpd pT5T33-Bu2 containing either pCWori.TdTer-trc.ALDH46.ADH2 (A), pCWori.TdTer-trc.ALDH46.ADH8 (B), or pCWori.TdTer-trc.ALDH21.ADH2 (C) was subjected to multiple round of dilution in M9 containing 10% (v/v) LB and 2.5% (w/v) glucose under anaerobic conditions. Individual clones were then isolated and characterized for their *n*-butanol titers compared to the parent strain. (B) Characterization of BDO and HB strains after adaptive evolution. *E. coli* DH1Δ*5* pBBR1-AceEF.Lpd pT533-phaA.phaB pCWO.trc-TdTer-aldh7.adh2 (BDO), *E. coli* DH1Δ*5* pBBR1-AceEF.Lpd pT533-phaA pCWO.trc-TdTer-aldh7.adh2 (HB). (C) List of genes found mutated in more than one evolved strain for either *n*-butanol, BDO, or HB.

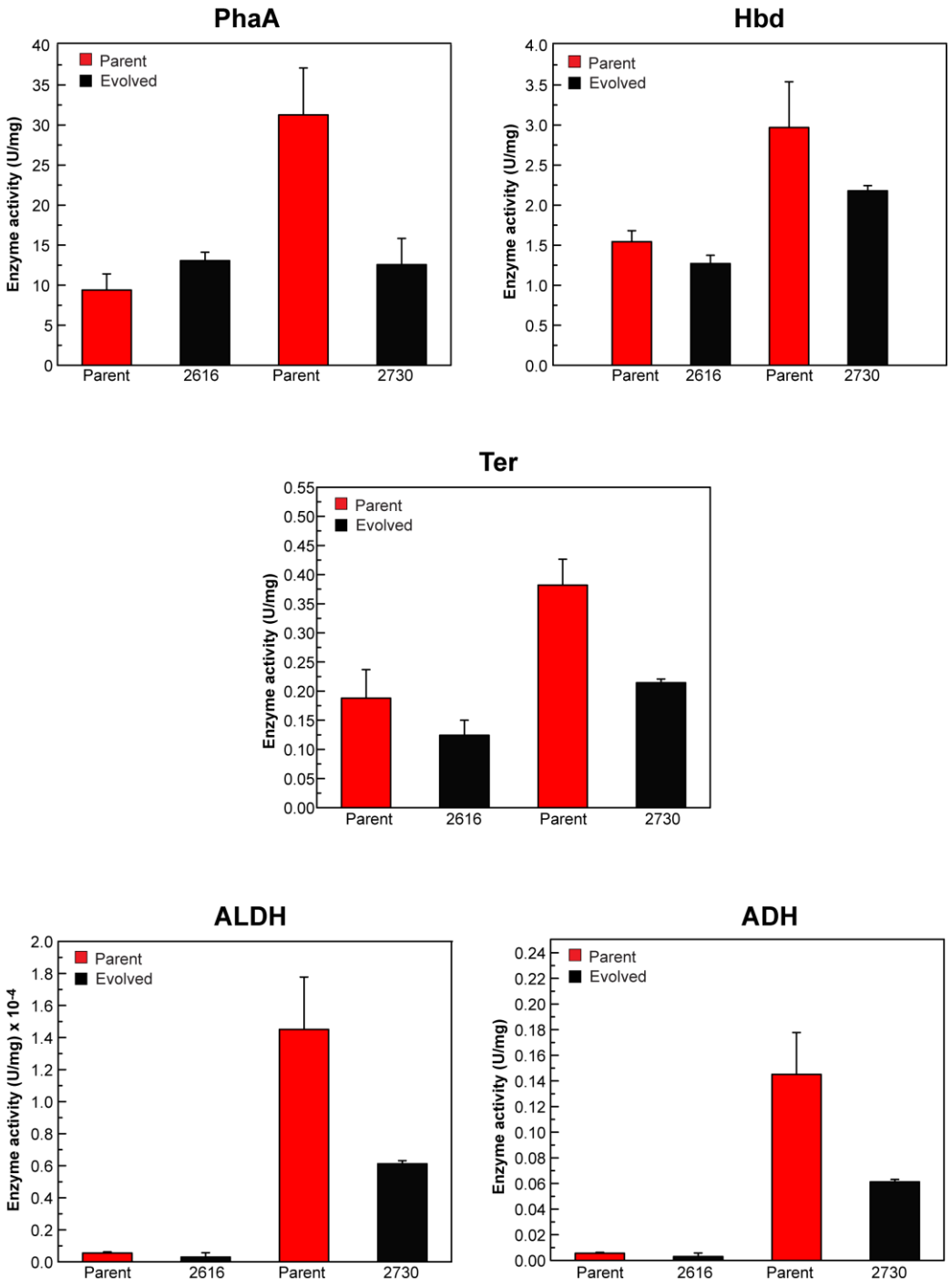
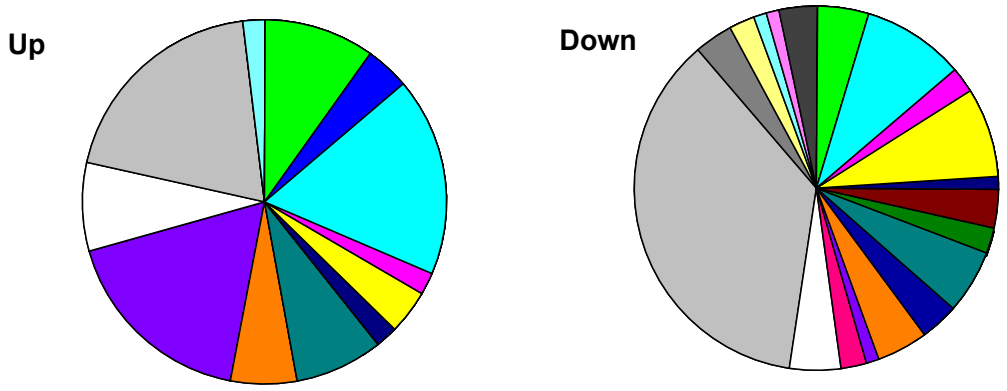


Figure 2.11. Cell lysate enzyme activities of *n*-butanol pathway enzymes for parent and evolved strains. Enzyme activities were measured in cell lysate of parent and evolved strains to examine whether increased heterologous expression of pathway enzymes could the source for increases in *n*-butanol titer. DH1Δ*5*.2616 was compared to its parent, DH1Δ*5* pBBR1-AceEF.Lpd pT5T33-Bu2 pCWori.TdTer-trc.ALDH46.ADH2. BW25113Δ*5*.2730 was compared to its parent, BW25113Δ*5* pBBR1-AceEF.Lpd pT5T33-Bu2 pCWori.TdTer-trc.ALDH21.ADH2. There is no significant activity differences for the four enzymes tested between the parent and evolved strains, leading us to conclude that differential pathway enzyme expression is not a major factor.



	Function	Genome	Up	Down
A	RNA processing and modification	2 (0.06)	0 (0)	0 (0)
C	Energy production and conversion	260 (7.72)	5 (10.64)	4 (5.06)
D	Cell cycle control, cell division, chromosome partitioning	38 (1.13)	2 (4.26)	0 (0)
E	Amino acid transport and metabolism	354 (10.51)	9 (19.15)	8 (10.13)
F	Nucleotide transport and metabolism	106 (3.15)	1 (2.13)	2 (2.53)
G	Carbohydrate transport and metabolism	380 (11.28)	2 (4.26)	7 (8.86)
H	Coenzyme transport and metabolism	179 (5.31)	0 (0)	0 (0)
I	Lipid transport and metabolism	123 (3.65)	1 (2.13)	1 (1.27)
J	Translation, ribosomal structure and biogenesis	227 (6.74)	0.00	0.00
K	Transcription	292 (8.67)	0.00	3 (3.80)
L	Replication, recombination and repair	137 (4.07)	0.00	2 (2.53)
M	Cell wall/membrane/envelope biogenesis	240 (7.12)	4 (8.51)	5 (6.33)
N	Cell motility	106 (3.15)	0 (0)	3 (3.80)
O	Posttranslational modification, protein turnover, chaperones	149 (4.42)	3 (6.38)	4 (5.06)
P	Inorganic ion transport and metabolism	207 (6.14)	9 (19.15)	1 (1.27)
Q	Secondary metabolites biosynthesis, transport and catabolism	57 (1.69)	0 (0)	2 (2.53)
R	General function prediction only	262 (7.78)	4 (8.51)	4 (5.06)
S	Function unknown	203 (6.03)	10 (21.28)	32 (40.51)
T	Signal transduction mechanisms	189 (5.61)	0 (0)	3 (3.80)
U	Intracellular trafficking, secretion, and vesicular transport	53 (1.57)	0 (0)	2 (2.53)
V	Defense mechanisms	88 (2.61)	1 (2.13)	1 (1.27)
W	Extracellular structures	32 (0.95)	0 (0)	1 (1.27)
X	Mobilome: prophages, transposons	60 (1.78)	0 (0)	3 (3.80)
	Total	**3369**	**47**	**79**

Figure 2.12. RNA-Seq profile of evolved BDO producing strain. Clusters of orthologous groups (COG) categories for genes differentially expressed between the parent (DH1Δ*5* pT533-PhaAB pCWori-trc.ALDH7.ADH2, pBBR2-aceE.F.lpd(WT)) and evolved BDO strain (DH1Δ*5*.2406). COG categories were identified by the IMG-ER annotation pipeline. COG categories represented by genes that are upregulated and downregulated 24 h after induction with IPTG. Comparison of COG category representation in the differentially expressed genes compared to the entire genome. The number of the open reading frames represented by each COG is given, and the percentage of total genes with COG categories is in parentheses. Since some genes fall into multiple COG categories, the percentage was calculated by dividing the total number of unique genes.

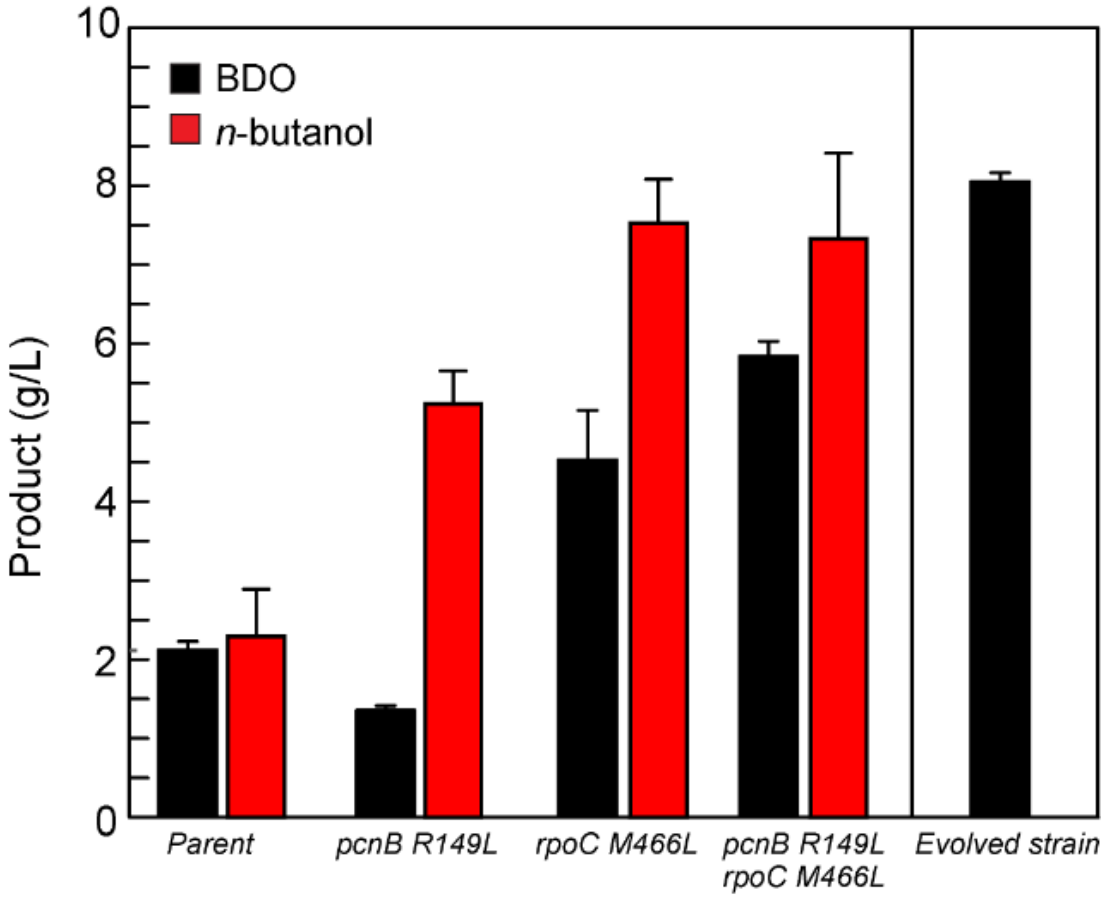


Figure 2.13. Validating mutations that arose from evolved strain. Generating the *pcnB* and *rpoC* mutations found in DH1Δ*5*.2406 in a clean genetic background (DH1Δ*5* parent) captures the majority of the improvement observed in the evolved strain, indicating that these two gene loci play an important role in enabling the increases in BDO production. Introduction of the *n*-butanol pathway into DH1Δ*5*.*pcnB*(R149L).*rpoC*(M366L) shows that some aspects of this phenotype can be transferred to other pathways.

rpoC. We found 126 differentially-expressed genes (β value > 2) between the parental and evolved strain (*Figure 2.12*), indicating that alterations in acetyl-CoA and central carbon homeostasis may require changes at many metabolic nodes. These genes fall into a broad range of categories, with the highest number assigned to energy production and conversion, amino acid transport and metabolism, cell envelope biogenesis, and carbohydrate transport and metabolism (*Figure 2.12*).

In order to validate the impact of the *pcnB* and *rpoC* mutations, the two mutations observed in this BDO strain (*pcnB* R149L/ *rpoC* M466L) were introduced into a clean genetic background. These experiments show that the mutations in *rpoC* and *pcnB* are synergistic, as both are required to achieve a substantive increase in BDO titer compared to the parent (*Figure 2.13*). Indeed, the double mutant demonstrations a 2.75-fold increase in BDO titers (parent, 2.1 ± 0.1 g L^{-1} ; DH1Δ*5*.2406, 5.8 ± 0.2 g L^{-1}) , which recapitulates 73% of the improvement observed in the fully evolved strain (8.1 ± 0.1 g L^{-1}) . We were also interested in the generality of these mutations and thus tested their ability to stimulate yield increases in a different pathway. When the *n*-butanol pathway is introduced into the double mutant, we observe a 3.2-fold increase in product titer from 2.3 ± 0.6 to 7.3 ± 1.1 g L^{-1}. (*Figure 2.13*). Altogether, these data show mutations in only two genes, *pcnB* and *rpoC*, can drive a large shift in central carbon metabolism that can be generalized to related pathways utilizing the acetyl-CoA building block.

2.4. Conclusions

Central carbon metabolism represents a key regulator and read-out of cellular state, both controlling and reporting on cell physiology [*43*]. Given its essential role in cell fitness and survival, these pathways are subject to tight homeostasis with multiple mechanisms to ensure robustness and reduce sensitivity to change [*44*]. As such, rational engineering of central carbon pathways for the purpose of re-routing flux to a synthetic product can be quite challenging as it opposes the cell's evolutionary impetus to direct carbon to growth or biomass. On the other hand, engineered pathways provide an interesting platform where product titer can be treated as a synthetic phenotype or marker for quantitative assessment of genetic traits. As such, they have the potential to identify and characterize factors that require complex changes at multiple nodes in the regulatory and metabolic network [*45*].

In this work, we have developed a genetic selection for the production of three different industrially relevant monomer precursors to 1-butene (*n*-butanol), 1,3-butadiene (BDO), and methyl vinyl ketone (HB). This selection probes a fundamental switch in central carbon pathway usage by requiring increased availability of key building block acetyl-CoA under anaerobic conditions, where it is not made at high levels because of low cell growth rates. Since anaerobic conversion of pyruvate to acetyl-CoA represents a differentiation of carbon away from ATP synthesis via fermentation towards wasteful growth pathways, homeostasis is strongly established at this node and not altered with the knockout of all the major fermentation pathways of the host (DH1Δ*5*). However, strains could be identified with up to 5-fold improvements in yield and near quantitative production using a design in which fitness is driven by the yield of products synthesized from acetyl-CoA.

Genome-level characterization of these strains revealed the surprising discovery that mutations in only two gene loci, *pcnB* and *rpoBC*, were sufficient to enable large shifts in carbon flow. Physiological studies indicate this effect relies on remodeling the transcriptome by influencing

RNA metabolism. Interestingly, a wide range of mutations were identified within these three genes, some of which have been found to important for activity in biochemical studies [*42*]. Furthermore, it was found that mutations found in the evolved BDO strain could be translated to significant increases in *n*-butanol yields, indicating that these strains could serve as a shared platform for production of a wide range of acetyl-CoA products such as fatty acids, polyketides, and isoprenoids.

In conclusion, living systems offer a unique advantage for chemical synthesis to increase product yields through evolution. By using evolution to solve difficult design challenges, we can also take advantage of synthetic pathways to identify new strategies to alter behaviors that are hard-wired into the systems-level behavior of the host.

2.5. References

1. Liao, J. C.; Mi, L.; Pontrelli, S.; Luo, S. Fuelling the Future: Microbial Engineering for the Production of Sustainable Biofuels. *Nat Rev Microbiol* **2016**, *14*, 288–304.
2. Nielsen, J.; Keasling, J. D. Engineering Cellular Metabolism. *Cell* **2016**, *164*, 1185–1197.
3. Paddon, C. J.; Keasling, J. D. Semi-Synthetic Artemisinin: A Model for the Use of Synthetic Biology in Pharmaceutical Development. *Nature Reviews Microbiology*. 2014, pp 355–367.
4. Stephanopoulos, G. Synthetic Biology and Metabolic Engineering. *ACS Synth. Biol.* **2012**, *1*, 514–525.
5. Lee, J. W.; Na, D.; Park, J. M.; Lee, J.; Choi, S.; Lee, S. Y. Systems Metabolic Engineering of Microorganisms for Natural and Non-Natural Chemicals. *Nature Chemical Biology*. 2012, pp 536–546.
6. Jeong, H.; Tombor, B.; Albert, R.; Oltval, Z. N.; Barabásl, A. L. The Large-Scale Organization of Metabolic Networks. *Nature* **2000**, *407*, 651–654.
7. Yamada, T.; Bork, P. Evolution of Biomolecular Networks Lessons from Metabolic and Protein Interactions. *Nature Reviews Molecular Cell Biology*. 2009, pp 791–803.
8. Bordbar, A.; Monk, J. M.; King, Z. A.; Palsson, B. O. Constraint-Based Models Predict Metabolic and Associated Cellular Functions. *Nature Reviews Genetics*. 2014, pp 107–120.
9. Meadows, A. L.; Hawkins, K. M.; Tsegaye, Y.; Antipov, E.; Kim, Y.; Raetz, L.; Dahl, R. H.; Tai, A.; Mahatdejkul-Meadows, T.; Xu, L.; et al. Rewriting Yeast Central Carbon Metabolism for Industrial Isoprenoid Production. *Nature* **2016**, *537*, 694–697.
10. Bogorad, I. W.; Lin, T. S.; Liao, J. C. Synthetic Non-Oxidative Glycolysis Enables Complete Carbon Conservation. *Nature* **2013**, *502*, 693–697.
11. Xu, P.; Qiao, K.; Ahn, W. S.; Stephanopoulos, G. Engineering *Yarrowia Lipolytica* as a Platform for Synthesis of Drop-in Transportation Fuels and Oleochemicals. *Proc Natl Acad Sci U S A* **2016**, *113*, 10848–10853.
12. Chubukov, V.; Mukhopadhyay, A.; Petzold, C. J.; Keasling, J. D.; Martín, H. G. Synthetic and Systems Biology for Microbial Production of Commodity Chemicals. *npj Syst. Biol. Appl.* **2016**, *2*, 16009.
13. Jantama, K.; Haupt, M. J.; Svoronos, S. A.; Zhang, X.; Moore, J. C.; Shanmugam, K. T.; Ingram, L. O. Combining Metabolic Engineering and Metabolic Evolution to Develop Nonrecombinant Strains of *Escherichia Coli* C That Produce Succinate and Malate. *Biotechnol. Bioeng.* **2008**, *99*, 1140–1153.
14. María, P. D. de. *Industrial Biorenewables: A Practical Viewpoint*, First.; Wiley, 2016.
15. Morrow, N. L. The Industrial Production and Use of 1,3-Butadiene. *Environ. Health Perspect.* **1990**, *86*, 7–8.
16. Madeira, L. M.; Portela, M. F. Catalytic Oxidative Dehydrogenation of n -Butane. *Catal. Rev.* **2002**, *44*, 247–286.
17. Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6*, 343–345.
18. Deatherage, D. E.; Barrick, J. E. Identification of Mutations in Laboratory-Evolved Microbes from next-Generation Sequencing Data Using Breseq. *Methods Mol. Biol.* **2014**, *1151* (Chapter 12), 165–188.
19. Davis, J. T.; Moore, R. N.; Imperiali, B.; Pratt, A. J.; Kobayashi, K.; Masamune, S.; Sinskey, A. J.; Walsh, C. T.; Fukui, T.; Tomita, K. Biosynthetic Thiolase from *Zoogloea Ramigera*. I. Preliminary Characterization and Analysis of Proton Transfer Reaction. *J. Biol. Chem.* **1987**, *262*, 82–89.

20. Bond-Watts, B. B.; Bellerose, R. J.; Chang, M. C. Enzyme Mechanism as a Kinetic Control Element for Designing Synthetic Biofuel Pathways. *Nat Chem Biol* **2011**, *7*, 222–227.
21. Bray, N. L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-Optimal Probabilistic RNA-Seq Quantification. *Nat. Biotechnol.* **2016**, *34*, 525–527.
22. Pimentel, H.; Bray, N. L.; Puente, S.; Melsted, P.; Pachter, L. Differential Analysis of RNA-Seq Incorporating Quantification Uncertainty. *Nat. Methods* **2017**, *14*, 687–690.
23. Markowitz, V. M.; Chen, I.-M. A.; Palaniappan, K.; Chu, K.; Szeto, E.; Grechkin, Y.; Ratner, A.; Jacob, B.; Huang, J.; Williams, P.; et al. IMG: The Integrated Microbial Genomes Database and Comparative Analysis System. *Nucleic Acids Res.* **2012**, *40* (Database issue), D115-22.
24. Jiang, W.; Bikard, D.; Cox, D.; Zhang, F.; Marraffini, L. A. RNA-Guided Editing of Bacterial Genomes Using CRISPR-Cas Systems. *Nat. Biotechnol.* **2013**, *31*, 233–239.
25. Chen, W.; Zhang, Y.; Yeo, W.-S.; Bae, T.; Ji, Q. Rapid and Efficient Genome Editing in *Staphylococcus Aureus* by Using an Engineered CRISPR/Cas9 System. *J. Am. Chem. Soc.* **2017**, *139*, 3790–3795.
26. Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*, Sixth.; W. H. Freeman: New York, NY, 2012.
27. van Dijken, J. P.; Scheffers, W. A. Redox Balances in the Metabolism of Sugars by Yeasts. *FEMS Microbiol. Lett.* **1986**, *32*, 199–224.
28. Humphrey, A. E.; Lee, S. E. *Industrial Fermentation: Principles, Processes, and Products. In: Kent J.A. (Eds) Riegel's Handbook of Industrial Chemistry*; Springer, Dordrecht, 1992.
29. Shi, L.; Tu, B. P. Acetyl-CoA and the Regulation of Metabolism: Mechanisms and Consequences. *Curr Opin Cell Biol* **2015**, *33*, 125–131.
30. Galdieri, L.; Zhang, T.; Rogerson, D.; Lleshi, R.; Vancura, A. Protein Acetylation and Acetyl Coenzyme a Metabolism in Budding Yeast. *Eukaryot Cell* **2014**, *13*, 1472–1483.
31. Clark, D. P. The Fermentation Pathways of *Escherichia Coli*. *FEMS Microbiol. Rev.* **1989**, *5*, 223–234.
32. Davis, Matthew A. ; Chang, M. C. Y. Exploring in Vivo Biochemistry with C_4 Fuel and Commodity Chemical Pathways, UC Berkeley, 2015.
33. Hillmyer D. K., M. A. S. 50th Anniversary Perspective: There Is a Great Future in Sustainable Polymers. *Macromolecules* **2017**, *50*, 3733–3749.
34. Global 1-Butene Market By Application (LLDPE, HDPE, Valeraldehyde, and Others), By Region (North America, Europe, Asia-Pacific, and Rest of the World) - Demand-Supply, Price, Analysis and Forecast To 2021 http://www.micromarketmonitor.com/market-report/1-butene-reports-6686141708.html (accessed Dec 27, 2017).
35. Production of Methyl-Vinyl Ketone from Levulinic Acid. **2010**.
36. Marvel, C. S.; Levesque, C. L. The Structure of Vinyl Polymers: The Polymer from Methyl Vinyl Ketone. *J. Am. Chem. Soc.* **1938**, *60*, 280–284.
37. Matsuyama, A.; Kobayashi, Y. Microbial Production of Optically Active 1, 3-Butanediol from 4-Hydroxy-2-Butanone. *Biosci. Biotechnol. Biochem.* **1993**.
38. Dragosits, M.; Mattanovich, D. Adaptive Laboratory Evolution -- Principles and Applications for Biotechnology. *Microb. Cell Fact.* **2013**, *12*, 64.
39. Winkler, J. D.; Kao, K. C. Recent Advances in the Evolutionary Engineering of Industrial Biocatalysts. *Genomics* **2014**, *104*, 406–411.
40. Zuo, Y.; Steitz, T. A.; Zuo, Y.; Steitz, T. A. Crystal Structures of the *E. Coli* Transcription Initiation Complexes with a Complete Bubble. *Mol. Cell* **2015**, *58*, 534–540.
41. Murakami, K. S. X-Ray Crystal Structure of *Escherichia Coli* RNA Polymerase. 70. **2013**, *288*, 9126–9134.

42. Toh, Y.; Takeshita, D.; Nagaike, T.; Numata, T.; Tomita, K. Mechanism for the Alteration of the Substrate Specificities of Template-Independent RNA Polymerases. *Structure* **2011**, *19*, 232–243.
43. Metallo, C. M.; Vander Heiden, M. G. Understanding Metabolic Regulation and Its Influence on Cell Physiology. *Molecular Cell*. 2013, pp 388–398.
44. Cairns, R.; Harris, I.; Mak, T. Regulation of Cancer Cell Metabolism. *Nat. Rev. Cancer* **2011**, *11*, 85–95.
45. Alper, H.; Moxley, J.; Nevoigt, E.; Fink, G. R.; Stephanopoulos, G. Engineering Yeast Transcription Machinery for Improved Ethanol Tolerance and Production. *Science.* **2006**, *314*, 1565–1568.

Chapter 3. *Characterizing the systems-level changes in Escherichia coli strains evolved for C_4 monomer production*

Portions of this work were performed in collaboration with the following:

Charles Berdan from Dan Nomura lab assisted with the metabolomics studies. Genome mutations construction and production experiments were performed in close collaboration with Dr. Hongjun Dong.

3.1. Introduction

Living organisms provide enormous synthetic potential for the production of molecules of interest from renewable feedstocks such as glucose. However, the targeted engineering of the complex coordinated diverse metabolic network of cells presents significant challenges as high product yields compete against cell growth. Much effort has been made to rewire central carbon flux using a broad range of approaches to rationally control flux to the target pathway while eliminating competing processes[*1–3*]. These include traditional protein and strain engineering and have more recently expanded to approaches that incorporate dynamic self-regulation, organelle compartmentalization, and cellular morphology engineering [*4, 5*]. Recently, we have developed a genetic selection to achieve high carbon flux to three of our engineered synthetic pathways for the production of C_4 monomers from acetyl-CoA through adaptive evolution (*Chapter 2*). We have successfully evolved and isolated strains carrying the *n*-butanol, butanediol, and hydroxybutanone pathways, which achieved greater than 95% theoretical yield (*Figure 2.1* and *Figure 2.10*). With this library of strains in hand, we seek to characterize the systems-level changes that enable the large changes in carbon flux.

While genome sequencing indicates that the genotypes of these strains are remarkably similar, with 83.9% carrying at least one mutation at three genetic loci (*rpoBC*, *pcnB*, and *rne*; *Table 2.3*) preliminary studies indicate that the molecular details of the changes between these strains may differ greatly. Our current working hypothesis is that the mutations in these genes provide a balanced remodeling of the transcriptome, sufficient to allow for multiple and synergistic changes in metabolism without resulting in cell death. Despite the shared overall phenotype of higher productivity, it is possible that the changes at each node in terms of transcriptional response and metabolic flux will differ and offer an opportunity to explore wanted to take an expansive genome approaches to survey changes at complete molecular levels within these evolved strain. In this chapter, we explore the profiling of these strains to understand their physiology in an effort to identify new regulatory mechanisms and metabolic control elements. We further seek to apply this knowledge to the development of new platform technologies for rapid engineering of cellular phenotypes.

3.2. Methods and materials

Commercial materials. Terrific Broth (TB), LB Broth Miller (LB), LB Agar Miller, and glycerol, and methylsulfoxide (DMSO) were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), Kanamycin (Km), chloramphenicol (Cm), isopropyl-ß-D-thiogalactopyranoside (IPTG), phenylmethanesulfonyl fluoride (PMSF), tris (hydroxymethyl) aminomethane hydrochloride (Tris-HCl), sodium chloride, dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), and magnesium chloride hexahydrate were purchased from Fisher Scientific (Pittsburgh, PA). Tris-(2-carboxyethyl)phosphine hydrochloride (TCEP) was purchased from Biosynth, Inc. (Itasca, IL). Sodium hydroxide was purchased from Avantor Performance Materials (Center Valley, PA). A sodium salt hydrate (CoA), acetyl-CoA, butyryl-CoA, acetoacetyl-CoA, β-nicotinamide adenine dinucleotide reduced dipotassium salt (NADH), β-nicotinamide adenine dinucleotide hydrate (NAD^+), formic acid,

trichloroacetic acid (TCA), Restriction enzymes, T4 DNA ligase, Phusion DNA polymerase, Q5 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. Amicon Ultra 10,000 centrifugal concentrators were purchased from EMD Millipore (Billerica, MA). MultiScreen$_{HTS}$ 0.22µm filter plates was purchased from Merck Millipore (Cork, Ireland). D-(+)-glucose was purchased from MP Biochemicals (Santa Ana, CA). 2,4-pentanediol, 1,3-butanediol, 4-hydroxy-2-butanone, trans-caryophyllene, dodecane, 3-hydroxy-butyrate acid were purchased from Sigma-Aldrich (St. Louis, MO). DNA purification kits, Ni-NTA agarose, genomic DNA isolation, and RNeasy RNA isolation kit were purchased from Qiagen (Valencia, CA). Illumina TruSeq RNA Sample Prep Kit was purchased from Illumina (Hayward, CA).

Bacterial strains. *E. coli* DH10B was used for DNA construction. All other strains were either developed in Chapter 2 (*Appendix 2.1*) or constructed as part of the work described in this chapter (*Appendix 3.1*).

E. coli DH1Δ5_2403_+TGG_pntAB (*Appendix 3.1*) was generated by the Cas9 system by introducing the indel from strain 2403 (*Table 2.2)* evolved for HB production. Strain 2403 was also cured of production plasmids by the Cas9 system. A series of pCRISPR_Tet_(guide) plasmids were constructed to express a guide to target the selection marker for the corresponding plasmid to be cured. (*Appendix 3.2)*. *E. coli* BW25113Δ5 was generated by Dr. Matthew A. Davis using standard λ_{red} protocol [6].

Introduction of various mutations or other genetic changes into a clean *E. coli* DH1Δ5 or BW25113Δ5 background was achieved using the Cas9 system described in Jiang *et al*. [7] (*Appendix 3.3*). The targeting vectors were constructed using the pTargetF vector as a template by reverse PCR using primer 459 and different primers in the XX-target family (*Appendix 3.3*) followed by self-ligation. The repair fragments were generated by SOE-PCR of two fragments derived from amplification of *E. coli* 799 genomic DNA using the XX-1/XX-2 and XX-3/XX--4 primer sets (*Appendix 3.3*).

Gene and plasmid construction. Plasmid construction was carried out using standard molecular biology techniques using the Gibson protocol [8]. PCR amplifications were carried out with Q5 DNA polymerase or Phusion DNA polymerase, following manufacturer instructions. Primer sequences are listed in *Appendix 3.2*. Constructs were verified by sequencing (Quintara Biosciences; Berkeley, CA).

Constructs for genome mutation. The pCRISPR-Gibson1 plasmids were constructed to clone constructs with specific guide sequence to target *E. coli* genome for introduction of point mutants. The parent plasmid, pCRISPR-Gibson1 (#2786), was generated from pCRISPR (Addgene 42875) to introduce cut sites between sgRNA promoter and the sgRNA to facilitate the use of Gibson assembly to introduce guide sequences for the target DNA. All guide sequences were generated using the Benchling CRISPR tool (*Appendix 3.3*).

pCRISPR-PcnB2409 (#2784) was constructed by insertion of the annealed oligonucleotides, P1155 and P1156, and inserted into the XbaI-HindIII site of pCRISPR-Gibson1 using the Gibson protocol.

pCRISPR-RpoC2406 (#2794) was constructed by insertion of the annealed oligonucleotides, P1232 and P1233, and inserted into the XbaI-HindIII site of pCRISPR-Gibson1 using the Gibson protocol.

pCRISPR_gibson_1guide_2403g2NADP (#2938) was constructed by the insertion of the annealed oligonucleotides, P1268 and P1269, and inserted into the XbaI-HindIII site of pCRISPR-Gibson1 using the Gibson protocol.

Constructs for curing plasmids. pKD46-Cas9-RecA-Cure_Sp (#2811) was constructed by switching the existing Cb^R market with the Sp^R on the pKD46-Cas9-RecA-Cure (#2416) plasmid constructed by Dr. Quanjiang Ji. Plasmid 2416 was double digested by NotI and SapI. Other parts of the backbone plasmid were amplified by the two sets of primers (906 and 1164, and 1167 and 1168). The Sp^R gene was amplified by primers 1165 and 1166 from the Sp^R bearing plasmid pTargetF (#2637). The pCRISPR_Tet (#2792) parent plasmid was constructed from pCRISPR by switching the existing Km^R marker with a Tet^R marker. The Tet^R marker was amplified from pCas_Tet^R using the 907/908 primer set and inserted into the SacI and EagI site of pCRISPR to replace the Km^R marker. The pCRISPR_Tet carries the XcmI and SacI sites for digestion to allow guide insertion between the sgRNA promoter and the sgRNA.

pCRISPR_Tet_g1Km (#2935) was constructed to target the plasmid bearing the Km^R marker in the evolved strains by insertion of the annealed oligonucleotides, P1256 and P1257, into the XcmI and SacI site of pCRISPR-Tet using the Gibson protocol.

pCRISPR_Tet_g3Cb (#2936) was constructed to target the plasmid bearing the Cb^R/Ap^R marker in the evolved strains by insertion of the annealed oligonucleotides, P1254 and P1255, into the XcmI and SacI site of pCRISPR-Tet using the Gibson protocol.

pCRISPR_Tet_g1Cm (#2937) was constructed to target the plasmid bearing the Cm^R marker in the evolved strains by insertion of the annealed oligonucleotides, P1273 and P1274, into the XcmI and SacI site of pCRISPR-Tet using the Gibson protocol.

Generation of chromosomal point mutations. Point mutations were made using the CRISPR-Cas9 system[9] [10]. Cells were transformed with the pKD46-Cas9-RecA-Cure which allows the expression of the Cas9 protein to generate a double-stranded DNA break and the RecA protein to assist homologous recombination. After growing overnight on LB Cb agar at 30^{o}C, a single colony was picked and inoculated in LB Cb (10 mL) for overnight growth at 30°C. This culture was used to inoculate LB Cb with 0.2% *w/v* arabinose (to induce RecA) to OD_{600} ~ 0.01, which was incubated at 30°C before harvesting at OD_{600} = 0.4 to make electrocompetent cells. Afterwards, electrocompetent transformants were transformed with the pCRISPR plasmid containing the guide as well as the appropriate repair fragments with the desired sequence. The repair fragments also carry a silent mutation to remove the PAM site and a phosphorothioate modification at both the 5′- and 3′-ends. Cells were recovered at 30°C for 1.5 h, plated on LB agar containing the appropriate antibiotic, and incubated at 30 °C overnight. At this point, strains were validated by

Sanger sequencing of the appropriate fragment amplified by colony PCR (Quintara Biosciences). PCR primers were at least 100 bp upstream and downstream from both the 5′- and 3′-ends the repair fragments to avoid false positive results. Once the desired mutations were confirmed, cells were grown at 30 ^{o}C in LB containing IPTG (0.05 mM) (10 mL) to cure the pCRISPR guide plasmid. Finally, these cells were plated onto LB agar and incubated at 37 ^{o}C to cure the pKD46-Cas9-RecA-Cure plasmid, which contains a temperature sensitive origin of replication.

DH1Δ*5*_2406_*pcnB*(R149L): The C<u>G</u>C → C<u>T</u>C mutation at position 446 that corresponds to the *pcnB*(R149L) mutation was made in DH1Δ5 using the CRISPR-Cas9 system. DH1Δ*5* pKD46-Cas9-RecA-Cure was transformed with pCRISPR_gibson_1guide_2409pcnB (#2784) plasmid and the appropriate repair fragments (P1227_2406_pcnB RF_R and P1226_2406_pcnB RF_F).

DH1Δ*5* _2406_*rpoC*(M466L): The <u>A</u>TG →<u>C</u>TG mutation at position 1396 that corresponds to the *rpoC*(M466L) mutation was made DH1Δ5 using the CRISPR-Cas9 method as described above with the pCRISPR_gibson_1guide_2406_rpoC (#2794) plasmid and the appropriate repair fragments (P1231_2406_rpoC_RF_R and P1230_2406_rpoC_RF_F).

DH1Δ*5* _2406_*pcnB*(R149L)_*rpoC*(M466L): The double mutant was made starting from DH1Δ*5*_2406_*pcnB*(R149L) using the CRISPR-Cas9 method described above with the pCRISPR_gibson_1guide_2406_rpoC (#2794) construct and the appropriate repair fragments (P1231_2406_rpoC_RF_R and P1230_2406_rpoC_RF_F).

DH1Δ*5* _2403_*pcnB*(G141A): The G<u>G</u>C →G<u>G</u>C mutation that corresponds to the *pcnB*(G141A) mutation was made in DH1Δ5 using the CRISPR-Cas9 method described above with the pCRISPR_gibson_1guide_2409pcnB (#2784) plasmid and the appropriate repair fragments (P1258_2403_pcnB_RF and P1275_2403_pcnB mutant RF_R).

DH1Δ*5* _2403_+TGG_pntA/B: The insertion of TGG at 38 bp upstream of *pntA* was made in DH1Δ5 using the CRISPR-Cas9 method described above with the pCRISPR_gibson_1guide_2403g2NADP (#2938) plasmid and appropriate repair fragments (P1267_2403_NADPH transhydrogenase RF and P1276_2403_NADPH transhydrogenase RF_R) were used.

DH1Δ*5*.2403*: All three plasmids from the strain 2403 evolved for HB production (DH1Δ*5*.2403) were cured using the CRISPR-Cas9 method to generate the DH1Δ*5*.2403* strain. DH1Δ*5*.2403 was transformed with the pKD46-Cas9-RecA-Cure_Sp (#2811) plasmid, made chemically competent at 30°C in the presence of 0.2% *w/v* arabinose, and transformed with pCRISPR_Tet_g1Km (#2935) to target the Km^{R} resistant plasmid in the host. This transformation was recovered at 30 ^{o}C for 2 hr and incubated at 30 ^{o}C overnight on LB Sp Tc agar plates. A single colony was picked and inoculated into LB Sp (5 mL) containing IPTG (0.5 mM) and grown at 30 ^{o}C overnight. As this point, the cells were plated separately onto LB Agar plates containing either Sp, Km, or Tc to confirm the loss of the both the original Km^{R} plasmid in the host as well as pCRISPR_Tet_g1Km (#2935). Once confirmed, the process was repeated with the appropriate plasmids with pCRISPR_Tet_g1Cm (#2937) and pCRISPR_Tet_g3Cb (#2936) to target the Cm^{R} and Cb^{R} plasmids, respectively. The pKD46-Cas9-RecA-Cure_Sp was cured from the host by growth at 37 ^{o}C. The culture was then plated

on a LB agar plate and grew at 37 ºC overnight. Finally, the single colony was picked, grew in LB overnight, and plated on LB agar plate and LB agar plates containing Sp, Tc, Km, Cb, and Cm. The DH1Δ5.2403* only grew on the LB agar plate.

Production of C_4 compounds in shake flasks. Overnight cultures of freshly transformed *E. coli* strains were grown for 12–16 h in TB at 37°C. These cultures were used to inoculate TB (30 ml) containing the appropriate antibiotics in which the standard glycerol supplement was replaced with glucose (aerobic, 2.5% *w/v*; anaerobic, 2.5% *w/v*) to a final OD_{600} = 0.05. A 250 mL-baffled flask (Kimble Glass; Chicago, IL) with a standard metal cap was used for aerobic cultures and a 250 mL-baffled anaerobic flask with GL45 threaded top (Chemglass) was used for anaerobic cultures. 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. The growth temperature was then reduced to 30°C. Cultures were sealed and the headspace was sparged with Ar for 3 min immediately follow induction. For isoprenoid production, cultures (40 mL) were grown at 37 °C for 3 h before induction with IPTG (1.0 mM). Dodecane (10 mL) was then added as an overlay to the culture. Cultures were sealed and the headspace was sparged with Ar for 3 min. At this time, the growth temperature was reduced to 30°C. Cultures were grown for 5 d before harvesting.

Quantification of 1,3-butanediol (BDO) and 4-hydroxy-2-butanone (HB) titers. Samples (2 mL) were removed from cell culture and cleared of biomass by centrifugation at 20,817*g* for 2 min using an Eppendorf 5417R centrifuge. The cleared medium samples, or standards prepared in TB medium, were diluted 1:1000 into water and filtered through a 0.22 μm filter (EMD Millipore MSGVN2210). Supernatants were diluted 1- to 1,000-fold fold with water containing 2,4-pentanediol (10 μM) added as internal standard and analyzed on an Agilent 1290 HPLC using a Rezex ROA-Organic Acid H^+ (8%) column (150 × 4.6 mm, Phenomenex) with isocratic elution (0.5% *v/v* formic acid, 0.6 mL min^{-1}, 55 °C). Samples were detected with an Agilent 6460C triple quadrupole MS with Jet Stream ESI source, operating in positive MRM mode (*m/z* 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 0.3125, 0.625, 1.25, 2.5, 5, 10 g L^{-1} 1,3-butanediol and 4-hydroxy-2-butanone.

Quantification of *n*-butanol titers. Samples (2 mL) were removed from cell culture and cleared of biomass by centrifugation at 20,817*g* 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 hexanol internal standard (10 g 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} H_2 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). Standard curve was normalized for injection volume using the internal standard.

Quantification of PHB. To analyze for PHB content, dry lyophilized cell samples of known weight were treated with concentrated H_2SO_4 (1 mL per 30 mg biomass) at 90 °C for 60 min to convert PHB into its monomer, crotonic acid. Samples were analyzed by LC-UV/Vis (Agilent 1200) using an Aminex HPX87H column (BioRad, Hercules, CA) with 7 mM H2SO4 as the mobile phase and acrylic acid as the internal standard. The eluent was monitored by UV at 214 nm [*11*].

Quantification of isoprenoid. For isoprenoid quantification, dodecane layer was removed and an aliquot (250 µL) was mixed ethyl acetate (250 µL) containing 5 mg/L *trans*-caryophyllene as an internal standard. 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: 7°C for 3 min, ramp to 300 °C at 45°C min^{-1}, 300°C for 1 min. Compounds were identified by comparison of the full mass spectrum to library compounds (isoprenoids). For quantification, the peak area of the compounds of interest was compared to the peak areas of the internal standard *trans*-caryopyhllene [*11*].

Quantification of hydroxy acid titers. Cell culture samples (1 mL) after 5 d of growth were cleared of biomass via centrifugation at 20,817 × g for 2 min with an Eppendorf 5417R Centrifuge (Hamburg, Germany). The supernatant (10 µL) was diluted in water (190 µL) containing 0.5 mM adipic acid as internal standard. Samples were filtered through a 96-well MultiScreenHTS plate before injecting onto an Agilent 1290 HPLC equipped with an auto-sampler, Phenomenex (Torrence, CA) Rezex-ROA Organic Acid H+ column (150 × 4.6 mm), and Carbo-H+ Security Guard cartridge. 0.5% *v/v* formic acid was used as mobile phase (0.3 mL/min, column temperature 55°C), and hydroxy acids were quantified by mass spectrometry on an Agilent 6460 triple quadrupole MS with ESI source, operating in negation ion MRM transition mode with fragmentor voltage set at 70V. Between 5-8 min, the following transition and collision energy were monitored: m/z 103.1→59.2, 5V (3-hydroxybutyric acid). Samples were quantified relative to a standard curve of 7.8125, 15.625, 31.25, 62.5, 125, 250, 500, and 1000 mg/L hydroxy acid [*12*].

RNA sequencing and analysis. Cells were harvested after 24 h post-induction for RNA extraction. Total RNA was isolated using the RNeasy RNA isolation kit (Qiagen). rRNA was then removed using the following protocol. Total RNA (5 µg) was treated with TURBO DNaseI (Thermo-Fisher, 4.5 µL) at 37ºC for 30 min in a 50 µL reaction containing 10× buffer (5 µL) to remove genomic DNA. The reaction was diluted with Buffer RLT (Qiagen, 100 µL) and 70% *v/v* ethanol (200 µL) and transferred to an RNeasy column (Qiagen) for RNA cleanup following the manufacturer instructions. This DNase-treated RNA (1 µg) was combined with 0.5 µM DNA probe (1 µL, *Appendix 2.4*) and Hybridization buffer (200 mM NaCl, 100 mM Tris-HCl, pH 7.5) was added to a final volume of 20 µL. Hybridization was carried out using the following program: Hold at 95°C for 2 min, gradient from 95°C to 45°C at -0.1 C/s. At this time, RNase H (5 U, Epicentre) in 10× Digestion buffer (2.5 µL; 0.5M Tris-HCl, pH 7.5, 1 M NaCl, 200 mM $MgCl_2$) was added, and the resulting mixture was incubated at 45 °C for 30 min. Following cleanup with the Qiagen RNeasy Kit, the sample was treated with TURBO DNaseI (3 U). The Qiagen RNeasy Kit was used again to clean up samples before RNA-Seq library prep. RNA-Seq libraries were prepared using the TruSeq RNA Sample Prep Kit (Illumina). Samples were sequenced on an Illumina HiSeq4000

at the DNA Technologies Core (UC Davis, CA). Reads were mapped using Kallisto[*13*] and Sleuth [*14*]. Functional enrichment analysis of differentially expressed genes is based on clusters of orthologous groups (COG) categories provided by the IMG-ER (https://img.jgi.doe.gov/cgi-bin/mer/main.cgi) annotation [*15*].

Metabolomics. Five replicates of cultures were grown as described previously for production in shake flasks and harvested 24 h after induction. Cultures were centrifuged at 20,817 × g for 1 min at 4 °C with an Eppendorf 5417R Centrifuge (Hamburg, Germany). The supernatants were decanted immediately and cell pellets were flash frozen with liquid nitrogen and stored at – 80°C until extraction. Pellets were extracted with 90% *v/v* methanol with 0.1% *v/v* formic acid containing d_3N^{15}-serine (0.01 mg/mL; Cambridge Isotope Laboratories, Inc., DNLM6863) to a final concentration of 1 mg biomass/µL of extraction buffer. The mass of the biomass was calculated using the standard value for *E. coli* of 23.8 mg/OD_{600nm}. Samplers were vortexed for 15 s, incubated at –80 °C for 30 min, and then thawed at – 20 °C for 30 min. The vortex-freeze-thaw cycle was repeated for total of five times. At the end of this procedure, the lysed cells were centrifuged at 20,817 × g at 4°C for 5 min and the supernatant collected for LC-MS/MS analysis. Samples were run in the Agilent LC-MS with the Luna 5 µm NH2 100 Å column.

3.3. Results and discussion

Evolved strains showed large transcriptome landscape changes. Genome sequencing of a total of 31 evolved strains that carried the HB, BDO, and the *n*-butanol pathway revealed genes involved in RNA metabolism were the dominant mutation hits from the limited number of total mutation (*Chapter 2*). In order to further characterize this finding, we decided to perform RNA sequencing on the evolved strains to investigate the changes in global transcriptome compared to the corresponding parent strains. To initiate these efforts, two different sets of strains were chosen. The first set included an HB-evolved strain (#2403) and its corresponding parent strain (*E. coli* DH1Δ*5* pBBR1-AceEF.Lpd pT533-phaA pCWO.trc-TdTer-aldh7.adh2) as it showed a 5-fold improvement in production titer (*Figure 2.10*) and carries a point mutation in the poly(A) polymerase (*pcnB* G141A). The second set of strains was comprised of a BDO-evolved strain (#2406) and its corresponding parental strain (*E. coli* DH1Δ*5* pBBR1-AceEF.Lpd pT533-phaA.phaB pCWO.trc-TdTer-aldh7.adh2) based on the near quantitative yield achieved by this strain compared to its parent (20%). This strain also carried a point mutation in both the poly(A) polymerase (*pcnB* R149L) as well as the RNA polymerase ß' subunit (*rpoC* M466L) (*Table 3.1*).

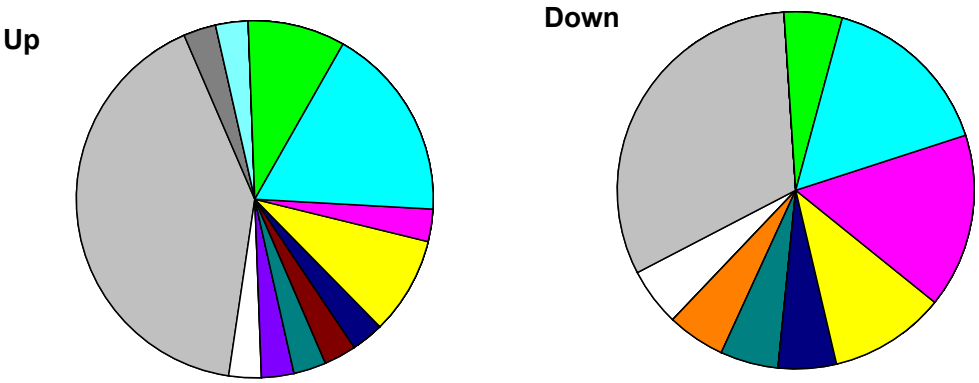
These strains were cultured and sampled 24h after induction of their respective production pathways. Total RNA was extracted from these samples and a method was developed to remove rRNA using annealing of complementary primers followed by RNAse H digestion. Libraries were generated for sequencing on a HiSeq4000 (with SR50 sequencing run; total of 24 samplers were pooled into one lane; total of 408,620,227 clusters were obtained for the entire lane). The reads were then mapped using Kallisto (>90% of reads were mapped) [*13*] and Sleuth [*14*] to an *E. coli* DH1 reference genome (Accession No. NC_017625). A ß value of 2 and p-value of 0.05 was used to determine differentially expressed genes. Analysis of the RNA-seq data reveals that there are indeed a number of changes occurring in the transcriptional landscape (HB, 49 differentially

Product	Host	Strain #	Gene	DNA Changes	Amino acid change
4-hydroxy-2-butanone	DH1Δ5	2403	*pcnB*	GGC → GCC	G141A
			pntA/B	Added GGT (intergenic (-38 / -486))	N/A
1,3-butanediol	DH1Δ5	2406	*pcnB*	CGC → CTC	R149L
			rpoC	ATG → CTG	M466L

Table 3.1. Strains characterized by RNA sequencing. Key mutations from the genome sequencing of strains evolved for production of HB and BDO. *E. coli* DH1Δ*5* pBBR1-AceEF.Lpd pT533-phaA pCWO.trc-TdTer-aldh7.adh2 (HB #2403), *E. coli* DH1Δ*5* pBBR1-AceEF.Lpd pT533-phaA.phaB pCWO.trc-TdTer-aldh7.adh2 (BDO #2406).

A

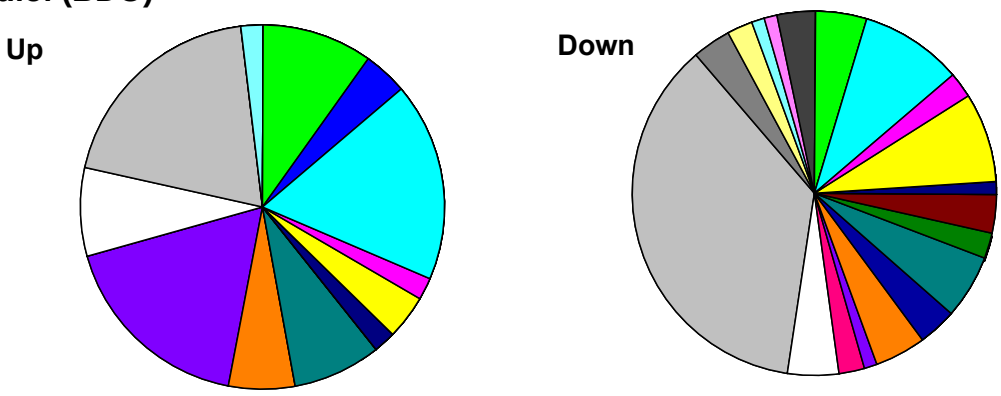
Hydroxybutanone (HB)



	Function	Genome	Up	Down
A	RNA processing and modification	2 (0.06)	0 (0)	0 (0)
C	Energy production and conversion	260 (7.72)	3 (9.09)	1 (6.25)
D	Cell cycle control, cell division, chromosome partitioning	38 (1.13)	0 (0)	0 (0)
E	Amino acid transport and metabolism	354 (10.51)	6 (18.18)	3 (18.75)
F	Nucleotide transport and metabolism	106 (3.15)	1 (3.03)	3 (18.75)
G	Carbohydrate transport and metabolism	380 (11.28)	3 (9.09)	2 (12.5)
H	Coenzyme transport and metabolism	179 (5.31)	0 (0)	0 (0)
I	Lipid transport and metabolism	123 (3.65)	1 (3.03)	1 (6.25)
J	Translation, ribosomal structure and biogenesis	227 (6.74)	0.00	0.00
K	Transcription	292 (8.67)	1 (3.03)	0.00
L	Replication, recombination and repair	137 (4.07)	0.00	0.00
M	Cell wall/membrane/envelope biogenesis	240 (7.12)	1 (3.03)	1 (6.25)
N	Cell motility	106 (3.15)	0 (0)	0.00
O	Posttranslational modification, protein turnover, chaperones	149 (4.42)	0 (0)	1 (6.25)
P	Inorganic ion transport and metabolism	207 (6.14)	1 (3.03)	0.00
Q	Secondary metabolites biosynthesis, transport and catabolism	57 (1.69)	0 (0)	0.00
R	General function prediction only	262 (7.78)	1 (3.03)	1 (6.25)
S	Function unknown	203 (6.03)	14 (42.42)	6 (37.5)
T	Signal transduction mechanisms	189 (5.61)	1 (3.03)	0.00
U	Intracellular trafficking, secretion, and vesicular transport	53 (1.57)	0 (0)	0.00
V	Defense mechanisms	88 (2.61)	1 (3.03)	0.00
W	Extracellular structures	32 (0.95)	0 (0)	0.00
X	Mobilome: prophages, transposons	60 (1.78)	0 (0)	0.00
	Total	**3369**	**33**	**16**

B

Butanediol (BDO)



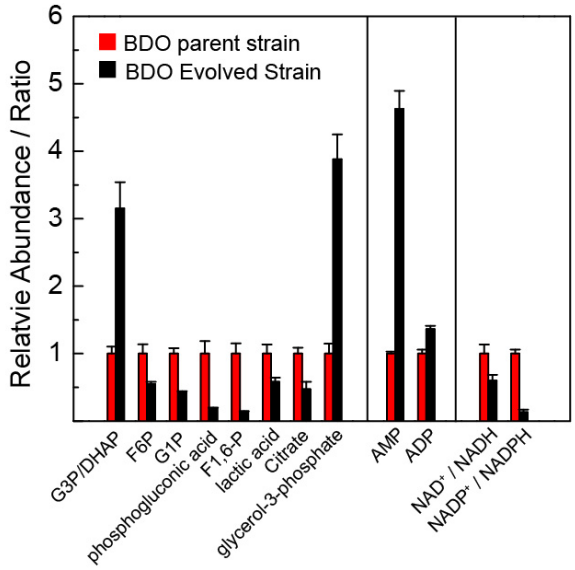
	Function	Genome	Up	Down
A	RNA processing and modification	2 (0.06)	0 (0)	0 (0)
C	Energy production and conversion	260 (7.72)	5 (10.64)	4 (5.06)
D	Cell cycle control, cell division, chromosome partitioning	38 (1.13)	2 (4.26)	0 (0)
E	Amino acid transport and metabolism	354 (10.51)	9 (19.15)	8 (10.13)
F	Nucleotide transport and metabolism	106 (3.15)	1 (2.13)	2 (2.53)
G	Carbohydrate transport and metabolism	380 (11.28)	2 (4.26)	7 (8.86)
H	Coenzyme transport and metabolism	179 (5.31)	0 (0)	0 (0)
I	Lipid transport and metabolism	123 (3.65)	1 (2.13)	1 (1.27)
J	Translation, ribosomal structure and biogenesis	227 (6.74)	0.00	0.00
K	Transcription	292 (8.67)	0.00	3 (3.80)
L	Replication, recombination and repair	137 (4.07)	0.00	2 (2.53)
M	Cell wall/membrane/envelope biogenesis	240 (7.12)	4 (8.51)	5 (6.33)
N	Cell motility	106 (3.15)	0 (0)	3 (3.80)
O	Posttranslational modification, protein turnover, chaperones	149 (4.42)	3 (6.38)	4 (5.06)
P	Inorganic ion transport and metabolism	207 (6.14)	9 (19.15)	1 (1.27)
Q	Secondary metabolites biosynthesis, transport and catabolism	57 (1.69)	0 (0)	2 (2.53)
R	General function prediction only	262 (7.78)	4 (8.51)	4 (5.06)
S	Function unknown	203 (6.03)	10 (21.28)	32 (40.51)
T	Signal transduction mechanisms	189 (5.61)	0 (0)	3 (3.80)
U	Intracellular trafficking, secretion, and vesicular transport	53 (1.57)	0 (0)	2 (2.53)
V	Defense mechanisms	88 (2.61)	1 (2.13)	1 (1.27)
W	Extracellular structures	32 (0.95)	0 (0)	1 (1.27)
X	Mobilome: prophages, transposons	60 (1.78)	0 (0)	3 (3.80)
	Total	**3369**	**47**	**79**

Figure 3.1. RNA-Seq profile of evolved HB and BDO producing strain. Clusters of orthologous groups (COG) categories for genes differentially expressed between the parent and evolved strains. COG categories were identified by the IMG-ER annotation pipeline. COG categories represented by genes that are upregulated and downregulated 24 h after induction with IPTG. Comparison of COG category representation in the differentially expressed genes compared to the entire genome. The number of the open reading frames represented by each COG is given, and the percentage of total genes with COG categories is in parentheses. Since some genes fall into multiple COG categories, the percentage was calculated by dividing the total number of unique genes. (A) DH1Δ*5* pT533-phaA, pCWori-trc.ALDH7.ADH2, pBBR2-aceE.F.lpd (parent) and evolved HB strain (DH1Δ*5*.2403). (B) DH1Δ*5* pT533-PhaAB pCWori-trc.ALDH7.ADH2, pBBR2-aceE.F.lpd (parent) and evolved BDO strain (DH1Δ*5*.2406).

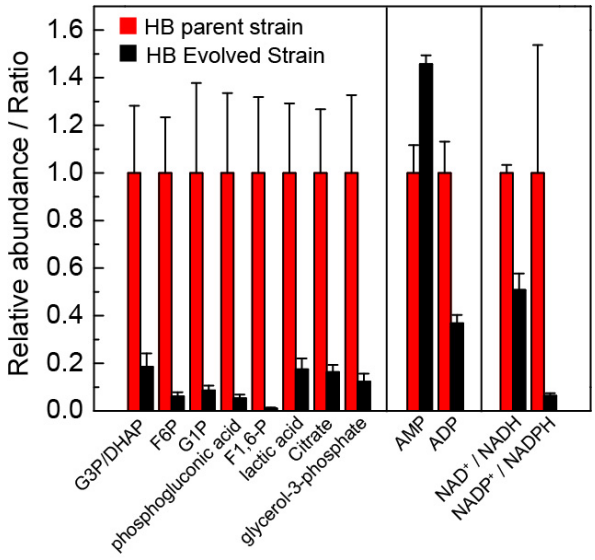
Product	Host	Strain #	Gene	DNA Changes	Amino acid change
4-hydroxy-2-butanone	DH1Δ5	2403	*pcnB*	GGC → GCC	G141A
			pntA/B	Added GGT (intergenic (-38 / -486))	N/A
1,3-butanediol	DH1Δ5	2406	*pcnB*	CGC → CTC	R149L
			rpoC	ATG → CTG	M466L
n-butanol	DH1Δ5	2622	*pcnB*	GAT → GAG	D194E
	BW25113Δ5	2731	rne	CGC → CAC	R488H

Table 3.2. Strains characterized by metabolomics. Key mutations from the genome sequencing of strains evolved for production of HB, BDO, and n-butanol. *E. coli* DH1Δ*5* pBBR1-AceEF.Lpd pT533-phaA pCWO.trc-TdTer-aldh7.adh2 (HB #2403), *E. coli* DH1Δ*5* pBBR1-AceEF.Lpd pT533-phaA.phaB pCWO.trc-TdTer-aldh7.adh2 (BDO #2406). *E. coli* DH1Δ*5* pBBR1-AceEF.Lpd pT5T33-Bu2 containing either pCWori.TdTer-trc.ALDH46.ADH2 (*n*-butanol #2622). *E. coli* BW25113Δ*5* pBBR1-AceEF.Lpd pT5T33-Bu2 containing either pCWori.TdTer-trc.ALDH21.ADH2 (*n*-butanol #2731).

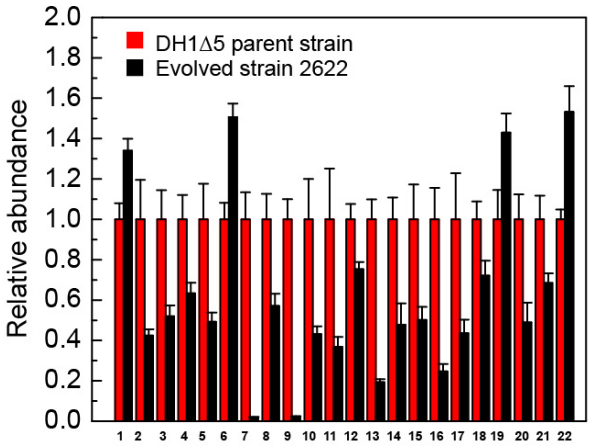
A


BDO parent strain
BDO Evolved Strain
Relatvie Abundance / Ratio
0
1
2
3
4
5
6
G3P/DHAP
F6P
G1P
phosphogluconic acid
F1,6-P
lactic acid
Citrate
glycerol-3-phosphate
AMP
ADP
NAD+ / NADH
NADP+ / NADPH

B

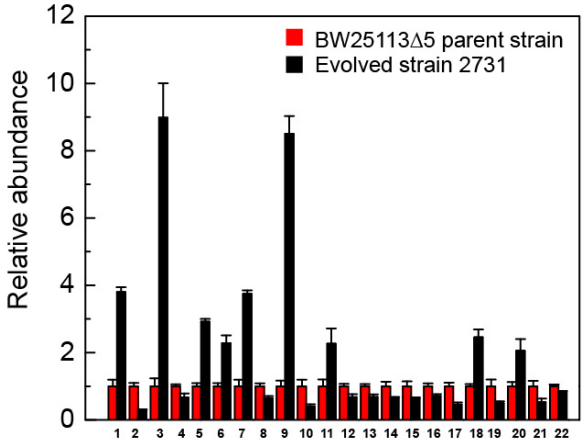

HB parent strain
HB Evolved Strain
Relative abundance / Ratio
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
G3P/DHAP
F6P
G1P
phosphogluconic acid
F1,6-P
lactic acid
Citrate
glycerol-3-phosphate
AMP
ADP
NAD+ / NADH
NADP+ / NADPH

C



Number	Metabolites
1	glyoxylic acid
2	pyruvate
3	uracil
4	succinate
5	oxaloacetate
6	malate
7	hypoxanthine
8	alpha ketoglutarate
9	xanthine
10	transaconitate
11	citrate
12	pantothenate
13	phosphonogluconic acid
14	glutathione, reduced GSH
15	dUTP
16	CTP
17	UTP
18	ATP
19	UMP
20	NADP
21	coenzyme A
22	acetyl-CoA

D



Number	Metabolites
1	glyoxylic acid
2	pyruvate
3	lactic acid
4	cytosine
5	fumarate
6	succinate
7	malate
8	phosphorylethanolamine
9	xanthine
10	phenyl pyruvate
11	inositol
12	glucose old
13	glucose new
14	inositol 4-phosphate
15	fructose-6-phosphate
16	UMP
17	cAMP
18	dUTP
19	CTP
20	uridine 5-disphosphoglucuronate
21	NADH
22	NADP

E

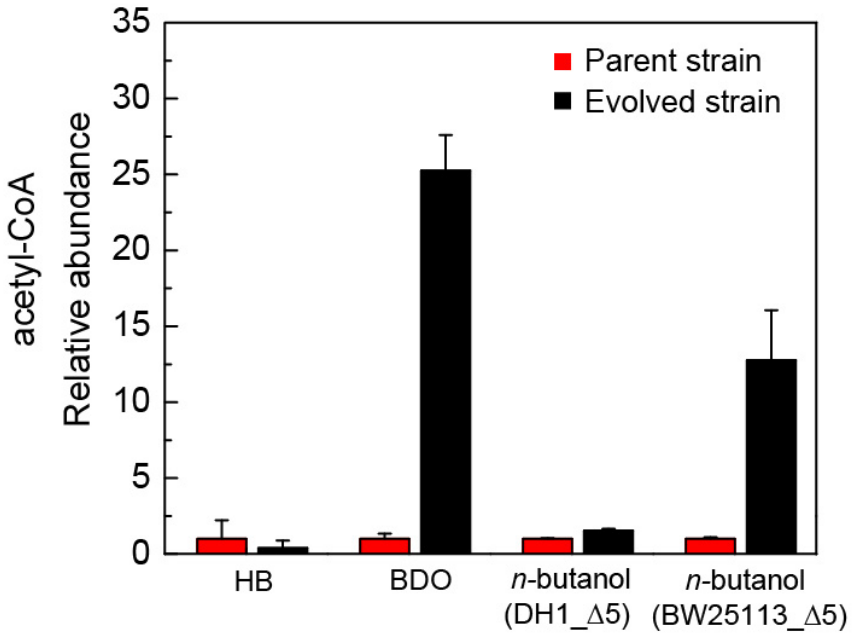


Figure. 3.2. Metabolomics analysis between parent strains and evolved strains. (A) (DH1Δ*5* pT533-PhaAB pCWori-trc.ALDH7.ADH2, pBBR2-aceE.F.lpd (WT)) and evolved BDO strain (DH1Δ*5*.2406). (B) (DH1Δ*5* pT533-phaA, pCWori-trc.ALDH7.ADH2, pBBR2-aceE.F.lpd(WT)) and evolved HB strain (DH1Δ*5*.2403). (C) n-butanol (DH1Δ*5)* pT5T33-phaA.HBD-crt (499), pBBR2-aceE.F.lpd (WT) (339), pCWO.trc-ter-aldh46.adh2(1866) and evolved strain (DH1Δ*5.*2622). (D) n-butanol (BW25113Δ*5)* pT5T33-phaA.HBD-crt (499), pBBR2-aceE.F.lpd (WT) (339), pCWO.trc-ter-aldh21.adh2 (2456) and evolved strain (BW25113Δ5.2731*).* (E) All strains are the same as described in A, B, C, and D.

regulated genes (33 up-regulated and 16 down-regulated; BDO, 126 differentially-regulated genes (47 up-regulated and 79 down-regulated; *Appendix 3.4*). However, it is interesting to note that despite the mutations to core genes in RNA metabolism, the number of changes are moderate and may indicate that they support sufficient change to alter homeostasis but not to incur cell death. We also observe that the differentially expressed genes from both sets of strains covered a wide range of biological process categories. The highest number of differentially expressed genes were assigned to energy production and conversion, amino acid transport and metabolism, cell envelope biogenesis, and carbohydrate transport and metabolism. Although both evolved strains displayed a similar phenotype of a large improvement in carbon conversion to product as well as similar genotype with mutations in genes involved in RNA metabolism, their transcriptome profiles were quite distinct (*Figure 3.1*). Indeed, the design of both the HB and BDO synthetic pathways are very similar, using similar chemistry and the same starting acetyl-CoA metabolite. However, the major differences between these two pathways do exist in terms of the number of reducing equivalents used and the chemical properties of the final product, which could contribute to the difference in product secretion and toxicity.

Metabolomics data revealed significant changes in central metabolism in evolved strains. Since the major phenotype selected for and observed is the increase in glucose conversion to product, we hypothesized that the changes in gene expression were related to changes in carbon flux through the metabolic network between the evolved and parent strains. In order to explore this possibility, four different sets of strains were chosen for metabolome profiling (*Table 3.2*) These strains represented the production of three C_4 monomers, 4-hydroxy-2-butanone (HB), 1,3-butanediol (BDO), and *n*-butanol. Both the HB and BDO strains are derived from DH1Δ*5* while *n*-butanol strains derived from both DH1Δ*5* (2622) and BW25113Δ*5* (2731) were chosen for characterization. These *n*-butanol strains also showed ~ 5-fold improvement in production titer compared with the parent strain and carried the key RNA processing mutations (*Figure 2.10, Table 3.2*). As with RNA sequencing, cells were grown under the standard production conditions and harvested for metabolomics analysis 24 h after pathway induction. Metabolites were extracted by mix-freeze-thaw cycles in 90% *v/v* methanol with 0.1% *v/v* formic acid. The cell extracts were analyzed by LC-MS/MS.

Preliminary data shows that metabolites in the central metabolism (glycolysis, TCA cycle) were significantly different between the parent and evolved strain for the HB, BDO, and the *n*-butanol pathways (*Figure 3.2ABCD, Appendix 3.5*). In addition, metabolites involved in energy conversion (ADP/AMP) and redox state (NAD(H) and NADP(H)) are quite different as well. Interestingly, the profiles of the HB and BDO evolved strains are quite different from each other, suggesting that there may be many solutions to the overall problem of increasing flux to these two pathways. Strikingly, when the acetyl-CoA pools are compared between all four sets of strains, only 2 out of 4 showed the expected large increases in the acetyl-CoA pool (2406; 25-fold increase; 2731, 12-fold increase; *Figure 3.2E*). These findings highlight the potential for a diverse set of approaches for breaking acetyl-CoA homeostasis in this system as well as the possibility for furthering our understanding of metabolic regulation.

Physiological characterization of the parent strains and evolved strains. Additional cell growth experiments were carried out to explore the role of these mutations in respect to cellular physiology. There was a very large difference on cell growth patterns between the HB parent and evolved strains. The evolved strain grew almost 5-fold better than the WT (*Figure 3.3A, left panel*).

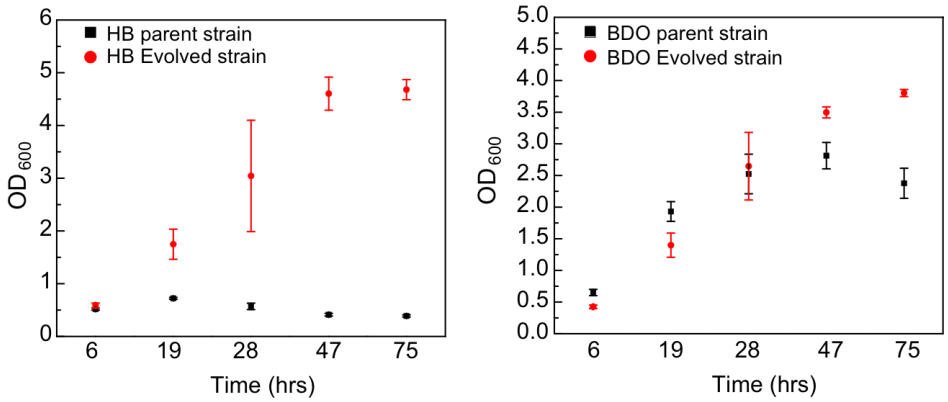
In addition to this significant growth enhancement, an obvious color change of production media was also observed between the HB WT and the evolved strain (*Figure 3. 4C*) and may be caused by a change in secreted products. This could be the result of reducing equivalent and redox potential differences between these strains. On the other hand, the growth difference between the BDO parent and evolved strains was not significant until 47 h induction. At 75 h, the evolved strain showed a 40% improvement on cell growth. This moderate increase could be related to the maintenance of pH by the evolved BDO strain (pH 7) compared to the parent (pH 6) (*Figure 3.3B*).

Metabolomics data showed that redox pools were different between the parent strains and the evolved strains (*Figure 3.2A*), which is not surprising given that redox usage is the basis of the selection for these strains [6] . We attempted to further characterize the redox status of the different strains by examining the growth and production profiles of these strain with carbon sources at different oxidation states. Three different carbons were selected in addition to the standard C_6 sugar, glucose. Sorbitol and gluconic acid were chosen as reduced and oxidized C_6 sugars, respectively. We also decided to include a standard reduced C_3 carbon source, glycerol. Althought the results are not definitive, it is interesting to note that the growth defect for the HB parent strain disappeared with all three new carbon sources, sorbitol, gluoconic acid and glycerol. Indeed, there was no difference in cell growth with these sugars. In contrast, the HB evolved strain grew ~4-fold better than the HB parent strain when glucose was fed (*Figure 3.4A, left panel*). Although glucose still yielded the highest production titer for HB, the HB evolved strain was able to reach higher product titers compared to the parent with all carbon sources (*Figure 3.4A, right panel*). For the BDO strains, there were significant cell growth differences between the parent strain and the evolved strain under both glucose and sorbitol (~2-fold) (*Figure 3.4B, left panel*). However, the cell growth difference between the parent and evolved strains were much smaller when gluconic acid and glycerol were used as the carbon source. However, in terms of production it is clear that the evolved strains show a large advantage with all three carbon sources (*Figure 3.4B, right panel*).

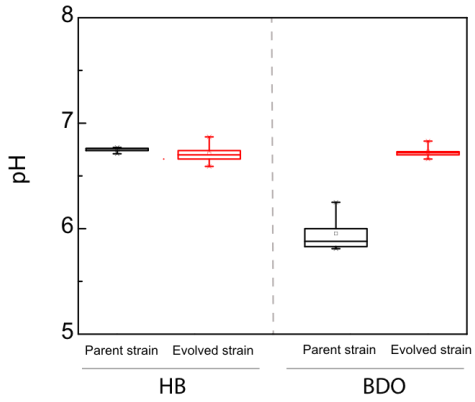
Taken together, these experiments indicate that there is a complex relationship between cell growth and productivity even though the evolved strains were originally selected for by adaptive evolution. It also suggests that details in how these sugars enter metabolism and are converted to acetyl-CoA as well as the different metabolic programs that may exist in these different hosts are also important for a more detailed understanding the outcome of this experiment. However, they show that the evolutionary reprogramming of these hosts can yield an advantage under many different conditions, showing a benefit for fermenting a wide range of carbon substrates. Furthermore, they suggest that these global RNA processors could be good targets for engineering to improve fermentation under different conditions.

Exploring the role of *pcnB* and *rpoC* in the evolved strains. To validate the impact of these key mutations that arose from the evolution experiments, these mutations were made in a clean genetic background and their production profiles were examined. Two key mutations from the BDO evolved strain, *pcnB*(R149L) and *rpoC*(M466L), as well as the double mutant *pcnB*(R149L)_ rpoC(M466L) were made. The plasmids corresponding to the BDO pathway were transformed into these mutants and conducted the standard BDO production experiment. The *pcnB*(R149L) mutant gave a lower production titer (~50% decreased) compare to the parent strain. However, mutations in *rpoC* and *pcnB* are synergistic, as both are required to achieve a substantive increase

A



B



C

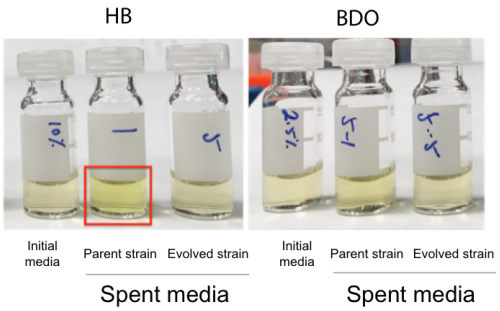


Figure 3.3. Physiology studies of the parent strains and evolved strains. (A)Time course experiment of cell growth for the HB WT and HB evolved strain (#2403) (Left) and the BDO WT and BDO evolved strain (#2406) (Right). (B) pH profile of spent media after 5 d of production. (C) Photograph of media after 5 d of production. The HB evolved strain grew almost 5 times better than the HB parent strain. There was no significant growth difference between the BDO parent and evolved strain. BDO evolved strain appeared to maintain a neutral starting pH after 5 days of production. Significant color changed was observed in the spent media for after 5 days of production experiment from the HB parent strain.

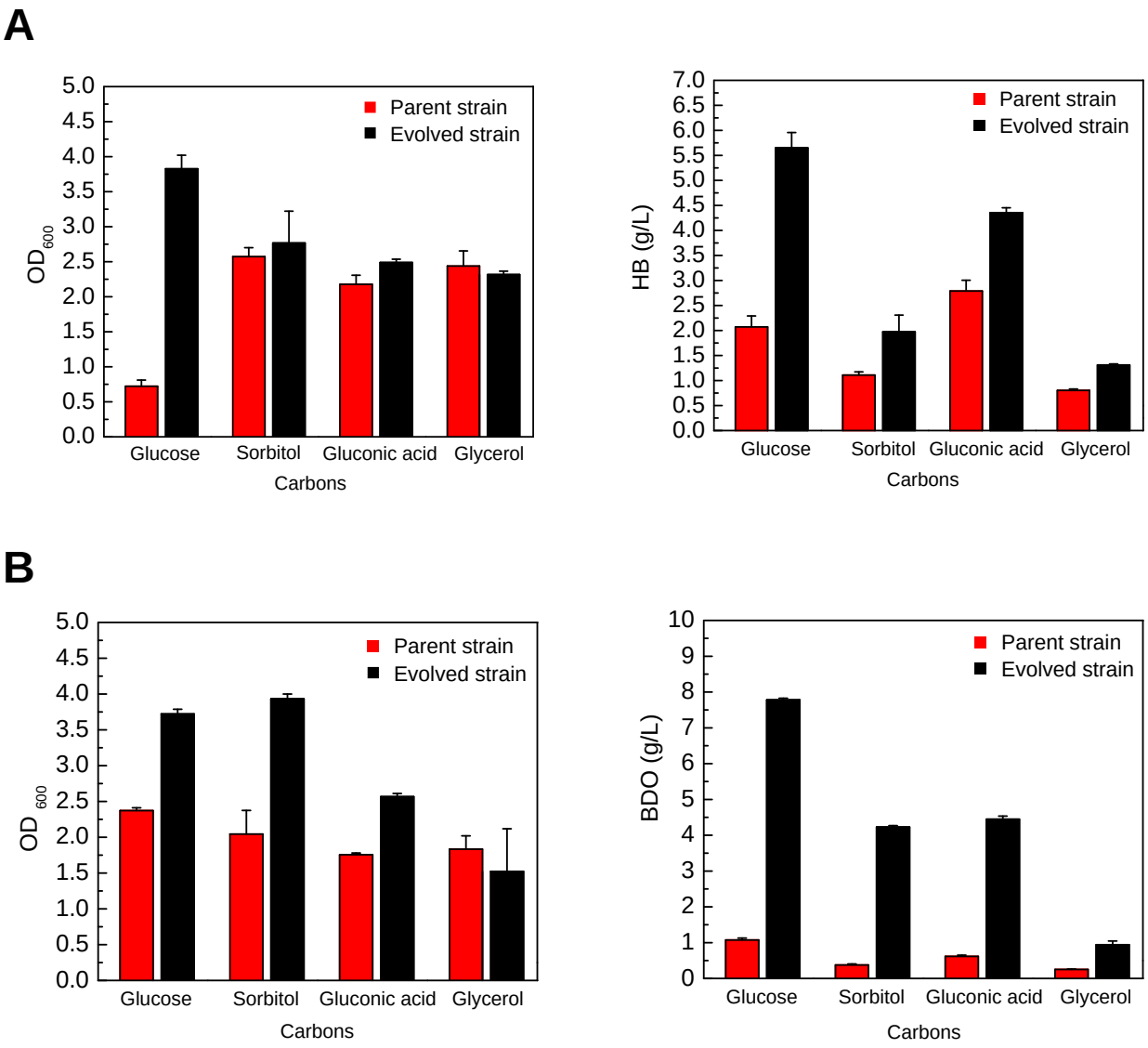


Figure 3.4. Cell growth and production profiles under different carbon sources. (A) Cell growth for the HB parent strain the HB evolved strain under different carbon sources and the corresponding production profile. HB parent strain (DH1Δ*5* pT533-phaA, pCWori-trc.ALDH7.ADH2, pBBR2-aceE.F.lpd(WT)) and evolved HB strain (DH1Δ*5*.2403). (B) Cell growth for the BDO parent strain the BDO evolved strain under different carbon sources and the corresponding production profile. BDO parent strain (DH1Δ*5* pT533-PhaAB pCWori-trc.ALDH7.ADH2, pBBR2-aceE.F.lpd (WT)) and evolved BDO strain (DH1Δ*5*.2406).

in BDO titer compared to the parent (*Figure 3.5A, left panel*). Indeed, the double mutant demonstrates a 2.75-fold increase in BDO titers (parent, 2.1 ± 0.1 g L^{-1} ; DH1Δ*5*.2406, 5.8 ± 0.2 g L^{-1}) , which recapitulates 73% of the improvement observed in the fully evolved strain (8.1 ± 0.1 g L^{-1}). Two key mutations from the HB evolved strain were reconstructed in the clean genetic background. One of them was the glycine to alanine mutation at the 141-amino acid residue for the poly(A) polymerase, *pcnB.* Another key interesting indel that arose from the HB evolved strain was the addition of three nucleotide TGG in the upstream sequence of the NAD(P) transhydrogenase alpha/beta subunits (pntA/B). Introducing the indel sequence from the upstream sequence of the pntA/B gave a 50% increase in production titer for HB, while the pcnB(G141A) mutant resulted a ~50% drop in production titer. Interestingly, the indel from the upstream pntA/B sequence the pcnB(G141A) have demonstrated a synergistic effect. This was demonstrated by the curing all the plasmids from the evolved strain (resulted the evolved strain*) and re-transformed the HB pathways back. The titer from the evolved strain* represented ~ 85% of the evolved strain production titer (*Figure 3.5A, right panel*). We were also interested in the generality of these mutations and thus tested their ability to stimulate yield increases in a different pathway. When the *n*-butanol pathway is introduced into the double mutant, we observed a 3.2-fold increase in product titer from 2.3 ± 0.6 to 7.3 ± 1.1 g L^{-1}. (*Figure 3.5B*).

Interestingly, none of the strains - *pcnB*(R149L), *rpoC*(M466L), and the double mutant – showed any growth difference in the absence of a synthetic pathway (*Figure 3.6A*). In the presence of the BDO pathway, the strains bearing single point mutants (*pcnB*(R149L) and *rpoC*(M466L)) showed a slight growth defect, while the double mutant gave a net positive effect (~30% improvement) (*Figure 3.6B*). This finding is potentially in contrast to previous reports, which showed that strains evolved for improved growth in minimal media (50%) were found to contain a deletion from the *rpoC* gene, implying that mutations in *rpoC* could support changes in growth phenotype[16].

These data have demonstrated that these key mutations in the *pcnB* and *rpoC* are capable of driving a large shift in central carbon metabolism that can be generalized to related pathways utilizing the acetyl-CoA building block. We set out to conduct production experiments with other acetyl-CoA dependent pathways using these strains as hosts. Three different acetyl-CoA dependent pathways were examined: the polyhydroxybutyrate (PHB) pathway, 3-hydroxy acid pathways, and the isoprenoid pathway. First, both the PHB production were conducted under both aerobic and anaerobic conditions (*Figure 3.7A*). Compared to the parent strain, the titer of monomer (crotonic acid) dropped by ~50% in the double mutant DH1Δ*5*.*pcnB*(R149L).*rpoC*(M466L) under anaerobic conditions (parent, 5.8 ± 0.5 g L^{-1}; double mutant, 3.1 ± 0.5 g L^{-1}). No product was observed under the aerobic condition, which suggested these mutants may be oxygen sensitive. Second, two different 3-hydroxy acid pathways were tested. One of them uses NADH as the cofactor, and one uses NADPH as the reducing equivalent. In addition to the parent strain and the double mutant, the cured HB-evolved strain (DH1Δ*5*.2403*) was also included. Under anaerobic conditions, there was essentially no difference on production titer for both pathways. However, production titer was decreased by ~50% under aerobic condition for both mutants (*Figure 3.7B*). Finally, the both mutants gave a lower titer of isoprenoid (amorphadiene) as compared to the parent strain under both aerobic and anaerobic conditions. (*Figure 3.7C*). Although these mutants did not show a positive effect on the production of the PHB, hydroxy acid, and the amorphadiene pathways, they demonstrate that the metabolic re-programming in these strains has occurred and is complex.

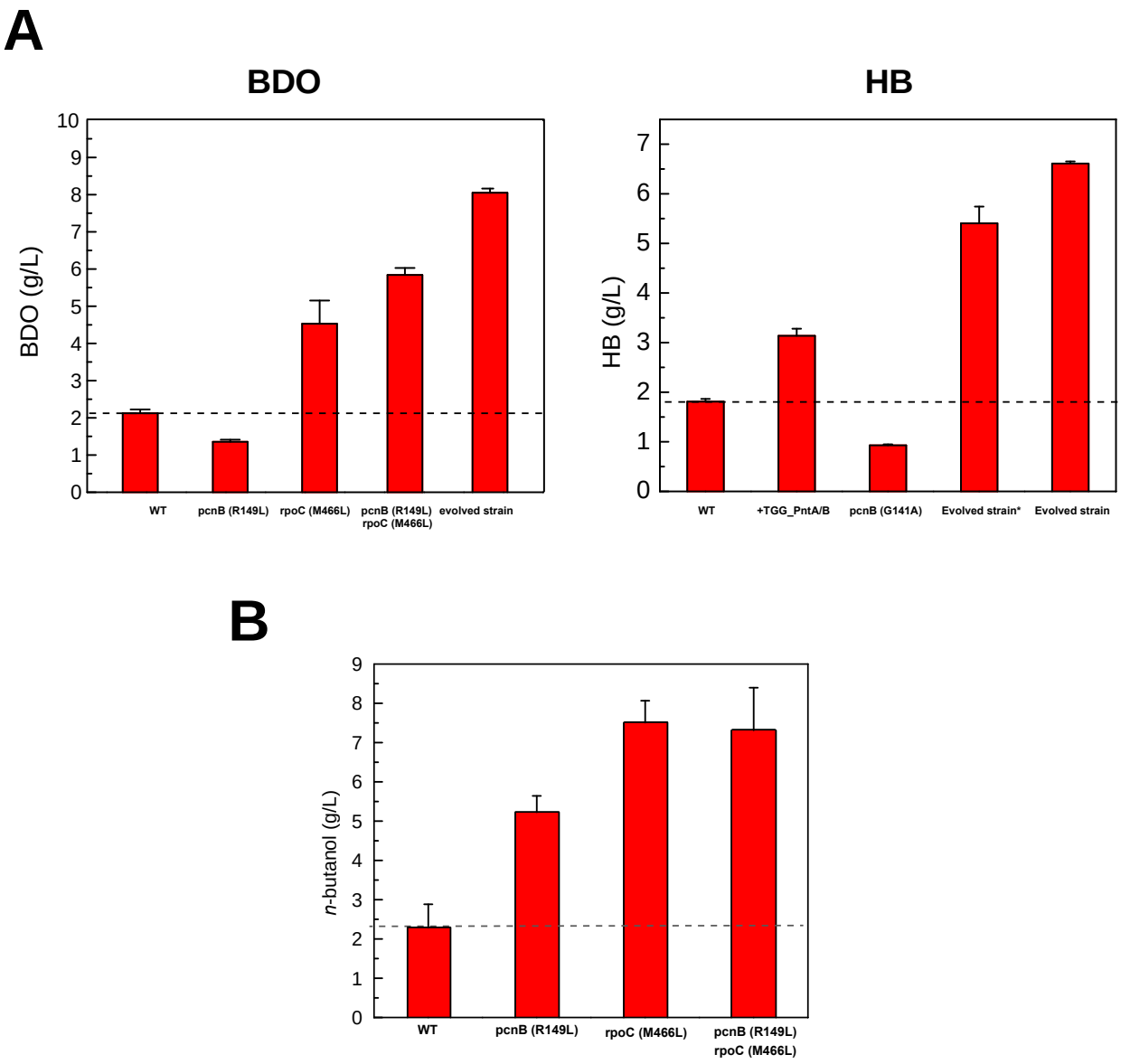


Figure 3.5. Validating mutations arose from evolved strain. (A) Generating the *pcnB* and *rpoC* mutations found in DH1Δ*5*.2406 in a clean genetic background (DH1Δ*5* parent) captures the majority of the improvement observed in the evolved strain, indicating that these two gene loci play an important role in enabling the increases in BDO production (left). Generating the *pcnB* mutation and upstream 3 nucleotides insertion in front of the *pntA/B* found in DH1Δ*5*.2403 in a clean genetic background (DH1Δ*5* parent) captures the majority of the improvement observed in the evolved strain, indicating that these indels play an important role in enabling the increases in HB production (right). (B) Introduction of the *n*-butanol pathway into DH1Δ*5*.*pcnB*(R149L).*rpoC*(M366L) shows that some aspects of this phenotype can be transferred to other pathways.

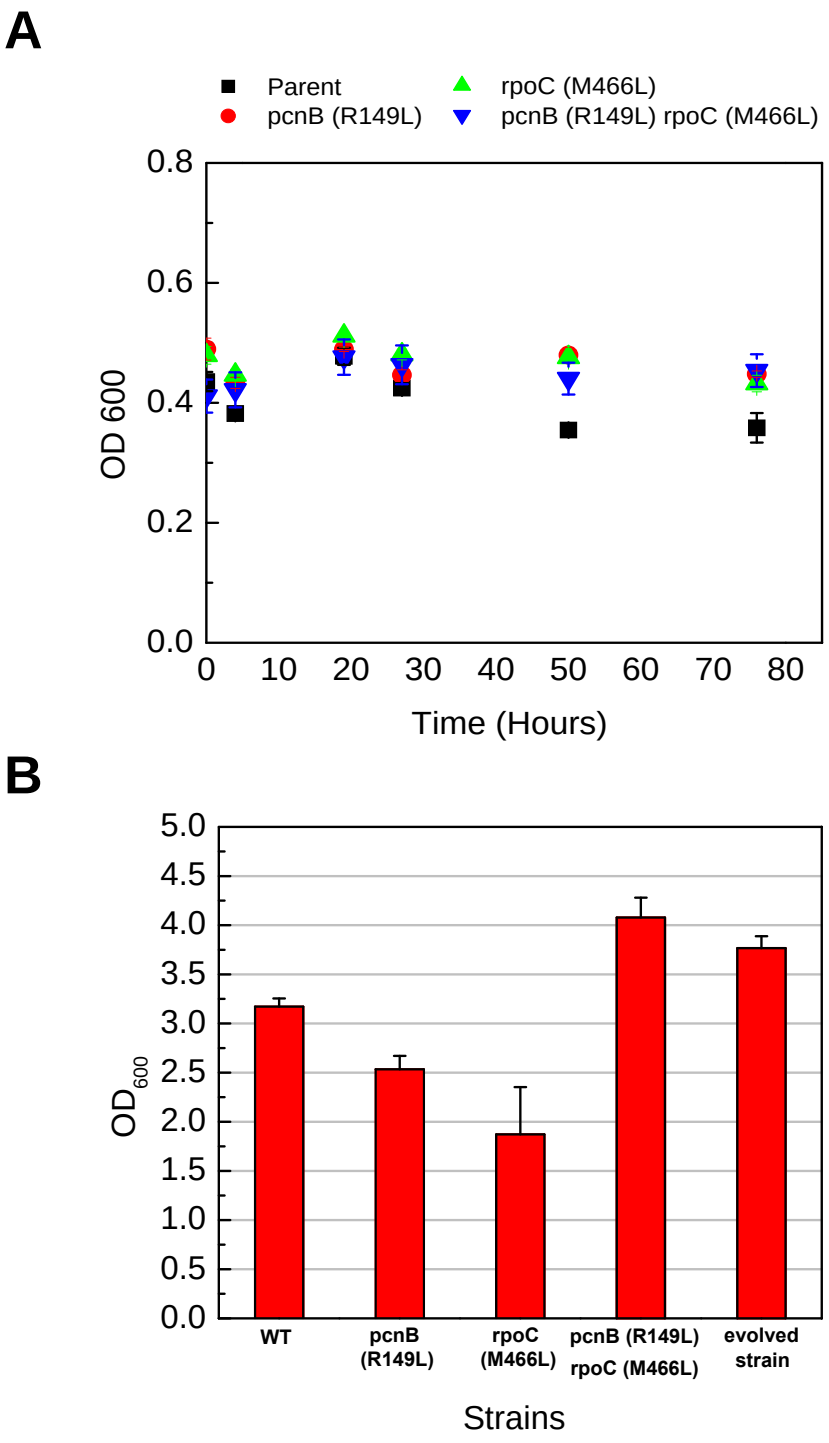
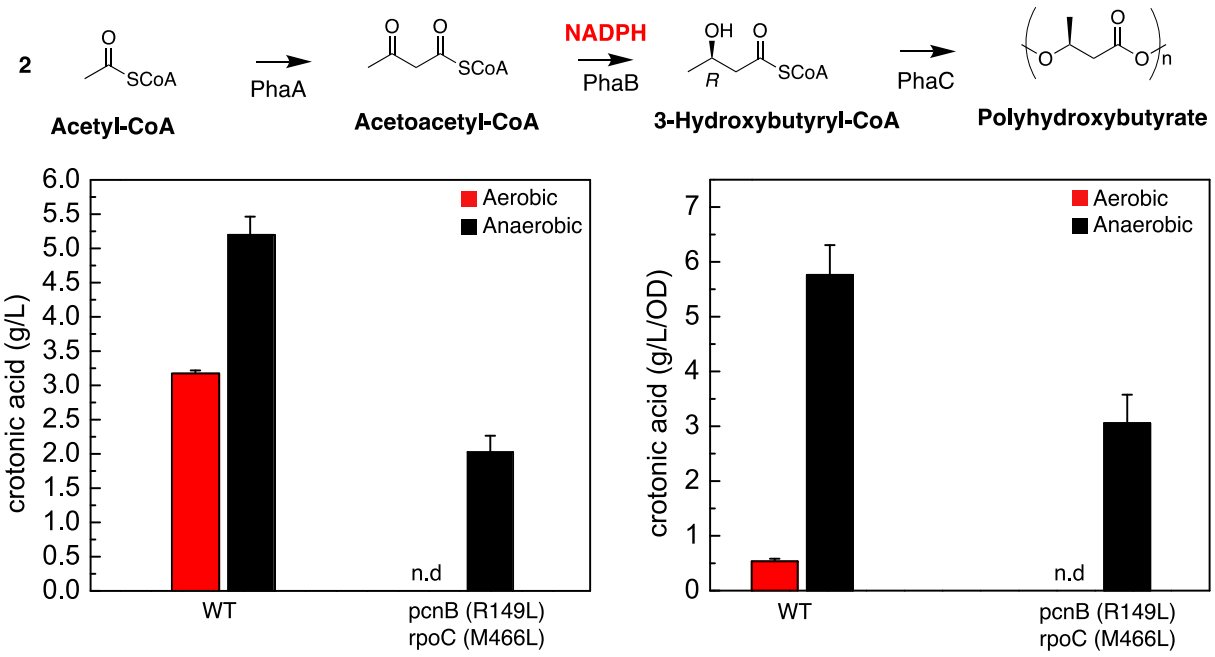
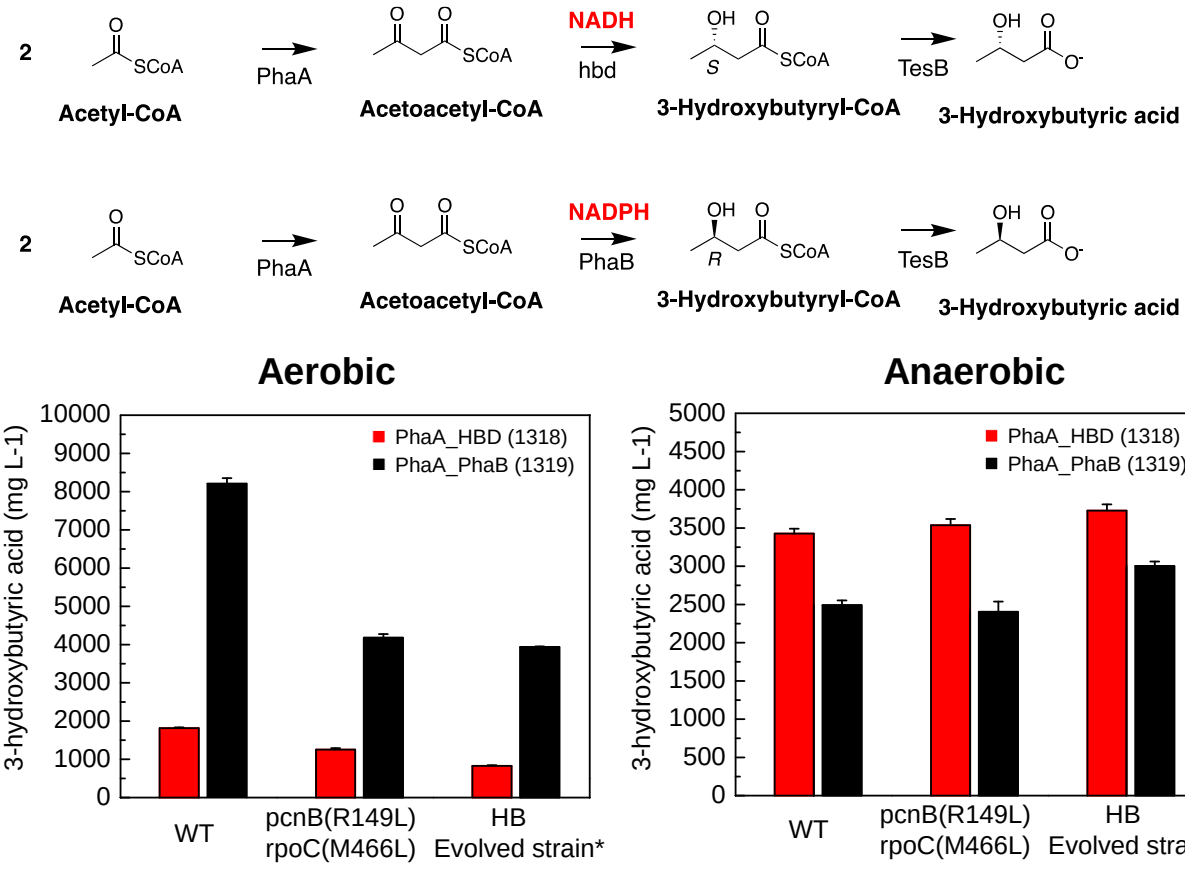


Figure 3.6. Physiology studies of the parent strains and *pcnB*(R149L) and *rpoC*(M466L) mutants. (A) Cell growth with host only. (B) Cell growth with the BDO pathway.

A


2
SCoA
Acetyl-CoA
PhaA
SCoA
Acetoacetyl-CoA
NADPH
PhaB
OH
O
R
SCoA
3-Hydroxybutyryl-CoA
PhaC
O
O
O
n
Polyhydroxybutyrate
crotonic acid (g/L)
Aerobic
Anaerobic
n.d
WT
pcnB (R149L)
rpoC (M466L)
crotonic acid (g/L/OD)
Aerobic
Anaerobic
n.d
WT
pcnB (R149L)
rpoC (M466L)

B


2
SCoA
Acetyl-CoA
PhaA
SCoA
Acetoacetyl-CoA
NADH
hbd
OH
O
S
SCoA
3-Hydroxybutyryl-CoA
TesB
OH
O
O-
3-Hydroxybutyric acid
2
SCoA
Acetyl-CoA
PhaA
SCoA
Acetoacetyl-CoA
NADPH
PhaB
OH
O
R
SCoA
3-Hydroxybutyryl-CoA
TesB
OH
O
O-
3-Hydroxybutyric acid
Aerobic
3-hydroxybutyric acid (mg L-1)
PhaA_HBD (1318)
PhaA_PhaB (1319)
WT
pcnB(R149L)
rpoC(M466L)
HB
Evolved strain*
Anaerobic
3-hydroxybutyric acid (mg L-1)
PhaA_HBD (1318)
PhaA_PhaB (1319)
WT
pcnB(R149L)
rpoC(M466L)
HB
Evolved strain*

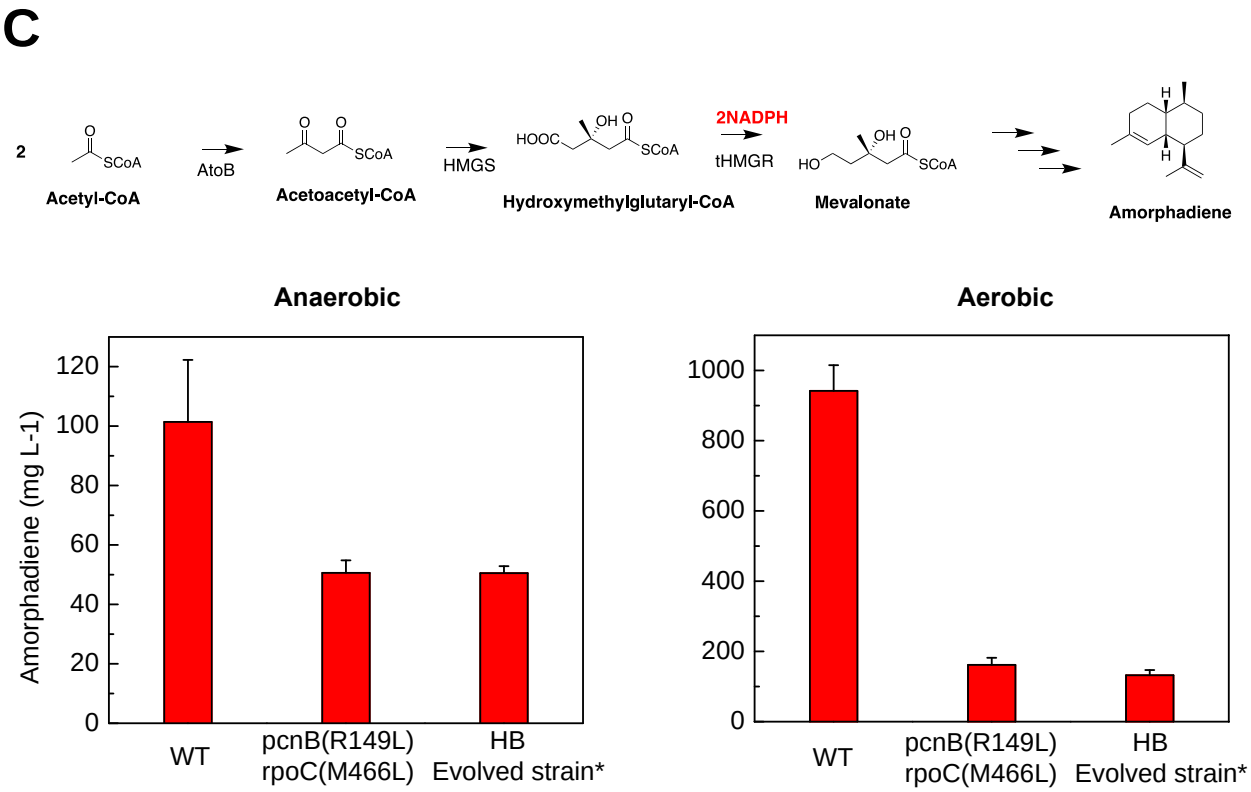


Figure 3.7. Production profile with key mutants from evolved strains. (A) The PHB pathway. pBT33-phaA.phaB.phaC (#2692). Production were conducted under TB media with 2.5% (*w/v*) glucose supplemented. Cultures were induced with 0.2% L arabinose and grew for 5 days before harvest for product quantification [*11*]. (B) The hydroxy acid pathway. The upper one: pT533-phaA.HBD (#1318), pX_Ter.tesB (#2717), and the bottom pathway: pT533-phaA.PhaB (#1319), pX_Ter.tesB (#2717). Production were conducted under TB media with 2.5% (w/v) glucose supplemented. Cultures were induced with 1 mM IPTG and grew for 5 days before harvest for product quantification. (C) Pathway encodes for the production of amorphadiene which consists for the following plasmids: pAM45 (#139) and pTrc-sADS (#122). Production were conducted under TB media with 2.5% (w/v) glucose supplemented. Cultures were induced with 1 mM IPTG and grew for 5 days before harvest for product quantification.

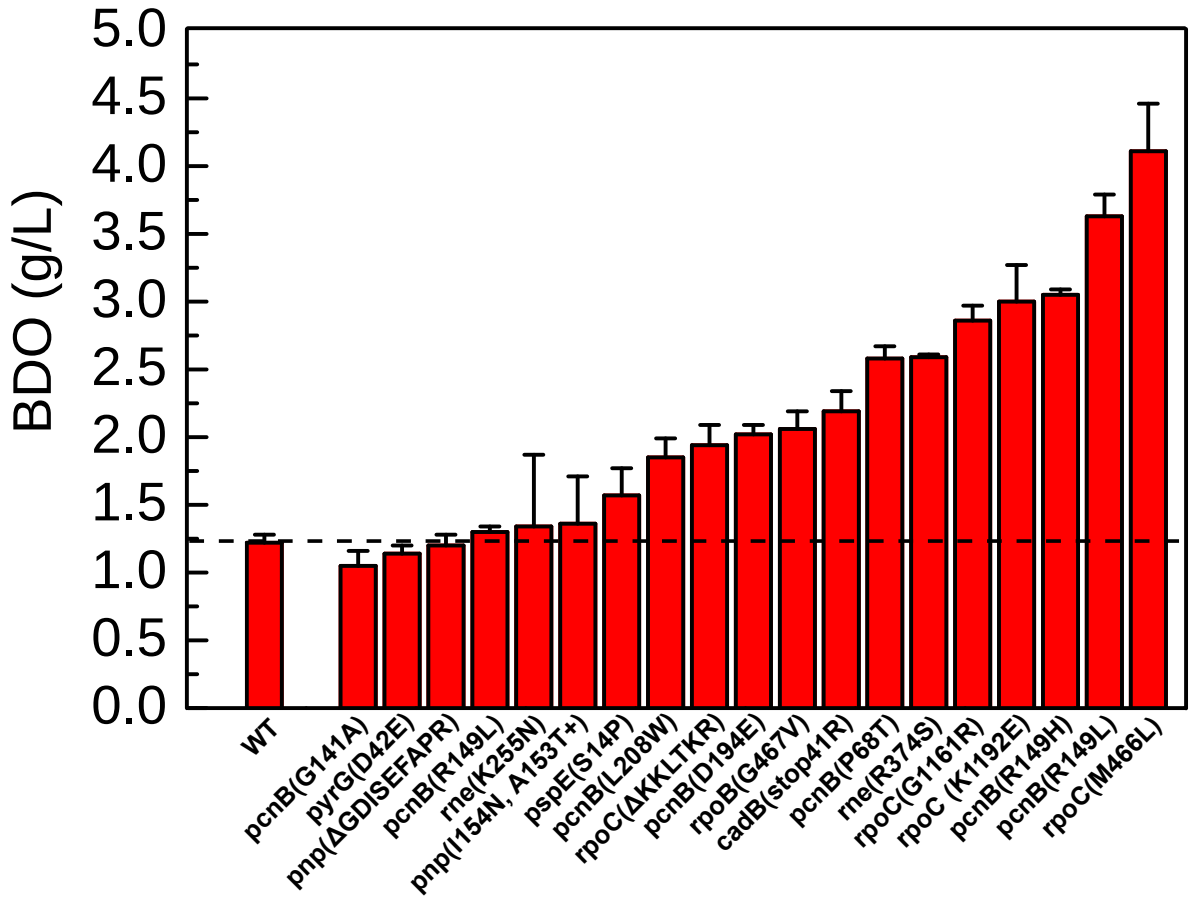


Figure 3.8. BDO production with strains that carried mutations from evolved pathways. All these strains were carried the following plasmids that correspond to the BDO pathway: pT533-phaA.phaB (#1319) and pCWO.trc-TdTer-aldh7.adh2 (#2076). Production were conducted in TB with 2.5% (*w/v*) glucose. Cultures were grown for 5 d before harvesting for product quantification.

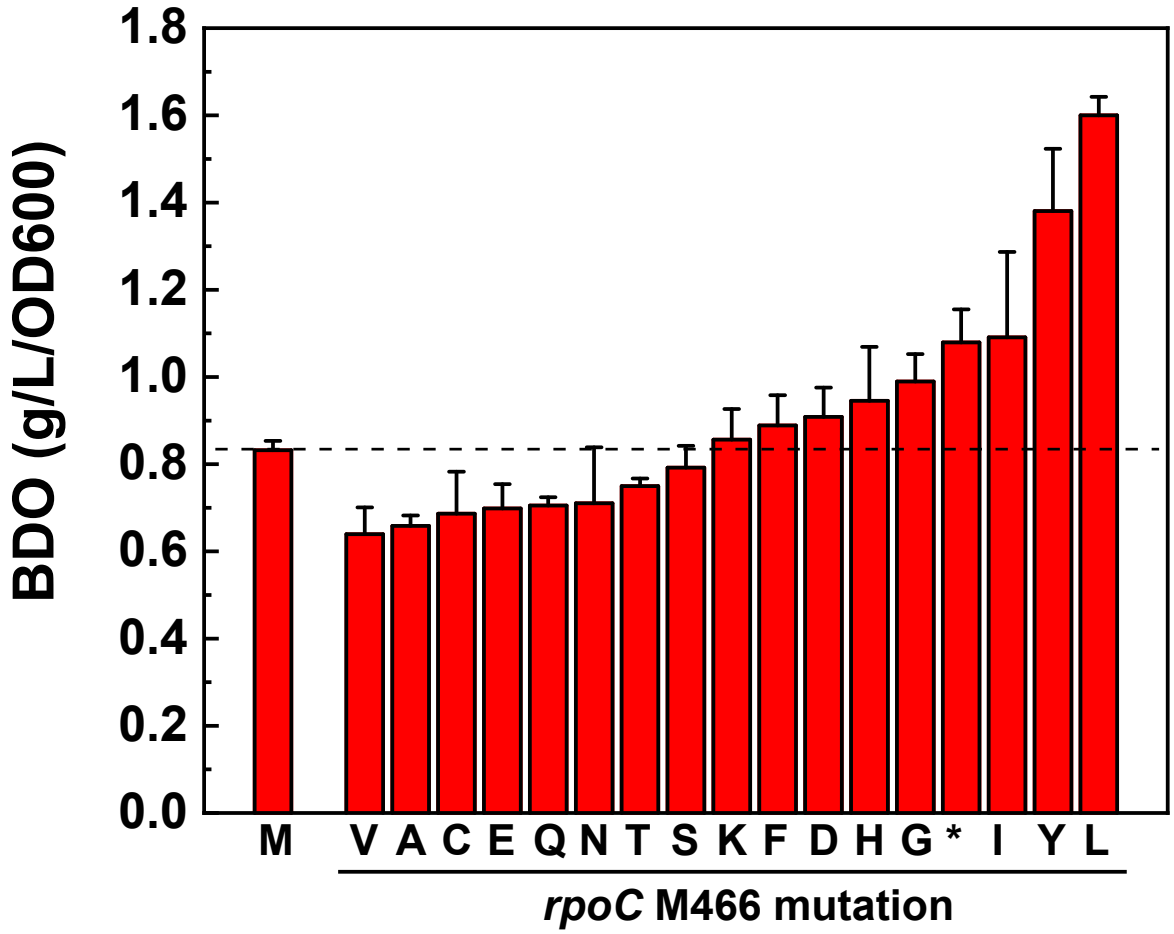


Figure 3.9. BDO production with the NNK library of rpoC M466. All these strains were carried the following plasmids that correspond to the BDO pathway: pT533-phaA.phaB (#1319) and pCWO.trc-TdTer-aldh7.adh2 (#2076). Production were conducted in TB with 2.5% (*w/v*) glucose. Cultures were grown for 5 d before harvesting for product quantification.

Exploring *pcnB*, *rpoC, rpoB, and rne* as targets for metabolic reprogramming. Given the results with the three biosynthetic pathways above, we designed a simpler experiment to examine the relationship between the different point mutants uncovered by the selection and potential differences in phenotype with respect to BDO production. We decided to reconstruct the mutations in the coding region from the 31 sequenced strains that were isolated from the BDO, HB, and *n*-butanol adaptive evolutions (*Table 2.3*). A total of 19 mutant strains were generated and characterized for BDO production (*Figure 3.8*). From this library, we have found 12 mutants showed a positive effect on BDO production. Interestingly, these single mutations display a range of effects on productivity up to four-fold with the best performer, *rpoC*(M466L). From this small screen, it appears as if mutations in *rpoC*, *pcnB*, and *rne* may have the largest general impact on BDO yield (*Figure 3.8*). This initial screen indicates that engineering these three global RNA processors may provide a useful platform for reprograming cell behavior.

Furthermore, these production experiments have shown *rpoC*(M466L) is the best performer, which arose from the BDO evolution experiment. We decided to mutate M466 to other amino acids and examine the corresponding production profiles. These mutants were constructed by the NNK library. A total of 17 mutants were isolated from the library. Among these 17 mutants, one of them encoded the stop codon, and the M466P, M466R, and M466W mutants were not obtained. These mutants were transformed with BDO pathway and conducted production. The *rpoC*(M466L) is the best performer measured by BDO production titer compared to the other 17 mutants (*Figure 3.9*). This highlights the power of evolution. Although the *rpoC*(M466L) is the best for the BDO production, it would be interesting to examine the production profile for the PHB, hydroxy acid, and the isoprenoid pathway with other 17 mutants.

3.4. Conclusion

Combining rational design and adaptive evolution, we have developed a system where adaptive evolution can be used to overcome and break homeostasis of carbon flux. Genome sequencing of 31 strains derived from three different pathways revealed that these phenotypes predominantly arise from point mutations in the global RNA processors, *rpoC*, *pcnB*, and *rne*, giving rise to the hypothesis that large-scale changes at the transcript level provide the necessary synergy to achieve global changes in carbon metabolism. RNA sequencing experiments of two different strains showed that a moderate number of changes are found (49 and 126 differentially expressed genes compared to their respective parents), indicating that alterations in the transcriptional landscape may be well balanced to enable systems-level changes in the tightly coupled carbon network while avoiding toxicity arising from too many changes. Consistent with this proposal, the functional categories of the differentially expressed genes found in this study are spread across a broad range of function.

Interestingly, the transcriptional profiles of these two strains differ greatly, raising the possibility that the microscopic metabolic states of these strains could differ even though the same outcome of high productivity is achieved. In order to further explore this possibility, we carried out metabolomics experiments on mutants from strains from each pathway, which showed that metabolite levels, energy charge, and redox state differ from strain to strain. Interestingly, even the levels of the shared building block, acetyl-CoA, span a range from similar to the parent strain up to 25-fold greater than the parent.

By making mutations in a clean background, we have validated that a large part of the phenotype can be recapitulated by just two mutations in *pcnB* and *rpoC*. In some cases, this phenotype can be transferred to another pathway from this family. However, the specific mutations do not appear to transfer directly to other acetyl-CoA-dependent pathways, such as those for the production of PHAs, isoprenoids, or 3-hydroxy acids. Altogether, these results suggest that the relationship between the metabolic microstate of each strain and the phenotype of high product yield is complex. As such, we believe that the profiling and study of these different strains can provide valuable new information about how carbon flux and metabolism are regulated. In addition, preliminary studies of the different mutations identified in this study show that even one mutation in these RNA processors is sufficient to see large gains, implying that like transcription factors [*17*]. They may be good candidates explore for systems-level engineering of cell behavior using a limited number control factors.

3.5. References

1. Nielsen, J.; Keasling, J. D. Engineering Cellular Metabolism. *Cell* **2016**, *164*, 1185–1197.
2. Liao, J. C.; Mi, L.; Pontrelli, S.; Luo, S. Fuelling the Future: Microbial Engineering for the Production of Sustainable Biofuels. *Nat Rev Microbiol* **2016**, *14*, 288–304.
3. Paddon, C. J.; Westfall, P. J.; Pitera, D. J.; Benjamin, K.; Fisher, K.; McPhee, D.; Leavell, M. D.; Tai, A.; Main, A.; Eng, D.; et al. High-Level Semi-Synthetic Production of the Potent Antimalarial Artemisinin. *Nature* **2013**, *496*, 528–532.
4. Brockman, I. M.; Prather, K. L. J. Dynamic Metabolic Engineering: New Strategies for Developing Responsive Cell Factories. *Biotechnol. J.* **2015**, *10*, 1360–1369.
5. Huttanus, H. M.; Feng, X. Compartmentalized Metabolic Engineering for Biochemical and Biofuel Production. *Biotechnology Journal*. 2017.
6. Davis, Matthew A. ; Chang, M. C. Y. Exploring in Vivo Biochemistry with C4 Fuel and Commodity Chemical Pathways, UC Berkeley, 2015.
7. Jiang, Y.; Chen, B.; Duan, C.; Sun, B.; Yang, J.; Yang, S. Multigene Editing in the *Escherichia Coli* Genome via the CRISPR-Cas9 System. *Appl. Environ. Microbiol.* **2015**, *81*, 2506–2514.
8. Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6*, 343–345.
9. Jiang, W.; Bikard, D.; Cox, D.; Zhang, F.; Marraffini, L. A. RNA-Guided Editing of Bacterial Genomes Using CRISPR-Cas Systems. *Nat. Biotechnol.* **2013**, *31*, 233–239.
10. Chen, W.; Zhang, Y.; Yeo, W.-S.; Bae, T.; Ji, Q. Rapid and Efficient Genome Editing in *Staphylococcus Aureus* by Using an Engineered CRISPR/Cas9 System. *J. Am. Chem. Soc.* **2017**, *139*, 3790–3795.
11. Liu, C.; Gallagher, J. J.; Sakimoto, K. K.; Nichols, E. M.; Chang, C. J.; Chang, M. C. Y.; Yang, P. Nanowire–Bacteria Hybrids for Unassisted Solar Carbon Dioxide Fixation to Value-Added Chemicals. *Nano Lett.* **2015**, *15*, 3634–3639.
12. Blaisse, M. R.; Dong, H.; Fu, B.; Chang, M. C. Y. Discovery and Engineering of Pathways for Production of α-Branched Organic Acids. *J. Am. Chem. Soc.* **2017**, *139*, 14526–14532.
13. Bray, N. L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-Optimal Probabilistic RNA-Seq Quantification. *Nat. Biotechnol.* **2016**, *34*, 525–527.
14. Pimentel, H.; Bray, N. L.; Puente, S.; Melsted, P.; Pachter, L. Differential Analysis of RNA-Seq Incorporating Quantification Uncertainty. *Nat. Methods* **2017**, *14*, 687–690.
15. Markowitz, V. M.; Chen, I.-M. A.; Palaniappan, K.; Chu, K.; Szeto, E.; Grechkin, Y.; Ratner, A.; Jacob, B.; Huang, J.; Williams, P.; et al. IMG: The Integrated Microbial Genomes Database and Comparative Analysis System. *Nucleic Acids Res.* **2012**, *40* (Database issue), D115-22.
16. Conrad, T. M.; Frazier, M.; Joyce, A. R.; Cho, B.-K.; Knight, E. M.; Lewis, N. E.; Landick, R.; Palsson, B. O. RNA Polymerase Mutants Found through Adaptive Evolution Reprogram *Escherichia Coli* for Optimal Growth in Minimal Media. *Proc. Natl. Acad. Sci.* **2010**, *107*, 20500–20505.
17. Alper, H.; Moxley, J.; Nevoigt, E.; Fink, G. R.; Stephanopoulos, G. Engineering Yeast

Transcription Machinery for Improved Ethanol Tolerance and Production. *Science (80-.).* **2006**, *314*, 1565–1568.

Chapter 4. *Engineering Saccharomyces cerevisiae for the production of n-butanol*

4.1. Introduction

Microbial fermentation provides an effective platform for developing single-stage fermentation processes to achieve complex and multi-step synthesis. While there are many possible hosts, *Saccharomyces cerevisiae* (Baker's Yeast) provides both practical and scientific advantages for study. On the practical side, the tools for synthetic pathway construction are quite advanced in the model bacterium, *E. coli* but issues with phage attack and other liabilities create expensive roadblocks for strain commercialization, especially for low cost point, high-volume commodity chemicals. In contrast, *S. cerevisiae* is a preferred industrial host organism that is Generally Regarded As Safe (GRAS) and can also be grown at much lower cost compared to *E. coli*, as it does not require antibiotic selection during fermentation. *S. cerevisiae* can also grow at a low pH which greatly reduces the susceptibility of contamination. In addition, yeast biomass from fermentation can be sold or reused in subsequent fermentations, eliminating expensive disposal costs. On the scientific side, *S. cerevisiae* provides many interesting areas for study when engineering synthetic pathways in this host, based on the need for increased understanding of the requirements for robust heterologous gene expression and eukaryotic compartmentalization of metabolism within different organelles [*1*, *2*].

We approach these questions by constructing a synthetic pathway for *n*-butanol production in *S. cerevisiae* as a model system for examining heterologous protein production and metabolic engineering (*Figure. 4.1*). *n*-Butanol is a second-generation biofuel [*3*], with improved properties compared to bioethanol. It is also the immediate precursor to an important C_4 feedstock, 1-butene [*4*]. In addition, the precursor for the *n*-butanol pathway is the central building block, acetyl-CoA. Acetyl-CoA has been reported as the starting precursor for many high value chemicals including, isoprenoids, polyketides, and fatty acids. It has been well documented that cytosolic acetyl-CoA pool is limited in *S. cerevisiae*, making it challenging to engineer high flux acetyl-CoA dependent pathways. Thus, using the chimeric *n*-butanol pathway as model, we could synthesize a high value chemical from renewable feedstocks. Furthermore, we would gain knowledge on both fundamental understanding on heterologous protein expression and improving carbon flux to acetyl-CoA in *S. cerevisiae*, which could be adapted to optimize other synthetic pathways.

With *E. coli* as a host, near quantitative yields have been achieved from glucose at titers >8000 mg/L at the lab-scale (*Chapter 2*), which is industrially relevant [*5*]. However, product titers drop over three orders of magnitude when the same pathway was introduced into *S. cerevisiae*. Preliminary experiments indicate that this drop is related to low heterologous protein production. We therefore used this pathway to explore different factors that affect product titer in *S. cerevisiae* with the long-term goal of developing a framework for understanding heterologous gene expression and post-transcriptional gene regulation in *S. cerevisiae* (*Figure 4.2*). We focused both on known factors, optimizing codon usage, promoters, 5′- and 3′-untranslated regions (UTRs), and enzyme homologs, as well as on elucidating the molecular mechanisms that lead to high translational efficiency and by which poorly expressed transcripts are derailed. Our strategy was to quantify the behavior of highly-expressed native yeast transcripts as compared to non-native transcripts and begin identifying factors in both the coding and non-coding regions of the transcript that affect the efficiency of various steps in mRNA processing, translation, and protein quality control.

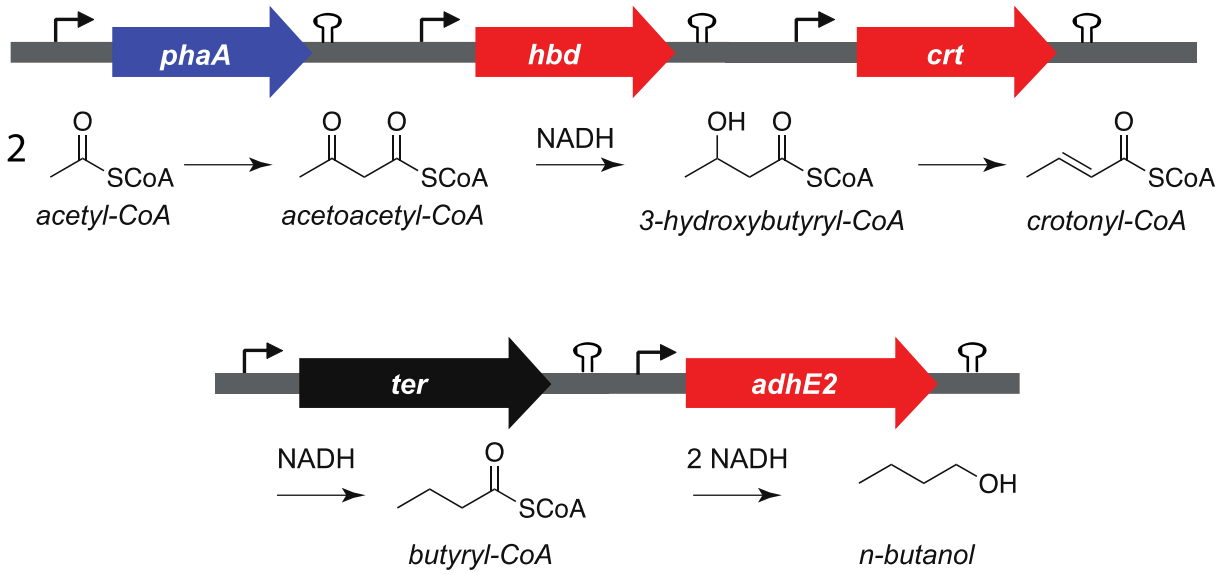


Figure 4.1. *n*-Butanol pathway assembled from three different organisms. The *n*-butanol pathway consists of five heterologous expressed genes from a broad range of microbial hosts (blue, *R. eutrophus*; red, *C. acetobutylicum*; black, *T. denticola*). *n*-Butanol is produced by the condensation of two monomers (acetyl-CoA) and subsequent rounds of reduction and dehydration. *phaA*, acetoacetyl-CoA thiolase/synthase; *hbd*, 3-hydroxybutyryl-CoA dehydrogenase; *crt,* crotonase*; ter,* trans-enyol-CoA reductase*; adhE2,* bifunctional butyraldehyde and butanol dehydrogenase.

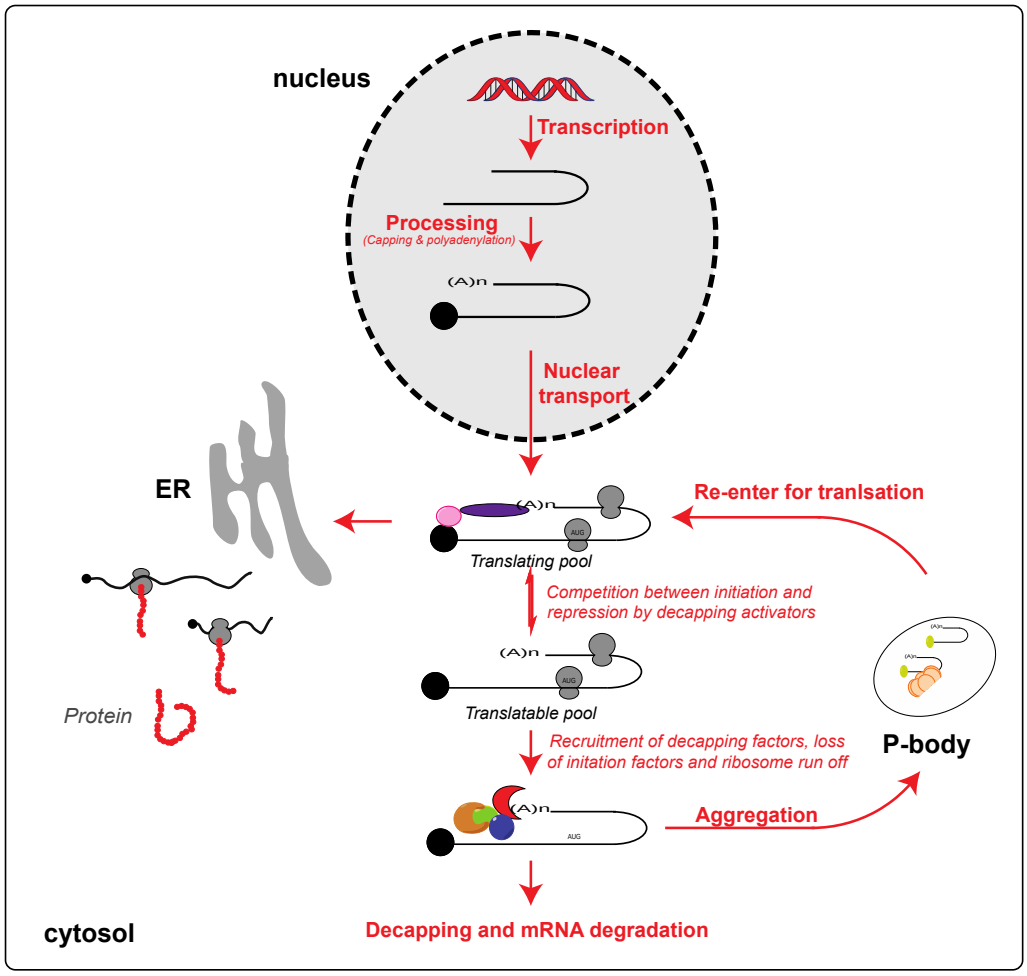


Figure 4.2. Schematic of post-transcriptional processing of eukaryotic mRNAs. RNAs were synthesized and modified in the nucleus. Matured mRNAs are then transported to the from the nucleus to the cytosol for downstream processing. The fate of transcripts is determined by the recruitment of additional factors. Transcripts can either enter for translation upon the binding of translation initiation factors, or targeted for degradation or P-body aggregation when initiation factors were lost or recruitment of decapping factors.

4.2 Materials and methods

Commercial materials. Terrific Broth (TB), LB Broth Miller (LB), LB Agar Miller, and glycerol, and methylsulfoxide (DMSO) were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), Kanamycin (Km), chloramphenicol (Cm), isopropyl-ß-D-thiogalactopyranoside (IPTG), phenylmethanesulfonyl fluoride (PMSF), tris (hydroxymethyl) aminomethane hydrochloride (Tris-HCl), sodium chloride, dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), and magnesium chloride hexahydrate were purchased from Fisher Scientific (Pittsburgh, PA). Nourseothricin Sulfate (Streptothricin Sulfate) (Nat) was purchased from Gold Biotechnology (St. Louis, MO). Cycloheximide was purchased from Sigma-Aldrich (St. Louis, MO). Tris-(2-carboxyethyl)phosphine hydrochloride (TCEP) was purchased from Biosynth, Inc. (Itasca, IL). Imidazole was purchased from Acros Organics (Morris Plains, NJ). Sodium hydroxide was purchased from Avantor Performance Materials (Center Valley, PA). A sodium salt hydrate (CoA), acetyl-CoA, butyryl-CoA, acetoacetyl-CoA, β-nicotinamide adenine dinucleotide reduced dipotassium salt (NADH), β-nicotinamide adenine dinucleotide hydrate (NAD^+), formic acid, trichloroacetic acid (TCA), β-mercaptoethanol (BME), lysozyme from chicken egg white, and bovine serum albumin (BSA) were purchased from Sigma-Aldrich (St. Louis, MO). 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, Phusion DNA polymerase, Q5 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. Amicon Ultra 10,000 centrifugal concentrators were purchased from EMD Millipore (Billerica, MA). cOmplete EDTA-free protease inhibitor was purchased from Roche Applied Science (Penzberg, Germany). Amicon Ultra spin concentrators and MultiScreen$_{HTS}$ 0.22µm filter plates were purchased from Merck Millipore (Cork, Ireland). Ter and AdhE2 antibodies were raised by ProSci Inc. (Poway, CA). Western Lighning Plus-ECL was purchased from PerkinElmer, Inc. (Waltham, MA). SYBR Green Master Mix was purchased from Bio-Rad (Hercules, CA). D-(+)-glucose was purchased from MP Biochemicals (Santa Ana, CA). D(+)-Galactose, 99+%, ACROS Organics™ was purchased from Fischer Chemicals (Pittsburgh, PA). SC powders were purchased from Sunrise Science Products (San Diego, CA). Difco yeast nitrogen base w/o amino acids was purchased from BD Bioscience (San Jose, CA). DNA purification kits, Ni-NTA agarose, genomic DNA isolation, and RNeasy RNA isolation kit were purchased from Qiagen (Valencia, CA). Illumina TruSeq RNA Sample Prep Kit was purchased from Illumina (Hayward, CA).

Host strains. *Escherichia coli* DH10B was used for DNA construction and BL21(de3) Star-T1^R was used for heterologous production of proteins for purification. *Saccharomyces cerevisiae BY4741* (MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) and BY4742 (MATα *his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) were used as the parent for all yeast strains generated in this study. BY4741 was obtained from J. Rine Lab. BY4742 and all heat shock protein knockouts were provided by the J. Thorner

Lab. Protease knockout strains (BJ1991 and BJ5457) were gifts from J. Cate Lab. (*Appendix 4.1*). Additional modifications to these strains were generated using the CRISPR-Cas9 system [*6*].

Construction of plasmids. Plasmid construction was carried out using standard molecular biology techniques using the Gibson protocol [*7*] and found in *Appendix 4.2*. PCR amplifications were carried out with Q5 DNA polymerase or Phusion DNA polymerase, following manufacturer instructions. Oligonucleotide sequences are listed in *Appendix 4.3*. Constructs were verified by sequencing (Quintara Biosciences; Berkeley, CA). Synthetic genes were assembled using gBlock sequences (Integrated DNA Technologies, *Appendix 4.4*). gBlocks were resuspended at 10 ng/μL in 10 mM Tris-HCl, pH 8.5 and used directly for assembly of vectors.

The initial base plasmids were constructed by Dr. Brooks Bond-Watts. pESCHis-Bu2 (#800) contains *phaA, hbd*, and *crt* under the control of the *S. cerevisiae adh1, tef1, pdc* promoters, respectively. pESCLeu2d-ter.adhE2 (#795) contains ter and adhE2 under the control of the *S. cerevisiae gal10* and *gal1* promoters, respectively. Additionally, pRS413-Bu2 (#932) contains *phaA, hbd*, and *crt* under the control of the *S. cerevisiae adh1, tef1, pdc* promoters, respectively, with the CEN ARS origin was constructed by Dr. Michael Blaisse. pESCUra-(Pcons)PDCzm.eutE (#903) contains pdc and eutE under the was constructed by FBA1 and PYK1 promoter, respectively with the Ura3 selection and 2 micron origin of replication was constructed by Dr. Michiei Sho.

Constructs for screening thiolase homologs.

pESCHis-Erg10.hbd.crt (#1383). The *erg10* gene (Gene Accession ID NM_001022609.2) was amplified from *Schizosaccharomyces pombe* genomic DNA (ATCC 24843) using primers P1_Erg10F1 and P2_Erg10R1. The *tef1* promoter was amplified from pESCHis-Bu2 (#800) using P3_P(Tef1)F1 and P4_P(Tef1)R1. These two PCR products were used to set up a Gibson reaction with pESC.His-Bu2 (#800) digested with Bam HI HF and Sac I.

pESCHis-Erg10His_{10}.hbd.crt (#1384). The *erg10* gene was amplified from *S. pombe* genomic DNA using P1_Erg10F1 and P23_Erg10_HisR4. The *tef1* promoter was amplified from pESCHis-Bu2 (#800) using P3_P(Tef1)F1 and P4_P(Tef1)R1. These two PCR products were used to set up a Gibson reaction with pESC.His-Bu2 (#800) digested with Bam HI HF and Sac I.

Constructs for 5'- and 3'-untranslated region (UTR) screening. All constructs were constructed using pESCLeu2d-ter.adhE2 (#795) as the parent using two different approaches. In the first approach, pESCLeu2d-ter.adhE2 (#795) was digested with Not I HF and Spe I HF to remove the Ter cassette. TdTer was then amplified from the same plasmid with primers containing the desired UTR sequences and combined with the parent plasmid by Gibson assembly. In the cases where the UTR sequences were too long, a second approach was used. gBlocks were ordered with a 25 bp overlap with the parent plasmid for direct use in the Gibson reaction. The parent plasmid, pESCLeu2d-ter.adhE2 (#795) was digested with Not I HF and Pst I, removing part of the N-terminal of TdTer. The missing part of N-terminal TdTer was replaced using the gBlock.

pESCLeu2d-AdhE2.(5'UTR-TPI1)TdTer (#1413). Primers P30_5'UTRTPI1_F and P37_5'UTR R were used to amplify TdTer and combined in a Gibson assembly with the Not I- and Spe I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-TDH2-YJR009C)TdTer (#1414). Primers P31_5'UTRTDH2_F and P37_5'UTR R were used to amplify TdTer and combined in a Gibson assembly with the Not I- and Spe I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-FBA1-YKL060C)TdTer (#1415). Primers P32_5'UTRFBA1_F and P37_5'UTR R were used to amplify TdTer and combined in a Gibson assembly with the Not I- and Spe I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-GPM1-YKL152C)TdTer (#1416). Primers P33_5'UTRGPM1_F and P37_5'UTR R were used to amplify TdTer and combined in a Gibson assembly with the Not I- and Spe I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-YLR075W)TdTer (#1417). Primers P34_5'UTRYLRO75W_F and P37_5'UTR R were used to amplify TdTer and combined in a Gibson assembly with the Not I- and Spe I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-YHL001W)TdTer (#1418). Primers P35_5'UTRYHL001W_F and P37_5'UTR R were used to amplify TdTer and combined in a Gibson assembly with the Not I- and Spe I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-YJL177W)TdTer (#1419). Primers P36_5'UTRYJL177W_F and P37_5'UTR R were used to amplify TdTer and combined in a Gibson assembly with the Not I- and Spe I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.TdTer(3'UTR-FBA1) (#1424). Primers P39_3'UTR F and P38_3'UTR FBA1R were used to amplify TdTer and combined in a Gibson assembly with the Not I- and Spe I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.TdTer(3'UTR-YJL177W) (#1425). Primers P39_3'UTR F and P44_3'UTR YJL177WgDNA_TerR were used to amplify TdTer. Primers P43_3'UTR YJL177WgDNA_R and P42_3'UTR YJL177WgDNA_F were used to amplify the YJL177W 3'–UTR from *S. cerevisiae* genomic DNA. These two PCR products were used in a Gibson assembly with the Not I- and Spe I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-FBA)TdTer(3'UTR-FBA1) (#1426). Primers P38_3'UTR FBA1R and P32_5'UTRFBA1_F were used to amplify TdTer and combined in a Gibson assembly with the Not I- and Spe I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-FBA)TdTer(3'UTR-YJL177W) (#1427). Primers P32_5'UTRFBA1_F and P44_3'UTR YJL177WgDNA_TerR were used to amplify TdTer. Primers P43_3'UTR YJL177WgDNA_R and P42_3'UTR YJL177WgDNA_F the YJL177W

3'–UTR from *S. cerevisiae* genomic DNA. The two PCR products were used in a Gibson assembly with the Not I- and Pst I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-TDH1)TdTer (#1453). gBlock 5'UTR_TDH1TdTer was used in a Gibson assembly with the Not I and Pst I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454). gBlock 5'UTR_PYK2TdTer was used in a Gibson assembly with the Not I and Pst I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-PGI1)TdTer (#1455). gBlock 5'UTR_PGI1TdTer was used in a Gibson assembly with the Not I and Pst I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-PFK1)TdTer (#1456). gBlock 5'UTR_PFK1TdTer was used in a Gibson assembly with the Not I- and Pst I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-PFK2)TdTer (#1457). gBlock 5'UTR_PFK2TdTer was used in a Gibson assembly with the Not I- and Pst I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2. (5'UTR-ENO1)TdTer (#1458). gBlock 5'UTR_ENO1TdTer was used in a Gibson assembly with the Not I- and Pst I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-ENO2)TdTer (#1459). gBlock 5'UTR_ENO2TdTer was used in a Gibson assembly with the Not I and Pst I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.(5'UTR-CDC19)TdTer (#1460). Primers P51_5'UTR CDC19_F and P37_5'UTR R were used to amplify TdTer and combined in a Gibson assembly with the Not I- and Spe I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-AdhE2.5'UTR-TDH3_TdTer (#1464). Primers P52_5'UTR TDH3_F and P37_5'UTR R were used to amplify TdTer and combined in a Gibson assembly with the Not I- and Spe I-digested pESCLeu2d-ter.adhE2 (#795) backbone.

pESCLeu2d-(5'UTR-PYK2)AdhE2.(5'UTR-PYK2)TdTer (#2401). Primers P657_YPK2_AdhE2_R and P656_YPK2_AdhE2_F were used to amplify AdhE2 and combined in a Gibson assembly with the Sma I-digested pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454) backbone.

Plasmids for promoter and codon usage screening. Gene sequences were optimized using either *S. cerevisiae* standard (sTdTer) or glycolytic codon usage (sTdTer(gly) and sAdhE2(gly)).

pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)TdTer (#1525). pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454) was digested with Bam HI and Not I to remove both the GAL1 and GAL10 promoters. Primers P84_pCCW12 for 1558 F and P63_gal1454_TDH3_R were used to amplify the intact CCW12 and GAL1 promoter fragment from pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)sTdTer(gly) (#1556) and combined in a Gibson assembly with the digested backbone.

pESCLeu2d-AdhE2.TDH3p(5'UTR-PYK2)TdTer (#1534). pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454) was digested with Bam HI and Not I to remove both the GAL1 and GAL10 promoters. Primers P62_gal1454_TDH3_F and P63_gal1454_TDH3_R were used to amplify the intact TDH3 and GAL1 promoter fragment from pESCLeu2d-AdhE2.TDH3p(5'UTR-PYK2)sTdTer(gly) (#1557) and combined in a Gibson assembly with the digested backbone.

pESCLeu2d-(CCW12p)TdTer-(TDH3p)ALD5-(FBA1p)ADH2 (#2391). pVYY1.5.1 (#1998) was digested with Pvu I HF and Bam HI to obtain a fragment containing TdTer, ALD5, and ADH2 with the corresponding promoters and terminators. Primers P638-Leu_BackbondR and P639_903_eutE_Seq were used to amplify pESC-Leu2d (#69) to obtain the backbone and combined in a Gibson assembly with the PCR product.

pESCLeu2d-AdhE2.(5'UTR-PYK2)sTdTer(gly) (#1551). TdTer codon-optimized for *S. cerevisiae* glycolytic codon usage was ordered in two gBlocks with the PYK2 5'-UTR (g21_TdTer (S.c gly) with 5'UTR PYK2 gBlock 1 and g22_TdTer (S.c gly) with 5'UTR PYK2 gBlock 2) and used in a Gibson assembly with Not I HF- and Spe I-digested pESCLeu2d.ter-adhE2 (#795).

pESCLeu2d-AdhE2.(5'UTR-PYK2)sTdTer (#1552). TdTer codon-optimized for *S. cerevisiae* standard codon usage was ordered in two gBlocks with the PYK2 5'-UTR (g23_TdTer (S.c) with 5'UTR PYK2 gBlock 1 and g24_TdTer (S.c) with 5'UTR PYK2 gBlock 2) and used in a Gibson assembly with Not I HF- and Spe I-digested pESCLeu2d.ter-adhE2 (#795).

pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)sTdTer(gly) (#1556). pESCLeu2d-AdhE2.(5'UTR-PYK2)sTdTer(gly) (#1551) was digested with Not I HF and Bam HI to remove both the GAL1 and GAL10 promoters. Primers 66_gal1454_CCW12_F and 63_gal1454_TDH3_R were used to amplify GAL1p from pESCLeu2d.ter-adhE2 (#795). Primers 84_pCCW12 for 1558 F and 85_pCCW12 for 1558 R were used to amplify CCW12p from *S. cerevisiae* genomic DNA. These two PCR products were used in a Gibson assembly with the digested backbone.

pESCLeu2d-AdhE2.TDH3p(5'UTR-PYK2)sTdTer(gly) (#1557). pESCLeu2d-AdhE2.(5'UTR-PYK2)sTdTer(gly) (#1551) was digested with Not I HF and Bam HI to remove both the GAL1 and GAL10 promoters. Primers 62_gal1454_TDH3_F and 63_gal1454_TDH3_R. pTDH3 were used to amplify GAL1p from pESCLeu2d.ter-adhE2 (#795). Primers 64_TDH3l1454_TDH3_F and 65_TDH3l1454_TDH3_R were used to amplify TDH3p from *S. cerevisiae* genomic DNA. These two PCR products were used in a Gibson assembly with the digested backbone.

pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)sTdTer (#1558). pESCLeu2d-AdhE2.(5'UTR-PYK2)sTdTer (#1552) was digested with Not I HF and Bam HI to remove both the GAL1 and GAL10 promoters. Primers 66_gal1454_CCW12_F and 63_gal1454_TDH3_R were used to amplify GAL1p from pESCLeu2d-sAdhE2(gly).TDH3p(5'UTR-PYK2)TdTer (#1557). Primers 84_pCCW12 for 1558 F and 85_pCCW12 for 1558 R were used to amplify CCW12p

from *S. cerevisiae* genomic DNA. These two PCR products were used in a Gibson assembly with the digested backbone.

pESCLeu2d-AdhE2.TDH3p(5'UTR-PYK2)sTdTer (#1559). pESCLeu2d-AdhE2.(5'UTR-PYK2)sTdTer (#1552) was digested with Not I HF and Bam HI to remove both the GAL1 and GAL10 promoters. Primers 62_gal1454_TDH3_F and 63_gal1454_TDH3_R. pTDH3 were used to amplify GAL1p from pESCLeu2d-sAdhE2(gly).TDH3p(5'UTR-PYK2)TdTer (#1557). Primers 64_TDH3l1454_TDH3_F and 65_TDH3l1454_TDH3_R were used to amplify TDH3p from *S. cerevisiae* genomic DNA. These two PCR products were used in a Gibson assembly with the digested backbone.

pESCLeu2d-sAdhE2(gly).TDH3p(5'UTR-PYK2)TdTer (#1560). AdhE2 codon-optimized for *S. cerevisiae* glycolytic codon usage was ordered in six gBlocks (Adhe2_YCO_G1 - Adhe2_YCO_G6) along with 30 bp upstream and downstream homology with the cut sites of the backbone plasmid. These gblocks were used in a Gibson assembly with Xho I- and Xma I-digested pESCLeu2d.ter-adhE2 (#795).

Constructs for Ter homolog screening.

pESCLeu2d-Adhe2.EgTer (#1124). This plasmid was constructed by Dr. Michael Blaisse with TdTer replaced with the native gene sequence for the Ter homolog from *Euglena gracilis* (EgTer, ATCC 12716) in pESCLeu2d-ter.adhE2 (#795)[8].

pESCLeu2d-Adhe2.sEgTer(EC) (#1067). This plasmid was constructed by Dr. Michei Sho with TdTer replaced with the synthetic gene sequence for the Ter homolog from *Euglena gracilis* (EgTer) optimized for *E. coli* codon usage (*Appendix 4.6*).

pESCLeu2d-AdhE2.sEgTer(YCO) (#1328). EgTer codon-optimized for *S. cerevisiae* standard codon usage was ordered in three gBlocks (EgTer_Yeast_G1, EgTer_Yeast_G2, EgTer_Yeast_G3) with 40 bp overlap and used in a Gibson assembly with Spe I- and Not I-digested pESCLeu2d-Adhe2.EgTer (#1124).

pESCLeu2d-AdhE2.MECR1 (#1428). Primers P45_MECR1_F and P49_MECR1_R were used to amplify MECR1 from *Euglena gracilis* from pET16b-EgMECR1 (#1424) [*8*] [and used in a Gibson assembly with Spe I- and Not I- digested pESCLeu2d-Adhe2.EgTer (#1124).

pESCLeu2d-AdhE2.His_{10}MECR1 (#1429). Primers P46_MECR1 His_F and P49_MECR1_R were used to amplify MECR1 from *Euglena gracilis* (ATCC 12716) from pET16b-EgMECR1 (#1424) [*8*] and used in a Gibson assembly with Spe I- and Not I- digested pESCLeu2d-Adhe2.EgTer (#1124).

Constructs for Aldh and Adh homolog screening. All ALDHs and ADHs were amplified from the collection of Dr. Matthew Davis (*Appendix 2.5, 2.6*) [9]. The pESCLeu2d-(5'UTR-PYK2)TdTer.Aldh21.Adh2 (#2759) parent plasmid was constructed by Dr. Zhen Wang and generated by removing the adhE2 cassette from pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454) and replacing it with the gal1p-ADH2.gal7p-ALDH21 cassettes.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh21.Adh plasmids for ADH screening. pESCLeu2d-(5'UTR-PYK2)TdTer.Aldh21.Adh2 (#2759) was digested with Bam HI and Apa I in order to insert various Adh genes between GAL1p and the TPS3 terminator by Gibson assembly.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh21.Adh3 (#2796). ADH3 was amplified using P1206_ADH3_aldh21_F and P1207_ADH3_aldh21_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh21.Adh4 (#2797). ADH4 was amplified using P1208_ADH4_aldh21_F and P1209_ADH4_aldh21_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh21.Adh5 (#2798). ADH5 was amplified using P1210_ADH5_aldh21_F and P1211_ADH5_aldh21_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh21.Adh6 (#2799). ADH6 was amplified using P1212_ADH6_aldh21_F and P1213_ADH6_aldh21_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh21.Adh7 (#2800). ADH7 was amplified using P1214_ADH7_aldh21_F and P1215_ADH7_aldh21_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh21.Adh9 (#2801). ADH9 was amplified using P1216_ADH9_aldh21_F and P1217_ADH9_aldh21_R.

pESCLeu2d-(5'UTR-PYK2)TdTer.Aldh21.Adh10 (#2802). ADH10 was amplified using P1218_ADH10_aldh21_F and P1219_ADH10_aldh21_R.

pESCLeu2d-(5'UTR-PYK2)TdTer.Aldh21.Adh12 (#2803). ADH12 was amplified using P1220_ADH12_aldh21_F and P1221_ADH12_aldh21_R.

pESCLeu2d-(5'UTR-PYK2)TdTer.Aldh21.Adh13 (#2804). ADH13 was amplified using P1222_ADH13_aldh21_F and P1223_ADH13_aldh21_R.

pESCLeu2d-(5'UTR-PYK2)TdTer.Aldh21.Adh14 (#2805). ADH14 was amplified using P1224_ADH14_aldh21_F and P1225_ADH14_aldh21_R.

pESCLeu2d-(5'UTR-PYK2)TdTer.Aldh.Adh plasmids for dual Aldh and Adh screening.

pESCLeu2d-(5'UTR-PYK2)Tdter.ADH(AdhE2).Aldh5 (#1574). This plasmid was constructed by Dr. Zhen Wang. It carries the ADH domain from AdhE2 and Aldh5 and was used as the template for the following Aldh and Adh pair cloning by combining the PCR products in a Gibson assembly with Bam HI- and Xho I-digested backbone.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh5.Adh2 (#2556). Adh2 was amplified from pCWO.trc-ter-aldh23.adh2 (#2460) [*9*] with P716_Adh2_Aldh5_F and P717_Adh2_Aldh5_R. The TPS 1 terminator was amplified from plasmid using P718_TPS3t_Adh2_Aldh5_F and P719_TPS3t_Adh2_Aldh5_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh5.Adh8 (#2557). Adh8 was amplified from pCWO.trc-ter-aldh23.adh8 (#2461) [*9*] with P720_Adh8_Aldh5_F and P721_Adh8_Aldh5_R. The TPS 1 terminator was amplified from plasmid #1574 using P722_TPS3t_Adh8_Aldh5_F and P719_TPS3t_Adh2_Aldh5_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh5.Adh22 (#2558). Adh22 was amplified from pCWO.trc-ter-aldh23.adh22 (#2468) [*9*] with P723_Adh22_Aldh5_F and P724_Adh22_Aldh5_R. The TPS 1 terminator was amplified from plasmid #1574 using P725_TPS3t_Adh22_Aldh5_F and P719_TPS3t_Adh2_Aldh5_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.ADH(AdhE2).Aldh6 (#1575). This plasmid was constructed by Dr. Zhen Wang. It carries the ADH domain from AdhE2 and Aldh6 and was used as the template for the following Aldh and Adh pair cloning by combining the PCR products in a Gibson assembly with Bam HI- and Xho I-digested backbone.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh6.Adh2 (#2559). Adh2 was amplified from pCWO.trc-ter-aldh23.adh2 (#2460) [*9*] with P726_Adh2_Aldh6_F and P727_Adh2_Aldh6_R. The TPS 1 terminator was amplified from plasmid #1574 using P728_TPS3t_Adh2_Aldh6_F and P719_TPS3t_Adh2_Aldh5_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh6.Adh8 (#2560). Adh8 was amplified from pCWO.trc-ter-aldh23.adh8 (#2461) [*9*] with P729_Adh8_Aldh6_F and P730_Adh8_Aldh6_R. The TPS 1 terminator was amplified from plasmid #1574 using P731_TPS3t_Adh8_Aldh6_F and P719_TPS3t_Adh2_Aldh5_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh6.Adh22 (#2561). Adh22 was amplified from pCWO.trc-ter-aldh3.adh22 (#2468) [*9*] with P732_Adh22_Aldh6_F and P733_Adh22_Aldh6_R. The TPS 1 terminator was amplified from plasmid #1574 using P734_TPS3t_Adh22_Aldh6_F and P719_TPS3t_Adh2_Aldh5_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.ADH(AdhE2).Aldh7 (#1576). This plasmid was constructed by Dr. Zhen Wang. It carries the ADH domain from AdhE2 and Aldh7 and was used as the template for the following Aldh and Adh pair cloning by combining the PCR products in a Gibson assembly with Bam HI- and Xho I-digested backbone.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh7.Adh2 (#2562). Adh2 was amplified from pCWO.trc-ter-aldh23.adh2 (#2460) [*9*] with P735_Adh2_Aldh7_F and P736_Adh2_Aldh7_R. The TPS 1 terminator was amplified from plasmid #1574 using P737_TPS3t_Adh2_Aldh7_F and P719_TPS3t_Adh2_Aldh5_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh7.Adh8 (#2563). Adh8 was amplified from pCWO.trc-ter-aldh23.adh8 (#2461) [*9*] with P738_Adh8_Aldh7_F and P739_Adh8_Aldh7_R. The TPS 1 terminator was amplified from plasmid #1574 using P740_TPS3t_Adh8_Aldh7_F and P719_TPS3t_Adh2_Aldh5_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh7.Adh22 (#2564). Adh22 was amplified from pCWO.trc-ter-aldh3.adh22 (#2468) [*9*] with P741_Adh22_Aldh7_F and P742_Adh22_Aldh7_R. The TPS 1 terminator was amplified from plasmid #1574 using P743_TPS3t_Adh22_Aldh7_F and P719_TPS3t_Adh2_Aldh5_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.ADH(AdhE2).Aldh10 (#1579). This plasmid was constructed by Dr. Zhen Wang. It carries the ADH domain from AdhE2 and Aldh10 and was used as the template for the following Aldh and Adh pair cloning by combining the PCR products in a Gibson assembly with Bam HI- and Xho I-digested backbone.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh10.Adh2 (#2565). Adh2 was amplified from pCWO.trc-ter-aldh23.adh2 (#2460) [*9*] with P744_Adh2_Aldh10_F and P745_Adh2_Aldh10_R. The TPS 1 terminator was amplified from plasmid #1574 using P746_TPS3t_Adh2_Aldh10_F and P719_TPS3t_Adh2_Aldh5_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh10.Adh8 (#2566). Adh8 was amplified from pCWO.trc-ter-aldh23.adh8 (#2461) [*9*] with P747_Adh8_Aldh10_F and P748_Adh8_Aldh10_R. The TPS 1 terminator was amplified from plasmid #1574 using P749_TPS3t_Adh8_Aldh10_F and P719_TPS3t_Adh2_Aldh5_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh10.Adh22 (#2567). Adh22 was amplified from pCWO.trc-ter-aldh3.adh22 (#2468) [*9*] with P750_Adh22_Aldh10_F and P751_Adh22_Aldh10_R. The TPS 1 terminator was amplified from plasmid #1574 using P752_TPS3t_Adh22_Aldh10_F and P719_TPS3t_Adh2_Aldh5_R.

The pESCLeu2d-(5'UTR-PYK2)TdTer.ADH(AdhE2).Aldh12 (#1581). This plasmid was constructed by Dr. Zhen Wang. It carries the ADH domain from AdhE2 and Aldh12 and was used as the template for the following Aldh and Adh pair cloning by combining the PCR products in a Gibson assembly with Bam HI- and Xho I-digested backbone.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh12.Adh2 (#2568). Adh2 was amplified from pCWO.trc-ter-aldh23.adh2 (#2460) [*9*] with P753_Adh2_Aldh12_F and P754_Adh2_Aldh12_R. The TPS 1 terminator was amplified from plasmid #1574 using P755_TPS3t_Adh2_Aldh12_F and P719_TPS3t_Adh2_Aldh5_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh12.Adh8 (#2569). Adh8 was amplified from pCWO.trc-ter-aldh23.adh8 (#2461) [*9*] with P756_Adh8_Aldh12_F and P757_Adh8_Aldh12_R. The TPS 1 terminator was amplified from plasmid #1574 using P758_TPS3t_Adh8_Aldh12_F and P719_TPS3t_Adh2_Aldh5_R.

pESCLeu2d-(5'UTR-PYK2)Tdter.Aldh12.Adh22 (#2570). Adh22 was amplified from pCWO.trc-ter-aldh3.adh22 (#2468) [*9*] with P759_Adh22_Aldh12_F and P760_Adh22_Aldh12_R. The TPS 1 terminator was amplified from plasmid #1574 using P761_TPS3t_Adh22_Aldh12_F and P719_TPS3t_Adh2_Aldh5_R.

Constructs for multi-component optimization.

pVYY1.0.0_2 (#1799). This plasmid was constructed as a template to screen different UTRs, promoters, and terminators. Unique cut sites were introduced between the promoters and terminators. Backbone plasmid pESCUra (#70) was digested with Bam HI and Hind III. All promoters and terminators were amplified from *S. cerevisiae* genomic DNA. P152_CCW12P_F and P151_CCW12P_R were used to amplify the CCW12 promoter. P441_1.4a.1_PRM9F and P442_1.4a.1_PRM9R were used to amplify the PRM9 terminator. P153_TDH3F and P196_pVYY100_3TDH3R were used to amplify the TDH3 promoter. P193_pVYY100_2SPG5F and P194_pVYY100_2SPG5R were used to amplify SPG5 terminator. All these four PCR products were used in a Gibson assembly with the digested backbone.

pVYY1.0.0.5 (#1879). Aldh5 codon-optimized for *S. cerevisiae* glycolytic codon usage (*Appendix 4.6*) and the HIS5 terminator were inserted between the TDH3 promoter and the SPG5 terminator of pVYY1.0.0.2 (#1799). Aldh5 assembled from two gBlocks (g29_TDH3_ALD5-1_His5 and g30_TDH3ALD5-2_His5). The HIS5 terminator was amplified from *S. cerevisiae* genomic DNA using P246_HIS1 and P247_HIS2. The PCR product and the two gBlocks were used in a Gibson assembly with the Xma I-digested backbone.

pVYY1.C.0 (#1828). sTdTer(gly) was introduced between the CCW12 promoter and the PRM9 terminator of pVYY1.0.0.5 (#1879). sTdTer(gly) was amplified from pESCLeu2d-AdhE2.TDH3p(5'UTR-PYK2)sTdTer(gly) (#1557) using P161_110_PYK2R and P172_1C0_gTdTer and used in a Gibson assembly with the Bam HI-digested backbone.

pVYY1.1.0 (#1821). sTdTer(gly) with the PYK2 5'-UTR was introduced between the CCW12 promoter and the PRM9 terminator of pVYY1.0.0.5 (#1879). TdTer was amplified from pESCLeu2d-AdhE2.CCW12.(5'UTR-PYK2)sTdTer(gly) (#1556) using P161_110_PYK2R and P160_110_PYK2F. The PCR product was used in a Gibson assembly with the Bam HI-digested backbone.

pVYY1.2.0 (#1822). sTdTer(gly) with the PYK2 5'-UTR was introduced between the CCW12 promoter and the PRM9 terminator of pVYY1.0.0.5 (#1879). The PFK1 5'-UTR was amplified from plasmid #1456 using P162_120_PFK1F and P163_120_PFK1R. sTdTer(gly) was amplified from pESCLeu2d-AdhE2.CCW12.(5'UTR-PYK2)sTdTer(gly) (#1556) using P164_120_gTdTerF and P161_110_PYK2R. The PCR products were used in a Gibson assembly with the Bam HI-digested backbone.

pVYY1.3.0 (#1823). sTdTer(gly) with the PYK2 5'-UTR was introduced between the CCW12 promoter and the PRM9 terminator of pVYY1.0.0.5 (#1879). The PFK2 5'-UTR was amplified from plasmid #1457 using P165_130_PFK2F and P166_130_PFK2R. sTdTer(gly) was amplified from pESC_Leu_AdhE2_CCW12_5'UTR_PYK2_TdTer(gS.c) (#1556) using P167_130_gTdTer F and P161_110_PYK2R. PCR products were used in a Gibson assembly with the Bam HI-digested backbone.

pVYY1.4.0 (#1824). sTdTer(gly) with the YHL001W 5'-UTR was introduced between the CCW12 promoter and the PRM9 terminator of pVYY1.0.0.5 (#1879). sTdTer(gly) with the YHL001W 5' UTR was obtained by amplifying pESCLeu2d-AdhE2.CCW12.(5'UTR-PYK2)sTdTer(gly) (#1556) using P168_140_YHL001WF and P161_110_PYK2R. The PCR product was used in a Gibson assembly with the Bam HI-digested backbone.

pVYY1.5.0 (#1825). sTdTer(gly) with the TDH2 5'-UTR was introduced between the CCW12 promoter and the PRM9 terminator of pVYY1.0.0.5 (#1879). sTdTer(gly) with the TDH2 5'-UTR was obtained by amplifying pESCLeu2d-AdhE2.CCW12.(5'UTR-PYK2)sTdTer(gly) (#1556) using P169_150_TDH2F and P161_110_PYK2R. The PCR product was used in a Gibson assembly with the Bam HI-digested backbone.

pVYY1.6.0 (#1826). sTdTer(gly) with the TDH3 5'-UTR was introduced between the CCW12 promoter and the PRM9 terminator of pVYY1.0.0.5 (#1879). sTdTer(gly) with the TDH3 5'-UTR was obtained by amplifying pESCLeu2d-AdhE2.CCW12.(5'UTR-PYK2)sTdTer(gly) (#1556) using P170_160_TDH3F and P161_110_PYK2R. The PCR product was used in a Gibson assembly with the Bam HI-digested backbone.

pVYY1.7.0 (#1848). sTdTer(gly) with the VSV 5'-UTR was introduced between the CCW12 promoter and the PRM9 terminator of pVYY1.0.0.5 (#1879). sTdTer(gly) with the VSV 5'-UTR was obtained by amplifying pESCLeu2d-AdhE2.CCW12.(5'UTR-PYK2)sTdTer(gly) (#1556) using P171_170_VSVF and P161_110_PYK2R. PCR product was used in a Gibson assembly with the Bam HI-digested backbone.

pVYY1.8.0 (#1827). sTdTer(gly) with the VSV 5'- and 3'-UTRs was introduced between the CCW12 promoter and the PRM9 terminator of pVYY1.0.0.5 (#1879). sTdTer(gly) with VSV 5'-and 3'-UTRs was obtained by amplifying pESCLeu2d-AdhE2.CCW12.(5'UTR-PYK2)sTdTer(gly) (#1556) using P171_170_VSVF and P173_180_3'VSVR. The PCR product was used in a Gibson assembly with the Bam HI-digested backbone.

pVYY1.0.1_1 (#2001). The ADH domain from AdhE2 codon-optimized using *S. cerevisiae* glycolytic codon only (sADH(gly)(ADHE2), *Appendix 4.6*) with the FBA1 promoter was inserted between the HIS5 terminator and SPG5 terminator of pVYY1.0.0.5 (#1879). sTdTer(gly) with FBAp was amplified from two gBlocks (g31_FBA11_ADH-1_CPS1 and g32_FBA11_ADH-2_CPS1) using P207_gBlock32_SPG5F and P208_gBlock32_SPG5R. The PCR product was used in a Gibson assembly with the BglII-digested backbone.

pVYY1.X.1 series. This set of plasmids were constructed from the corresponding pVYY1.X.0 series by inserting sADH(AdhE2) with FBAp. The insert was amplified from pVYY1.0.1_1 (#2001) using P209_ALD5_ADH F and P208_gBlock32_SPG5R. The PCR product was used in a Gibson assembly with the Xma I-digested backbones.

Constructs for transcript processing studies.

pRS316-TDH3p.TDH3t (#2186) was constructed to allow the insertion of gene of interest between the TDH3 promoter and terminator to allow direct comparison between native and non-native transcripts. This plasmid carries a CEN origin and Ura selection marker. Both the

TDH3 promoter and TDH3 terminator were amplified from genomic DNA using P361_TDH3t_F / P204_pRS316_TDH3t_R and P199_pRS316_TDH3p_F / P360_TDH3p_R repectively. The Bam I- and Nhe I cut sites were introducing between the TDH3 promoter and terminator in the PCR primers sequences.

pRS316-TDH3p.sTdTer(gly).TDH3t (#1800) was constructed to compare the transcript abundance and translation efficiency between the abundant and highly transcribed and translated endogenous glycolytic transcript, TDH3 and the heterologous sTdTer(gly) transcript. The gBlocks corresponding to the sTdTer(gly) sequence (g21_TdTer (S.c gly) with 5'UTR PYK2 gBlock 1 and g22_TdTer (S.c gly) with 5'UTR PYK2 gBlock 2) were used in a Gibson assembly with Bam I- and Nhe I-digested pRS316-TDH3p.TDH3t (#2186).

The following plasmids were constructed for the overexpression of chaperons.

pRS316_TDH3_SSA1_TDH3 (#2303) and were constructed based on the parent construct pRS316-TDH3p.TDH3t (#2186). SSA1 was amplied using P584_SSA1-Ura_F and P585_SSA1-Ura_R. The PCR product was used in the Gibson assembly with the Bam I- and Nhe I -digested backbone.

pRS316_SSA1_YDJ1 (#2304) was constructed using pRS316_TDH3_SSA1_TDH3 (#2303) as the parent plasmid. The TEF1 promter and YDJ1 were amplied from the genome DNA using the P467_TEF1_YDJ1_F / P468_TEF1_YDJ1_R and P469_YDJ1F / P470_YDJ1R, respectively. PCR products were used in the Gibson assembly with the Sac I and Sac II-digested backbone.

pESC-Leu_YDJ1_SSA1 (#2326) was constructed using pESC-Leu (#69) as the parent plasmid. PTDH3_SSA1_TDH3t_pTEF1_YDJ1 cassette was amplified from pRS316_SSA1_YDJ1 (#2304) using P580_SSA1_YDJ1_Leu_F and P470_YDJ1R. The PCR product was used I the Gibson assembly with the Bam I and Hind III digested backbone.

The following constructs were used for co-overexpression of candidates from RNA-Seq data. All plasmids were aseembled using pESC-Ura (#70) as the parent. pESC-Ura (#70) was digested with BamH I and Xho I. Gene of interests were amplified from genomic DNA and used in the Gibson assembly reaction with the digested backbone. Gene of interests were driven by the *pGal10* promoter, along with the CYC1 terminator, Ura3 selection marker, and the 2 micron origin of replication.

pESCUra-ANB1 (#2590). ANB1 was amplied using P793_ANB1_F and P794_ANB1_R.

pESCUra-RPS14B (#2591). RPS14B was amplified using P795_RPS14B_F and P796_RPS14B_R.

pESCUra-TMA10 (#2592). TMA10 was amplified using P797_TMA10_F and P798_TMA10_R.

pESCUra-DBP2 (#2599). DBP2 was amplified using P791_DBP2_F and P792_DBP2_R

pESCUra-RLI1 (#2600). RLI1 was amplified using P848_RLI1_F and P849_RLI1_R.

Constructs for CRISPR-Cas9 genome editing.

pCas-Pphe-BsaI_NAT (#2046) was constructed from the pCAS_Pphe_BASI (#1943) parent plasmid from the J. Cate lab [10]. The original G418 selection marker was replaced by the NAT marker by Gibson assembly. pCAS_Pphe_BASI (#1943) was digested with Bgl II and Sap I to remove the G418 selection marker the pRNR2 promoter driven the expression of Cas9. The new selection marker, NAT, was amplified using P325_CAS_NAT_F and P326_CAS_NAT_R from a template plasmid with the NAT selection (gift from the J. Cate Lab). The pRNR2 promoter was amplified from the parent plasmid pCAS_Pphe_BASI (#1943) using P327_CAS_NAT_pRNR2_F and P328_CAS_NAT_pRNR2_R. All PCR products were used in the Gibson assembly with the digested backbone.

Guide sequences were inserted into the Bsa I site of pCas-Pphe-BsaI_NAT (#2046). All guide sequences were generated using the CRISPR function on Benchling [*11*] (*Appendix 4.3B*). Two 60-bp single-stranded oligonucleotides (forward and reverse) that contained the 20-bp guide sequence with 20-bp upstream and downstream homology arms were purchased (IDT) and annealed before using in a Gibson assembly with Bsa I-digested backbone. All constructs were verified by sequencing (Quintara Bioscience or UC Berkeley Barker Sequencing Facility).

Repair fragments were ordered as a single-stranded ultramer from IDT (*Appendix 4.3A*). They contain 50-bp upstream and downstream homology arms for recombination. A TAA stop codon was added after the upstream homology sequence. A 20-bp bar code sequence was added between the homology sequences. These single-stranded DNA sequences were then amplified with the corresponding primer (*Appendix 4.3A*) to generate double-stranded DNA fragments, which were used in a co-transformation with the corresponding Cas9 plasmid to generate different knockout strains.

Strain generation. All knockout strains and genome integration strains were generated using the CRISPR-Cas9 system [6]. Plasmids (1 μg) with the specific target guide (2–5 μg) were co-transformed with the linear repair fragment using the Frozen-EZ Yeast transformation kit (Zymo Research). The transformation was incubated at 30°C for 1 h before centrifuging at 4°C for 5 min at 20,817 × *g*. The cell pellet was then resuspended with YPGA (2 mL, 20 g L^{-1} peptone, 10 g L^{-1} yeast extract, 10 mg g L^{-1} adenine hemisulfate, 2% *w/v* galactose) and recovered at 30°C for 2 h. The cells were then centrifuged again at 4°C for 5 min at 20,817 × *g*, resuspended in ddH_2O (200 μL), and plated on YPG agar with NAT (100 μg. L^{-1}). Plates were then incubated at 37°C overnight for Cas9 expression before transferring to 30 °C incubator. Transformants were verified by amplification of the relevant junctions diagnostic for genome integration followed by sequencing of the PCR amplicon (*Appendix 4.3C*) (Quintara Biosciences). Verified strains were passage through 2 to 5 times in YPG media to cure the pCAS_Pphe-BsaI_NAT plasmid, which was confirmed by loss of resistance in YPG plate with NAT antibiotic.

***In vivo* production of *n*-butanol.** All yeast transformations were conducted using the Frozen-EZ yeast transformation kit (Zymo Research) following the manufacturer instructions. Overnight cultures of freshly-transformed *S. cerevisiae* strains were grown in defined dropout media (Yeast Nitrogen Base without amino acids and SC powder with the appropriate amino acid dropouts, Difco) with supplement of 2% *w/v* galactose, and buffered at pH 6.0 with 100 mM MES. Culture were grown at 30 °C and 200 rpm. Seed cultures were then used to inoculate media (30 mL) to an initial OD_{600} of 0.2 in either 250 mL non-baffled flasks for microaerobic conditions (Kimble Glass, Chicago, IL) or 250 mL non-baffled anaerobic flasks with GL45 threaded tops (Chemglass, Vineland, NJ). For microaerobic production, the culture flasks were sealed with Parafilm M (Pechiney Plastic Packaging) to prevent product evaporation. Anaerobic cultures were sealed and the headspace was sparged with argon for 5 min immediately after inoculation. Samples were quantified after either 3 or 5 d of cell culture.

Extraction and quantification of *n*-butanol. Samples (2 mL) were removed from cell culture and cleared of biomass by centrifugation at 20,000 rpm for 5 min using an Eppendorf 5417R centrifuge. The supernatant or cleared medium sample was extracted with toluene by mixing supernatant with 1:1 media toluene (with 100 mg L^{-1} heptanol as an internal standard) ratio using the 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. Standard curves were prepared freshly during each run and normalized for injection volume using the internal hexanol standard (100 mg L^{-1}). Standard curve was normalized for injection volume using the internal standard. 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).

Cell lysate enzyme assays. Biomass was harvested at the end of production and stored at -80 °C. Frozen cell pellets (from 2 mL culture) were thawed and resuspended in 500 μL of 100 mM Tris-HCl pH 7.5 containing DTT (5 mM) and PMSF (0.5 mM). The cell suspension wwas then transferred to a 2 mL Eppendorf tube with an O-ring to and glass beads (250 μL; 1 mm). Cells were lysed by two rounds of bead-beating (Biospec, 30 s each) with 5 min pause in between. The cell lysate was then centrifuged at 20,817*g* at 4°C for 5 min and the supernatant was removed for enzyme assays using Molecular Devices M2 plate reader.

PhaA. Thiolysis activity was measured by monitoring the enolate form of acetoacetyl CoA as previously described [12]. Assays were performed at 30°C in a 96-well plate in a total volume of 100 μL containing 100 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 10 μM CoA, and 20 μM acetoacetyl CoA.

Hbd, Ter, Aldh, and Adh. These activities were assayed as previously described [5]. All assays were perfomed at 30°C in a 96-well plate in a total volume of 100 μL.

Hbd. Assays contained 100 mM Tris-HCl, pH 7.5, 100 μM acetoacetyl-CoA, 100 μM NADH and were monitored by the oxidation of NADH at 340 nm.

Ter. Assays contained 100 mM Tris-HCl, pH 7.5, 100 µM NADH, and 50 µM crotonyl-CoA and were monitored by the oxidation of NADH at 340 nm.

Aldh. Assays for the Aldh domain of AdhE2 contained 100 mM Tris-HCl, pH 7.5, 0.5 mM DTT, 400 µM NAD^+, 400 µM CoA, and 10 mM butyraldehyde and monitored by the reduction of NAD^+ at 340 nm.

Adh. Assays for the Adh domain of AdhE2 contained 100 mM Tris-HCl, pH 7.5, 0.5 mM DTT, 400 µM NADH, and 10 mM butyraldehyde and monitored by the oxidation of NADH at 340 nm.

Purification of affinity-tagged proteins and antibody generation. 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, fresh freeze with liquid nitrogen and store at -80 °C.

Purification of His-tagged protein. Frozen cell pellets were thawed and resuspended in Buffer A1 (50 mM potassium phosphate, 300 mM NaCl, 20 mM imidazole, 50 µM PMSF, pH 8.0) supplemented with DNase (0.7 unit/g of cell pellet) at a final concentration of 5 mL per g cell paste. The cell suspension was homogenized by ten passes with a glass-Teflon homogenizer and was lysed with a Misonix 3000 sonicator at full power with a 15 s on/60 s off cycle for a total sonication time of 2.5 min. 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 the dropwise addition of 15% *v/v* polyethylenimine to a final concentration of 0.5% *v/v*. The precipitated DNA was removed by centrifugation at 15,300 × g for 20 min at 4°C. The cleared lysate was loaded by gravity flow onto a Ni-NTA column (Qiagen) pre-equilibrated with Buffer A1, and washed with Buffer A1 with 10 column volume. The protein was then eluted with Buffer B1 (50 mM potassium phosphate, 300 mM NaCl, 250 mM imidazole, pH 8.0). Eluted fractions were concentrated by Amicon using 10 KDa MWCO (UFC901024, Millipore) to 5 mL, which was then passed through a G-25 column (25 mL) for buffer exchange into Buffer C1 (20 mM Tris-HCl, 50 Mm NaCl, pH 7.5). Finally, glycerol was added to the eluted protein to a final concentration of 5% *v/v*. Protein concentration was measured by the Bradford assay with BSA as the standard. Total of 3 mg of purified TdTer protein (7.2 mg ml^{-1}) was sent to ProSci Inc. for polyclonal antibody generation in rabbit host.

Purification of Strep-tagged protein. Frozen cell pellets were thawed and resuspended in Buffer A2 (100mM Tris, 150 mM NaCl, 2 mM DTT, 50 µM PMSF, pH 7.5, 0.7 unit of DNase /g of cell pellet) to a final concentration of 5 mL per g of cell pellet. The cell suspension was homogenized by ten passes with a glass-Teflon homogenizer and was lysed with a Misonix 3000 sonicator at full power with a 15 s on / 60 s off cycle for a total sonication time of 2.5 min. 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 the dropwise addition of 15% *v/v* polyethylenimine to a final concentration of 0.5% *v/v*. The precipitated DNA was removed by centrifugation at 15,300 × g for 20 min at 4°C. The cleared lysate was loaded by

gravity flow onto a Strep-tactin Superflow High Capacity column (IBA) pre-equilibrated with Buffer A2 and washed with 10 column volme of Buffer A2. The protein was then eluted with Buffer B2 (100 mM Tris, pH 7.5, 150 mM NaCl, 2.5mM desthiobiotin, 2mM DTT). Eluted fractions were concentrated by Amicon using 10 KDa MWCO (UFC901024, Millipore) Glycerol was added to the eluted protein with a final concentration of 5% *v/v*. Protein concentration was measured by the Bradford assay with BSA as the standard. Total of 3 mg of purified AdhE2 protein (3 mg ml^{-1}) was sent to ProSci Inc. for antibody generation in rabbit host.

Western blot. Antibodies to both TdTer and AdhE2 were raise by ProSci Inc. (Poway, CA) in rabbits using purified proteins as described above. A 2 ml culture was harvested after it was grown for three days by centrifuging for 5 mins at 20,817 x g at 4 °C. The cell pellets was resuspended in 500 µL lysis buffer (100 mM Tris HCl, pH 7.5, 5mM DTT, 0.5mM PMSF). Lysate was then transferred to a 2 ml O-ring tube with 500 µL beads (1 mm) and cells were lysated by bead beating (BioSpec) twice with 30 seconds each with 5 mins pause in between at 4 °C. Cell debris was then transferred to a new eppie tube and spun at 20,817 x g for 2 mins at 4°C. The supernatant (soluble fraction) was transferred to a fresh tube and the cell pellet was then resuspended with 100 µL lysis buffer (insoluble fraction). Total protein was quantified using Bradford reagents with a BSA standard curve. Gel samples were prepared by mixing both the soluble and insoluble fractions with Laemmli loading buffer. The samples were boiled for 5 mins at 98 °C before being separated using SDS-PAGE gel electrophoresis. Once the gel run was complete, the content of the gel was transferred to a PVDF membrane. Membrane was then blocked with 5% BSA overnight in the cold room or at room temperature for 4 hrs. The membrane was blotted with either TdTer or AdhE2 antibodies (1 to 10,000 dilution) overnight in the cold room or 1 hr at room temperature. After straining, the membrane was washed three times with fresh TBST to remove unbound primary antibody. Finally, the membrane was blotted with secondary antibody (anti-rabbit HRP 1 to 10,000 dilution) for 2 hrs at room temperature, follow with the same washing procedures as the primary antibodies. The blot was developed using Western Lightening Plus-ECL (PerkinElmer) and imaged by the Bio-Rad gel doc under ChemIllu filter.

Real-time quantitative PCR. RNA was isolated using the RNeasy RNA isolation kit (Qiagen) following the manufactural protocol. Purified RNA (500 ng) was treated with iScript gDNA Clear cDNA Synthesis kit (Bio-Rad) to remove genomic DNA and performed cDNA synthesis according to the manufacturer protocol. cDNA was used for real-time PCR with the SYBR Green master mix (1725271) according the recommended protocol. Primers were designed using the RealTime qPCR tool from Integrated DNA Technologies (*Appendix 4.3D*). Reactions were analyzed using an iQ5 real-time PCR detection system (Bio-Rad).

Transcript 5′-cap characterization assay. RNA was isolated using the RNeasy RNA isolation kit (Qiagen) following the manufacturer protocol. Purified RNA (5 µg) was treated for 30 min at 37 ^{o}C with TURBO DNaseI (4.5 µL, Thermo-Fisher) in a 50 µL reaction to remove genomic DNA. The reaction was diluted with Buffer RLT (100 µL) and 70% *v/v* ethanol (200 µL) and transferred to an RNeasy column (Qiagen) for RNA cleanup following the manufacturer instructions. Treated RNA was then used for the enzymatic capping characterization assay[13]. All enzymatic treatments were conducted by following recommended protocols. Briefly, RNA was treated with Antarctic phosphatase (5 U) for 90 min at 37°C followed by a 10 min heat inactivation at 65°C. The reaction was then treated by T4 PNK (10 U) for 90 min at 37°C followed by a 20 min heat

inactivation at 65°C for inactivation. Finally, the reaction was split into two aliquots. Terminator exonuclease (XRN-1, 1 U) was added to one of the reactions and water was added to the other as a control. Both reactions were incubated at 30 °C for 90 min. The RNA was then purified by phenol-chloroform extraction, followed by ethanol precipitation. Reactions were run on a 1% agarose gel for diagnostic analysis.

RNA-Seq library preparation and analysis. Cells were grown under microaerobic conditions as described for *in vivo* production of *n*-butanol. Three strains were used in the RNA-Seq experiment (n = 3): host only (BY4741*adh1*-Δ), host with empty plasmids (BY4741*adh1*-Δ pESCLeu2d pESCHis pESC Ura; #68-#69-#70), and the host with the *n*-butanol pathway (*BY47adh1*-Δ ##; plasmid #800-#1454-#903). Cells were harvested 12 h after inoculation. The culture was sampled (2 mL) and centrifuged at 20,817g for 1 min at 4 °C. Cell pellets were flash frozen with liquid nitrogen and stored at -80 °C. Cell pellets were thawed on ice and RNA was isolated using the RNeasy RNA isolation kit (Qiagen) by following the manufacturer protocol. Purified RNA (4 μg) was then treated with TURBO DNaseI (4.5 μL, Theromo-Fisher) for 30 min at 37 ºC in a 50 μL reaction to remove genomic DNA. The reaction was diluted with Buffer RLT (100 μL) and 70% *v/v* ethanol (200 μL) and transferred to an RNeasy column (Qiagen) for RNA cleanup following the manufacturer instructions. RNA-Seq libraries were prepared using the Illumina TruSeq RNA Sample Prep Kit. Samples were sequenced with SR50 with the Illumina HiSeq2500 at UC Davis DNA Technologies Core. Sequence reads were assembled and analyzed in CLC Genomics Workbench 6.5 (CLC Bio, Aarhus, Denmark). The *S. cerevisiae* S288C genome was downloaded from RefSeq at the NCBI (sequence assembly version R64-1-1) (https://www.ncbi.nlm.nih.gov/refseq/) including 16 chromosomes and the mitochondrial genome. The genes for the *n*-butanol pathway (*PhaA, hbd, crt, ter, adhE2, pdc,* and *eutE*) were manually annotated and combined with the *S. cerevisiae* S288C genome as the reference (total size of 12.17 Mb). Expression values were normalized by calculating the reads per kb of mRNA exon per million mapped reads (reads per kb per million; RPKM), and further normalized using the option of "By totals" [*14*]. A mean of 45 million 50 bp single reads was generated for each library. Following the default parameters in the CLC Genomics Workbench, around 63% of reads per library was successfully imported, of which approximately 88% was mapped. Next, an unpaired two-group comparison of all nine libraries using the mapping results was used for quality control analysis. All annotations were derived from the SGD gene association file (http://www.geneontology.org/GO.current.annotations.shtml).

Polysome profile. Cells were grown under microaerobic conditions as described for *in vivo* production of *n*-butanol. Cells were harvested 12 h after inoculation. Cyclohexamide (100 μg/mL, final concentration; 50 mg /mL, stock solution in ethanol) was added three min before harvesting to immobilize ribosomes. Culture were sampled (10 mL) and centrifuged at 20,817xg for 2 min at 4 °C. Cell pellets were flash frozen with liquid nitrogen and stored at -80 °C. Cell pellets were thawed on ice and washed with 1.5 mL of polysome lysis buffer (20 mM Tris-HCl, 140 mM KCl, 1.5 mM $MgCl_2$, 1% (*w/v*) Triton X-100, pH 8.0 with 100 μg/mL cycloheximide). The washed cell pellet was then resuspended in polysome lysis buffer (500 μL) and transferred to 2 mL tube with an O-ring cap that contained glass beads (500 μL; 1 mm). Cells were lysed by bead beating (Biospec) with 6 cycles of 30 s on and 1 min off while chilling on ice. Finally, samples were centrifuged for 5 min at 20,817xg at 4 °C and the supernatant collected. The sample A_{260} was measured using a Nanodrop spectrophotometer to determine the amount of lysate to load in the

gradient (3.5 A_{260} units) [15] [16]. Samples (200 µL) were loaded to a 10 to 50% *w/v* linear sucrose gradient containing polysome gradient buffer (20 mM Tris-HCl, 140 mM KCl, 5 mM $MgCl_2$, pH 8.0, 100 µg/mL cycloheximide, 0.1 mM DTT, 20 U/ml SUPERase• In™ RNase Inhibitor). Gradients were centrifuged at 40, 000 × *g* for 2 h on a Beckman Ultracentrifuge and analyzed by the fractionator by monitoring at A_{254}nm. RNA from fractions were extracted using the RNeasy Kit (Qiagen) for transcript quantification and quantified as described in the real-time quantitative PCR section.

Adaptive evolution. A single colony was picked and inoculated in YPD media and grown at 30°C until it reached OD_{600} = 3 to 5. Cultures were then inoculated in 30 ml fresh YPD media in a 250 ml unbsaffled anaerobic flask with an initial OD of 0.05. Culture were then make anaerobic by purging with argon for 5 mins and grown at 30 °C at 200 rpm. Cultures were diluted in fresh media with an initial OD of 0.05 every 24 h.

4.3. Results and discussion

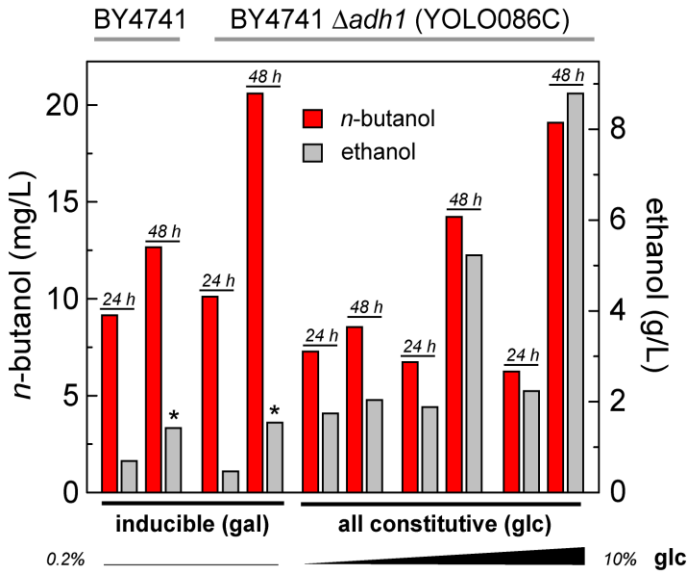
Identifying Ter as a bottleneck step. Preliminary efforts to translate the n-butanol pathway into S. cerevisiae indicated that productivity was low (~ 20 mg L^{-1}). Characterizing these strains using cell lysate enzymatic activity assays, transcript abundance quantification by real time PCR (qPCR), and *n*-butanol titer after promoter titration, the bottleneck appeared to be derived from low heterologous protein expression with the step catalyzed by the *trans*-enoyl reductase (Ter) serving as the limiting step (*Figure 4.3*, unpublished data collected by Dr. Michiei Sho). This limited step is catalyzed by the trans enyol CoA reductase (Ter) (*Figure. 4.1*). Indeed, compared to cell lysate activities in *E. coli*, each enzyme showed an order of magnitude drop or more when expressed in *S. cerevisiae*.

Introducing UTRs from native highly-expressed cytosolic genes on Ter increased *n*-butanol titer. While a significant amount of work has been carried out on examining the relationships between promoter strength and gene expression in yeast [*17, 18*], much less is known about how the sequence and structure of mRNAs contribute to protein production. Key features of eukaryotic mRNAs are untranslated regions (UTRs). These sequences may play an important role in the regulation of gene expression in yeast [*19, 20*] (*Figure 4.2*). In eukaryotes, UTRs control mRNA translation, degradation, and localization using various sequence elements, such as secondary structures, upstream initiation codons, upstream open reading frames, internal ribosome entry sites and various *cis*-acting elements that are bound by RNA-binding proteins. Moreover, UTRs regulate mRNA stability [*19, 21, 22*]. Therefore, we hypothesized that introducing the UTRs from highly-expressed cytosolic genes would increase mRNA stability and alleviate potential translation issues with heterologous transcripts. By analyzing combined data from yeast proteomics [*23*] (*Figure. 4.4A.*) and ribosome profiling studies [*24*] (*Figure. 4.4B.*), we identified glycolytic and ribosomal genes as the native genes with the highest protein levels and mRNA translation efficiencies in *S. cerevisiae*. We then reconstructed the UTR sequences of these native genes using RNA-seq data [*25*]. From these sequences, we designed and constructed a library of chimeric constructs with the *ter* open reading frame flanked by 5´- and 3´-UTRs from highly expressed genes (*Figure.4.5A.*). Using this *ter* mRNA library, we have discovered constructs with improved *n*-butanol titer. The highest tier that was achieved from this library screening showed a 3-fold (~120 to 350 mg L^{-1}) increase as compared with the original construct (*Figure.4.5B*).

Improving Ter expression by promoter screening. With the promising data from UTR library constructs screening, we wanted to improve the *n*-butanol titer further by promoter screening. Promoter screening has long served as a standard approach to improve synthetic pathways [*20*]. Three constructs were built to modulate the Ter expression. All three constructs contained a PYK2 *5′*-UTR sequence in front of Ter, which achieved the highest titer from the UTR screening. Three promoters that tested were *GAL10p, CCW12p,* and *TDH3p.* Both GAL10 and TDH3 promoter have showed as strong promoters [*26*, *27*]. The CCW12 promoter was chosen because Lin *et. al* have showed CCW12p was highly expressed under anaerobic fermentation [*28*]. We reasoned CCW12 promoter would be a strong promoter in anaerobic condition. Both the *TDH3* and *CCW12* promoters showed improved titers (~2-fold and 3-fold respectively) as compared the original *GAL10* promoter yielding up to 480 ± 5 mg L^{-1} *n*-butanol (*Figure.4.6*). This is consistent with the literature, where these two promoters have also been demonstrated as strong promoters [*28*]. Interestingly, *CCW12p* was identified from the anaerobic fermentation condition [*28*] and also demonstrated better performance under anaerobic conditions for *n*-butanol production.

Codon optimized *Ter* using glycolytic genes codon usage table improved *n*-butanol production. Codon usage is another potential factor to improve heterologous protein expression in *S. cerevisiae.* Codon bias has been extensively observed in both prokaryotes and eukaryotes. A significant amount of work has been conducted on investigating synonymous codon substitution and protein expression [*29*]. Traditional codon optimization has now become standard protocol for heterologous protein expression [*30*]. However, it is not guaranteed that codon optimization will improve protein expression despite extensive research in this area [*31, 32*]. The observation of inconsistent performance of codon optimization on protein expression is probably due to the generalization of the matrix that was used to generate the codon usage table. In other words, the matrix has taken into account too many parameters that are known to contribute to synonymous codon distribution. Recently, a promising codon optimization approach for heterologous gene expression in *S. cerevisiae* has been reported by the Alper group, which was termed "condition specific codon optimization" using growth stage as the main parameter for codon optimization [*34*]. In addition, the Boles group has reported a similar strategy where they took advantage of the naturally evolved high-flux glycolytic pathway and generated the codon usage table from glycolytic genes only. They have shown improvement on arabinose fermentation in *S. cerevisiae* by codon optimizing two of the genes using codons that are unique to glycolytic genes in the arabinose utilization pathway [*35*].

We have codon optimized two of the bottleneck genes, *ter* and *adhE2*, with various codon usage tables (*E. coli* codon usage table, *S. cerevisiae* codon usage table, and *S. cerevisiae*- glycolytic genes only codon usage table). Preliminary data suggests that codon-optimized *ter* and *adhE2* using the glycolytic genes only codon usage table modestly improves the final *n*-butanol titer by 1.4-fold. (*Figure.4.7*). We performed activity assays to characterize the functional expression of Ter in cell lysate. Using this assay, we were able to show that the increased *n*-butanol titer correlated with the increased activity of Ter (*Figure.4.8*).

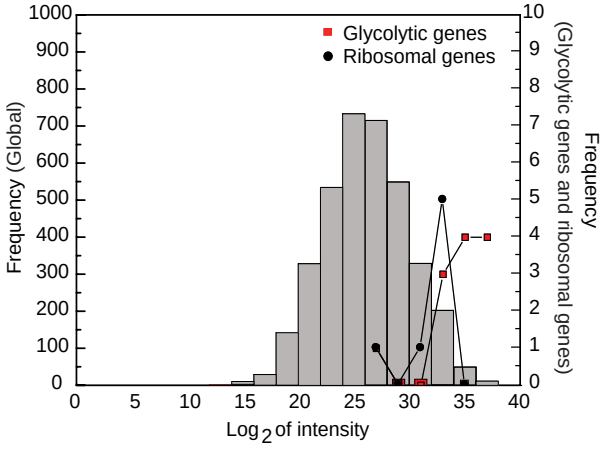


cell lysate enzyme activities

enzyme	*E. coli*	yeast
PhaA	0.55	0.03
HBD	1.68	nd
Crt	45.8	-
Ter	5.37	0.23
AdhE2 (AldH)	0.29	nd
AdhE2 (ADH)	0.14	0.12

Figure. 4.3. *n*-Butanol production titer and pathway enzymatic activities under different hosts. Left panel: Plasmids with both inducible gal promoters and constitutive promoters were constructed to examine the corresponding production profile under both the BY4741 and BY4741*adh1*-Δ hosts. Both the n-Butanol and ethanol titer were measured for all strains. This data suggested the n-Butanol pathway driven by the gal promoters under the BY4741*adh1*-Δ background gave the highest n-butanol to ethanol ratio. Red: n-butanol titer; Grey: ethanol titer. Right panel: The same n-butanol pathway was transformed and expressed under both *E. coli* and *S. cerevisiae* hosts. Cultures were harvested and used to performed cell lysate enzyme assays to access the activities for all pathway enzymes when they were expressed under the *E. coli* and *S. cerevisiae* hosts. Overall, when the pathway was expressed under the *S. cerevisiae* host, the enzyme activities were dramatically lower, except the activity from the alcohol dehydrogenase domain of the AdhE2.

A



B

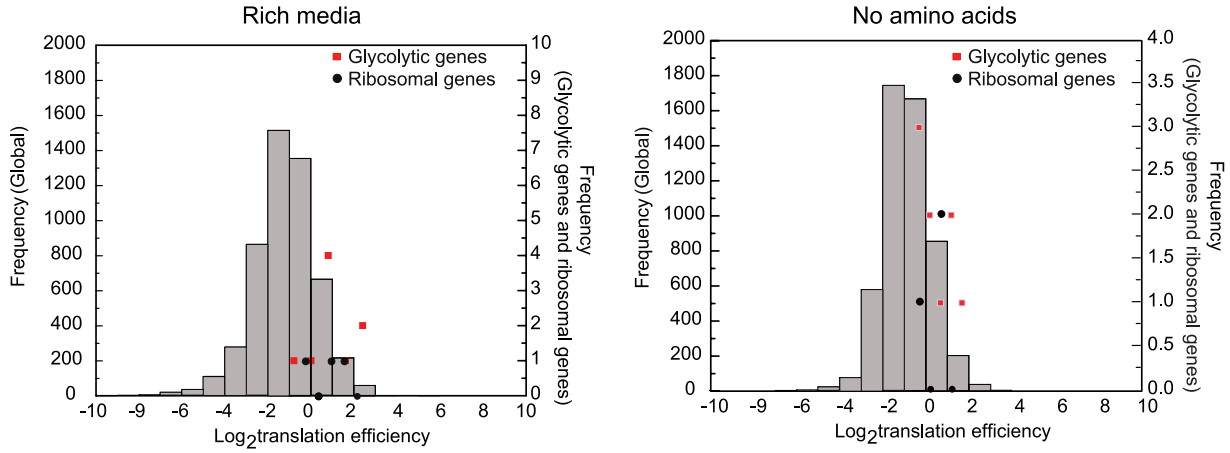
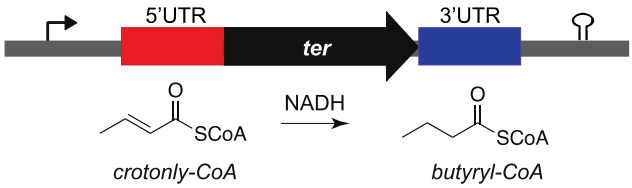
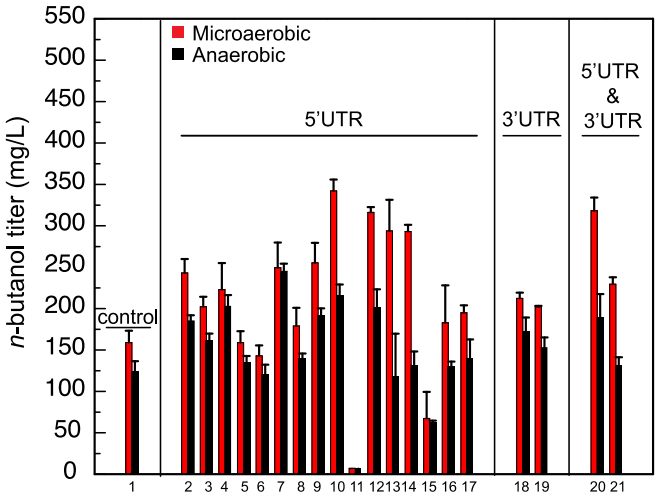


Figure 4.4. Protein abundance and translation efficiency under different media conditions. (A) Both glycolytic and ribosomal proteins are highly abundant based on the proteomic data collected by De Godoy *et. al*. [*23*]. Red: glycolytic proteins; Black: ribosomal proteins. (B) Both glycolytic and ribosomal genes demonstrate high translation efficiency compared to global transcripts under both rich (left) and no amino acids (right) mredia. Data was extracted from Ingolia *et. al.*[*33*].

A



B



Strain	UTR	Plasmid No.
1	no UTR	795
2	5' TPI1	1413
3	5' TDH2	1414
4	5' FBA1	1415
5	5' GPM1	1416
6	5' YLR075W	1417
7	5' YHL001W	1418
8	5' YJ177W	1419
9	5' TDH1	1453
10	5' PYK2	1454
11	5' PGI1	1455

Strain	UTR	Plasmid No.
12	5' PFK1	1456
13	5' PFK2	1457
14	5' ENO1	1458
15	5' ENO2	1459
16	5' CDC19	1460
17	5' TDH3	1464
18	3' FBA1	1424
19	3' YJL177W	1425
20	5' FBA1 and 3' FBA1	1426
21	5' FBA1 and 3' YJL177W	1427

Figure 4.5. Optimization of TdTer UTR sequences. (A) Design of TdTer with UTR sequences. **(**B) *n*-butanol titers from the chimeric pathway with engineered *ter* mRNA constructs. BY4741*adh1*-Δ was transformed with pESCHis-Bu2 (#800) and pESCUra-(Pcons)PDCzm.eutE (#903), and a pESCLeu2d-AdhE2.TdTer plasmid with various UTRs. Red, microaerobic; Black, anaerobic. Cells were grown in defined synthetic dropout media for 3 d with 2% (*w/v*) galactose (n = 3).

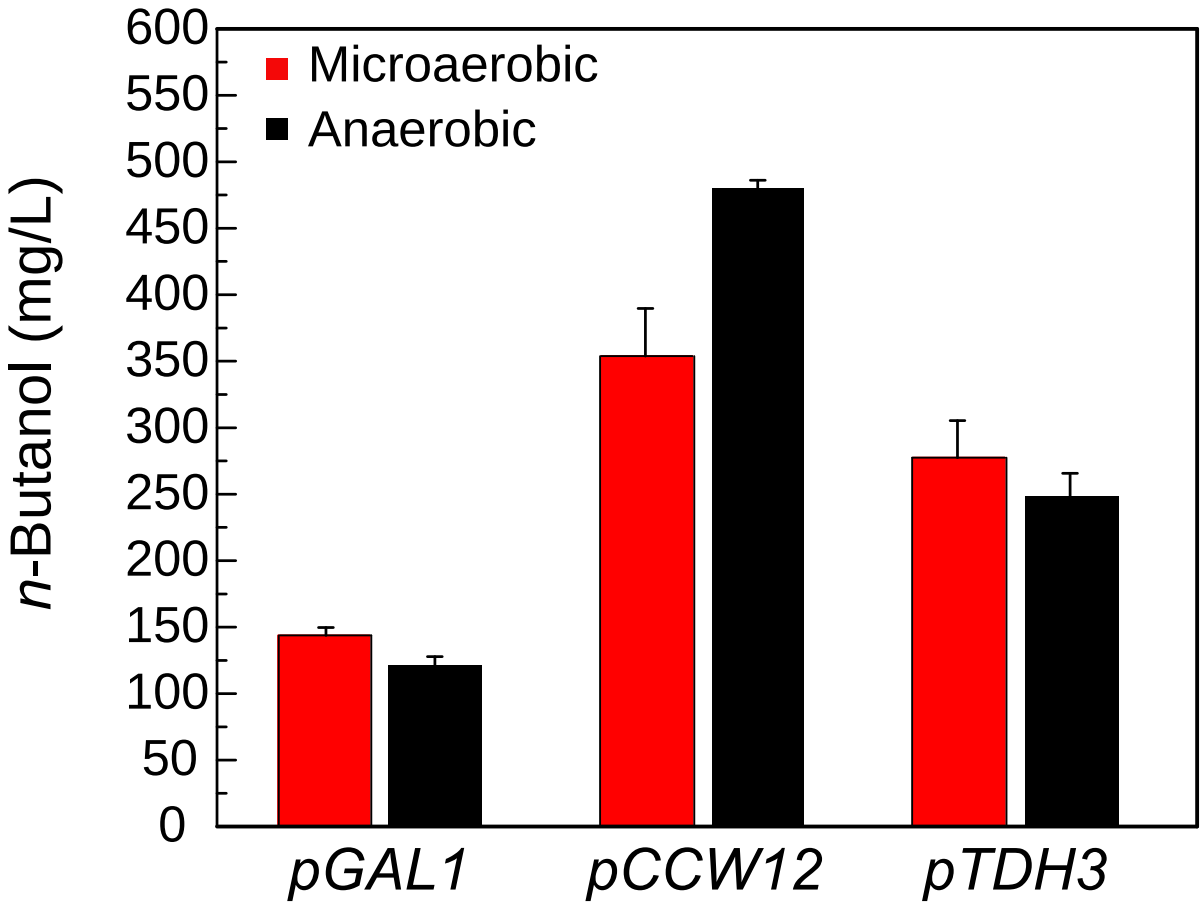


Figure 4.6. *n*-Butanol production with TdTer driven by different promoters. BY4741*adh1*-Δ was used as the production host. All hosts carried the pESCHis-Bu2 (#800) and pESCUra-(Pcons)PDCzm.eutE (#903) plasmids while varying the TdTer.AdhE2 plasmid. The following plasmids were used for promoter screening: p*GAL1*- pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454); pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)TdTer (#1525); pESCLeu2d-AdhE2.TDH3p(5'UTR-PYK2)TdTer (#1534). Cells were grown in defined synthetic dropout media with 2% (*w/v*) galactose for 3 d (n = 3).

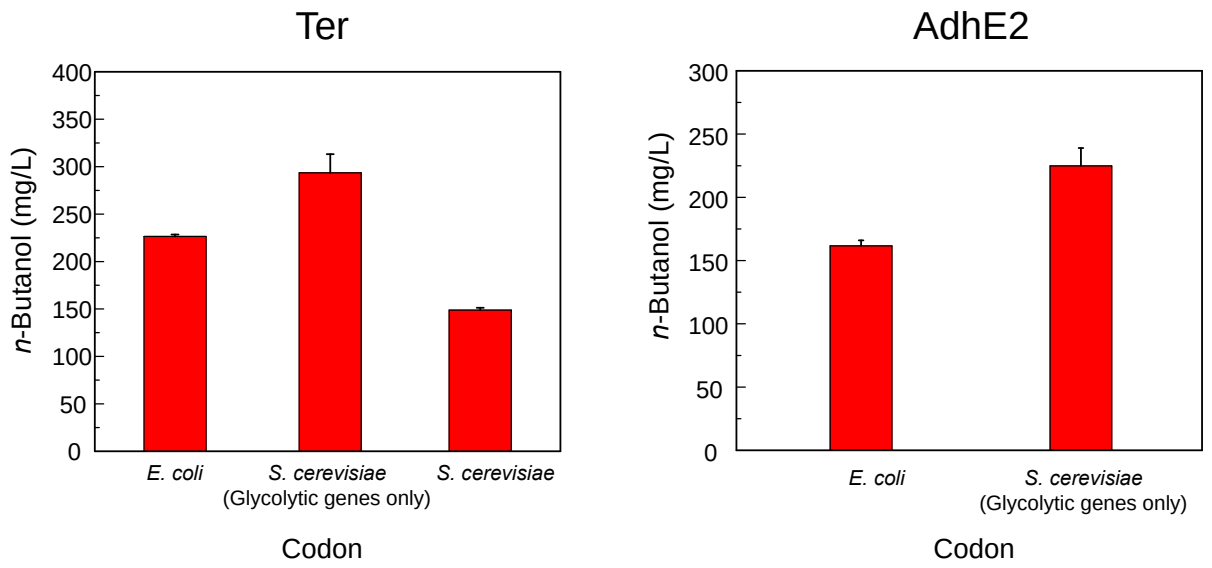
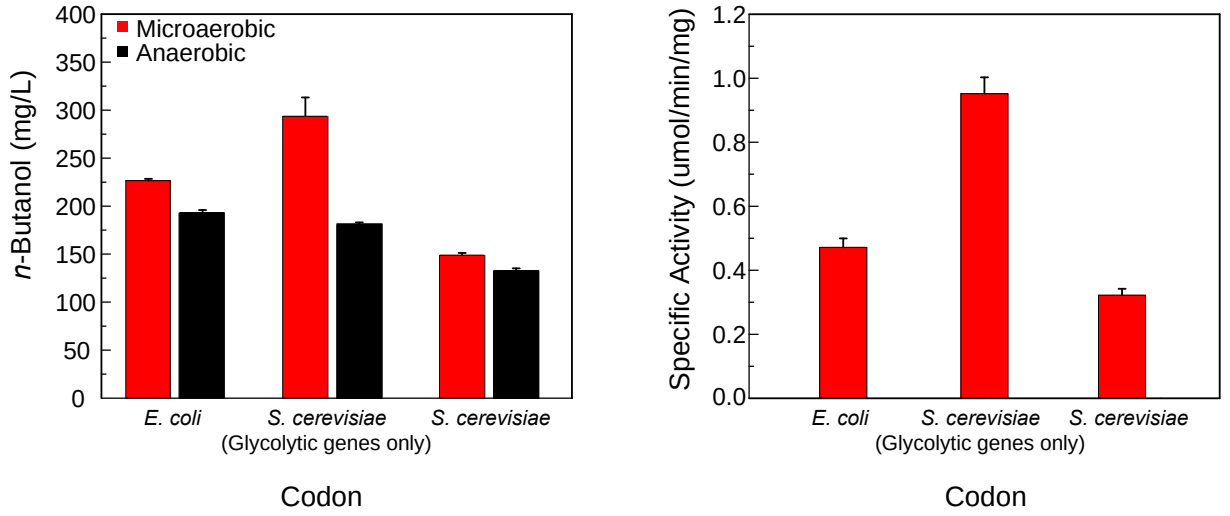
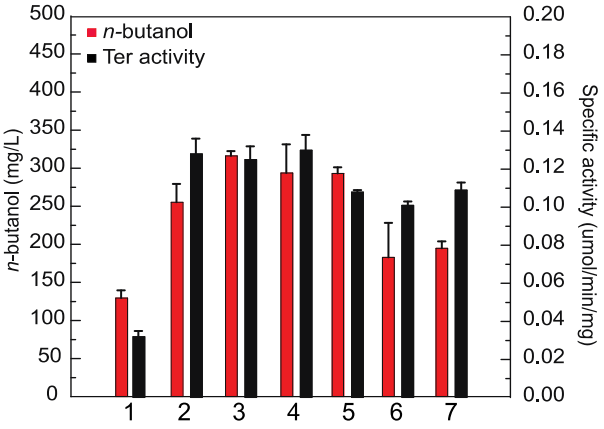


Figure 4.7. *n*-Butanol titer with different coding sequences of *ter* and *adhE2*. Both codon-optimized *ter* and *adhE2* using the glycolytic genes only codon usage table improves the final *n*-butanol titer. BY4741*adh1*-Δ was the production host. All cells carried the following plasmids: pESC.His-Bu2 (#800) and pESCUra.P(cons)PDCzm.eutE (#903). The following plasmids were used for different codon optimization versions of TdTer or AdhE2. *E. coli* codon optimized: pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454). *S. cerevisiae* codon optimized with glycolytic genes only for TdTer: pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)sTdTer(gly) (#1556); *S. cerevisiae* codon optimized TdTer: pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)sTdTer (#1558). *S. cerevisiae* codon optimized with glycolytic genes only for AdhE2 : pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)sTdTer. Cultures were grown in defined media with 2% (*w/v*) galactose under microarobic condition for 5 d (n = 3).

A



B



#	UTR	plasmid number
1	no UTR	795
2	5' TDH1	1453
3	5' PFK1	1456
4	5' PFK2	1457
5	5' ENO1	1458
6	5' CDD19	1460
7	5' TDH3	1464

Figure 4.8. Increased TdTer activity correlates with increased *n*-butanol titer. BY4741*adh1*-Δ pESCHis-Bu2 (#800) pESCUra-(Pcons)PDCzm.eutE (#903) was co-expressed with various plasmid variants containing TdTer-AdhE2. Cultures were grown in defined media with 2% (*w/v*) galactose under microaerobic conditions (n = 3). TdTer was assayed in cell lysates by monitoring the reduction of crotonyl-CoA by NADH. The assay mixture contained crotonyl-CoA (100 µM) and NADH (100 µM) in 100 mM Tris-HCl, pH 7.5 and was initiated by addition of crotonyl-CoA. No activity was observed in empty vector control. (A) The following plasmids were used to examine the effect of codon optimization of TdTer on *n*-butanol production (left) and TdTer enzyme activity (right) after 5 d: pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (*E. coli* codon-optimized, #1454), pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)sTdTer(gly) (*S. cerevisiae* codon-optimized for glycolytic usage, #1556), pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)sTdTer (*S. cerevisiae* codon-optimized for standard usage, (#1558). (B) pESCLeu2d-AdhE2.TdTer plasmid variants were used to examine the effect of 5′-UTRs on *n*-butanol production and TdTer enzyme activity after 3 d (left) according to the table (right).

Production of *n*-butanol with the integrated design construct. Our earlier data have shown promoter, codon optimization, and introducing UTRs altered the production profile of *n*-butanol. We decided to design another series of constructs for *n*-butanol production by combining all the elements we have investigated and others that factors that have been reported to affect heterologous protein expression [*20*]. Elements that were included in this design are promoters, terminators, selection markers, UTRs, and codon usage. The last two steps of the pathway which were catalyzed by Ter and AdhE2 were identified as the bottleneck steps based on preliminary data (*Figure 4.3*). Thus, these two enzymes were chosen as the initial target for optimization. It has been showed that Ura3 selection marker and terminators greatly affect protein expression level in *S. cerevisiae* [*36*]. Ura3 was chosen as the new selection marker instead of Leu2D. The bifunctional AdhE2 is now replaced by the monofunctional Ald5 and the ADH domain from AdhE2. Ald5 was driven by TDH3p and the His5 terminator. Adh was driven by FBA1p and SPRG5 terminator. Ter was driven by the pCCW12 and the PRM9 terminator with various 5'- and/or 3'-UTRs. All three enzymes, Ter , ALD5, and ADH were codon optimized by the codon table generated by the glycolytic enzymes in *S. cerevisiae* only (*Figure. 4.9A*). The result showed that introducing UTRs to the bottleneck step Ter greatly changed the production profile of *n*-butanol. Introducing the ribosomal YHL001W 5'-UTR gave the greatest effect, which showed a 10-fold increase up to 220 ± 15 mg L^{-1} *n*-butanol as compared the construct without any UTR (*Figure. 4.9B*). This highlights the potential of harnessing native UTRs for heterologous protein expression.

Screening ALDHs and ADHs. In addition to the preliminary data showed that last step of the pathway catalyzed by the bifunctional enzyme AdhE2 is one of the bottleneck steps (*Figure. 4.3*), *in vitro* kinetic data have demonstrated that AdhE2 is a promiscuous enzyme. AdhE2 reduces butyryl-CoA to butyraldehyde, which is then further reduced to *n*-butanol. However, AdhE2 can also reduce acetyl-CoA to acetaldehyde and ethanol, which is a side product that depletes the precursor from the targeted molecule (*Figure. 4.10A*). We decided to screen the ALDHs and ADH library developed by Dr. Matthew Davis [*9*] to identify a more C_4-specific ALDH and ADH using the *n*-butanol:ethanol ratio as a readout. Unfortunately, none of the ALDHs and ADHs pairs that were screened yielded improved selectivity or productivity for *n*-butanol compared to the bifunctional AdhE2 (*Figure. 4.10B*).

Exploring the expression of prokaryotic vs. eukaryotic proteins. Homolog screening is another typical approach to improve functional heterologous expression. Since all five enzymes of the *n*-butanol pathway were derived from prokaryote hosts, we wanted to explore if changing to eukaryotic homologs might improve functional expression given the molecular machineries are very different between prokaryotes and eukaryotes [*37*]. We decided to screen two different enoyl reductases from a eukaryote host, *Euglena gracillis*, EgTer and MecR1. In addition, EgTer has been observed to use either NADH or NADPH as a cofactor [*8*] whereas MecR1 uses NADPH as the reducing equivalent. TdTer was isolated from the bacterial host, *Treponema denticola*, and uses NADH as its cofactor. *n*-Butanol titers dropped from ~150 mg L^{-1} to ~ 40 to 120 mg L^{-1} when TdTer was replaced with EgTer depending on the different codons, whereas titer increased to 240 ± 23 mg L^{-1} when TdTer was replaced with MecR1 (*Figure. 4.11*). One possibility is that this enzyme is better expressed but another possibility is that co-factor usage may play a role in production titers. Cells have evolved intricate self-balance systems to maintain redox homeostasis and the consumption of NADH raher than NADPH could possibly lead to cell stress as it is still relying on ethanol production in this system for fermentation [*38*].

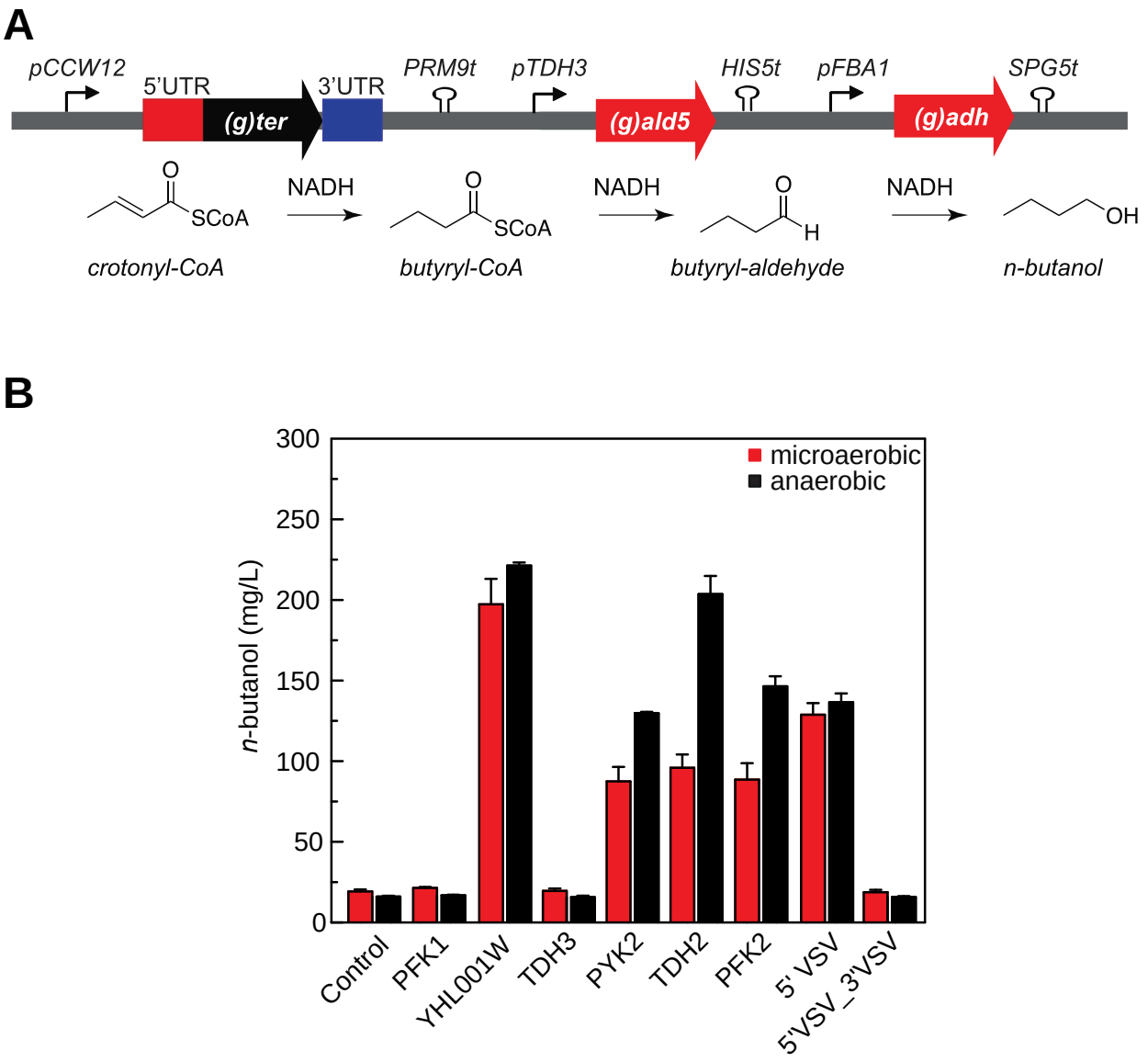
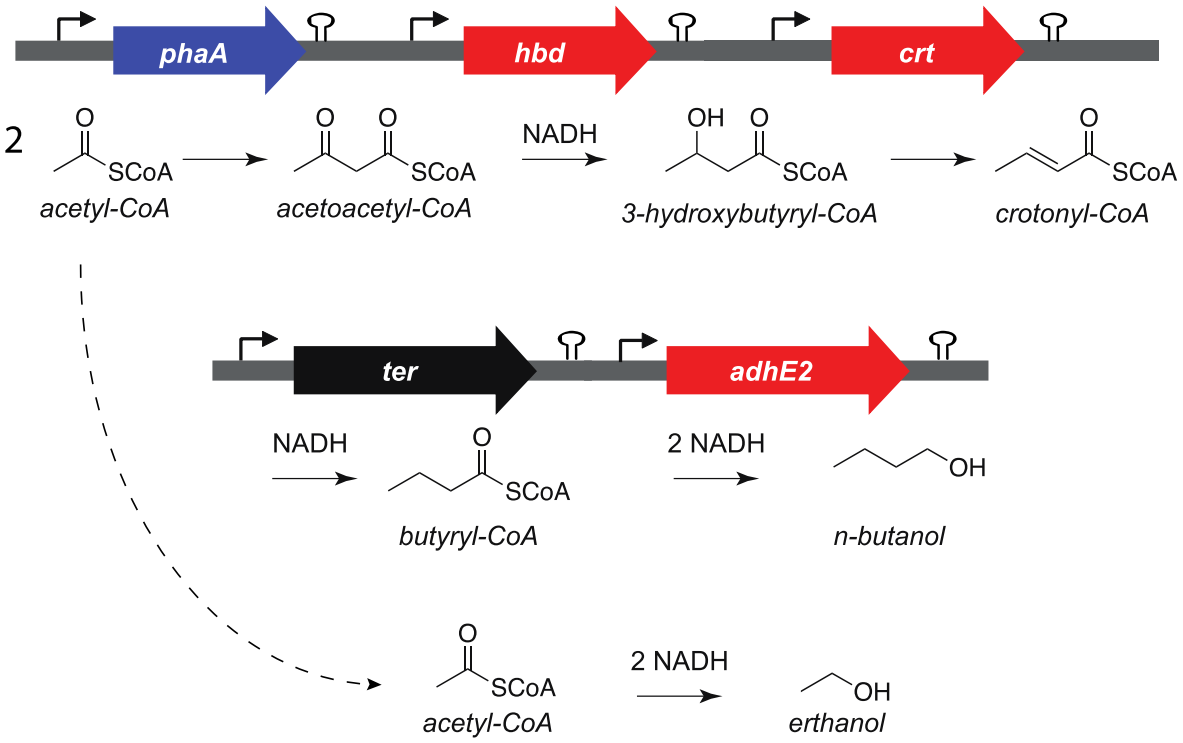
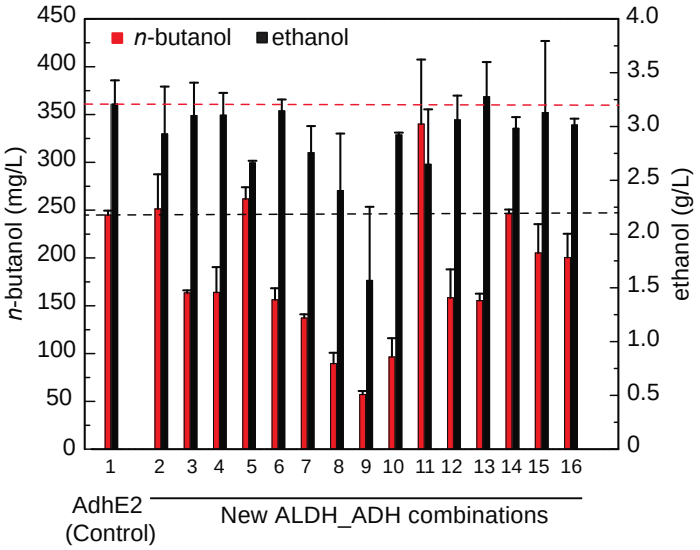


Figure 4.9. Integrating optimization of promoters, terminators, and UTRs. (A) Design of plasmid for optimization of TdTer, Aldh, and Adh gene expression. All three genes were driven by strong constitutive promoters and known terminators gave higher expression. The selection marker for the construct was Ura3. Ald5 and the ADH domain from AdhE2 were used instead of the bifunctional AdhE2 to reduce butylry-CoA to *n*-butanol. (B) Production of *n*-butanol with integrated design plasmids. BY4741*adh1*-Δ pESCHis-Bu2 (#800) with the various downstream pathways were the production hosts. Cultures were grown in defined drop out media with 2% (*w/v*) galactose for 5 d (n = 3). Downstream pathways were: Control was pVYY1.C.1 (# 1977) no UTR. Constructs for UTR screening were: pVYY1.2.1_PFK1 (#1972), pVYY1.4.1_YHL001W (#1973), pVYY1.6.1_TDH3 (#1974), pVYY1.1.1_PYK2 (#1997), pVYY1.5.1_TDH2 (#1998), pVYY1.7.1_VSV (#1975), pVYY1.8.1_5'VSV_3'VSV (#1976), pVYY1.3.1_PFK2 (#2002)..

A

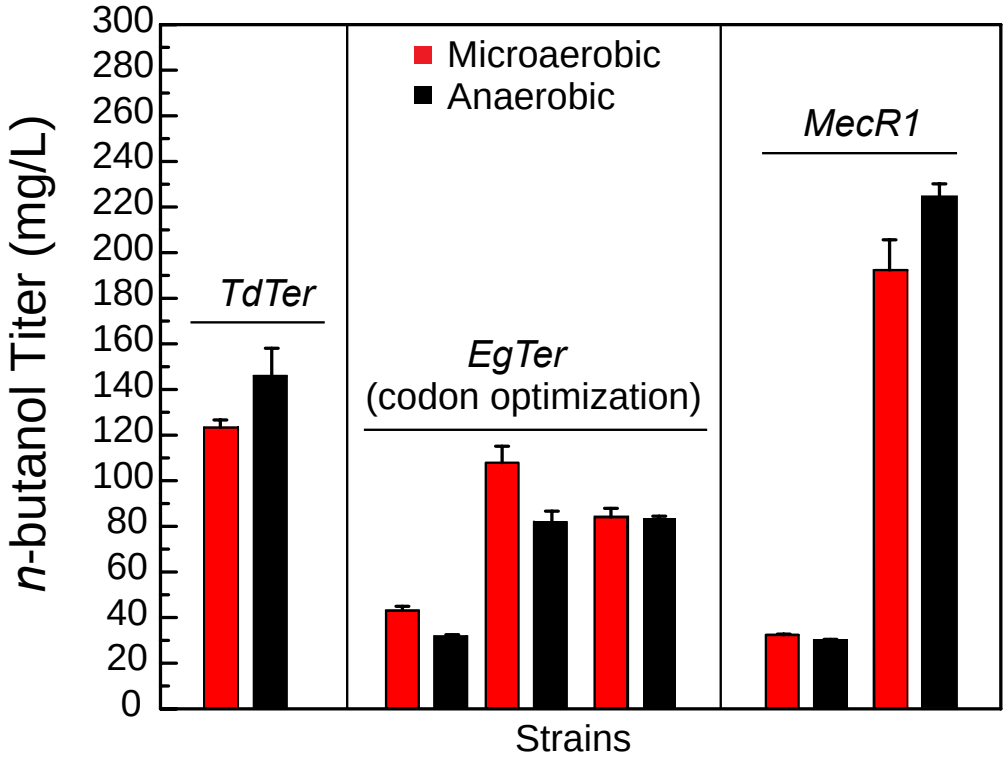

phaA
hbd
crt
2
SCoA
acetyl-CoA
acetoacetyl-CoA
NADH
OH
3-hydroxybutyryl-CoA
crotonyl-CoA
ter
adhE2
NADH
butyryl-CoA
2 NADH
OH
n-butanol
SCoA
acetyl-CoA
2 NADH
OH
erthanol

B



Strain	Aldh_Adh combination	Plasmid No.
1	AdhE2	1454
2	ALDH5-ADH2	2556
3	ALDH5-ADH8	2557
4	ALDH5-ADH22	2558
5	ALDH6-ADH2	2559
6	ALDH6-ADH8	2560
7	ALDH6-ADH22	2561
8	ALDH7-ADH2	2562
9	ALDH7-ADH8	2563
10	ALDH7-ADH22	2564
11	ALDH10-ADH2	2565
12	ALDH10-ADH8	2566
13	ALDH10-ADH22	2567
14	ALDH12-ADH2	2568
15	ALDH12-ADH8	2569
16	ALDH12-ADH22	2570

Figure 4.10. Production of *n*-butanol with different ALDH-ADH pairs. (A) The promiscuity of AdhE2 in accepting acetyl-CoA as a substrate both enables a short-circuit of the *n*-butanol pathway and complements the DAdhE phenotype of the parent strain, producing ethanol as a byproduct. (B) Screening different combinations of monofunctional aldehyde dehydrogenases and alcohol dehydrogenases to alter the *n*-butanol and ethanol ratio. BY4741*adh1*-Δ was transformed with pESCHis-Bu2 (#800) and pESCUra-P(cons)PDCzm.eutE (#903), and pESCLeu 2d plasmid that carried TdTer and various combinations of Aldh-Adh pairs. Cultures were grown in defined dropout media with 2% (*w/v*) galactose for 5 d under anaerobic conditions (n = 3).



Organism	*Treponema denticola*	*Euglena gracillis*	*Euglena gracillis*
Cofactor(s)	NADH	NADH / NADPH	NADPH

Figure. 4.11. Production of *n*-butanol with various enoyl-CoA reductase. TdTer is derived from *Treponema denticola*. Both EgTer and MecR1 are derived from *Euglena gracillis*. TdTer uses NADH as a cofactor while mecR1 uses NADPH. EgTer can use either NADH or NADPH. BY4741*adh1*-Δ was transformed with pESCHis-Bu2 (#800), and pESCUra-P(cons)PDCzm.eutE (#903) and another plasmid carrying AdhE2 and Ter from different hosts. TdTer plasmid: pESCLeu2d-ter.adhE2 (#795). EgTer plasmids from from left to right : pESCLeu2d-Adhe2.EgTer (#1124), pESCLeu2D-Adhe2.sEgTer(*E.coli*) (#1067), pESCLeu2d-AdhE2.sEgTer(YCO) (#1328). MecR1 plasmids from left to right: pESCLeu2d-AdhE2.MecR1 (#1428), pESCLeu2d-AdhE2-His_{10}MecR1 (#1429). Cultures were grown in defined dropout media with 2% (*w/v*) galactose for 5 d under both aerobic and anaerobic conditions (n = 3).

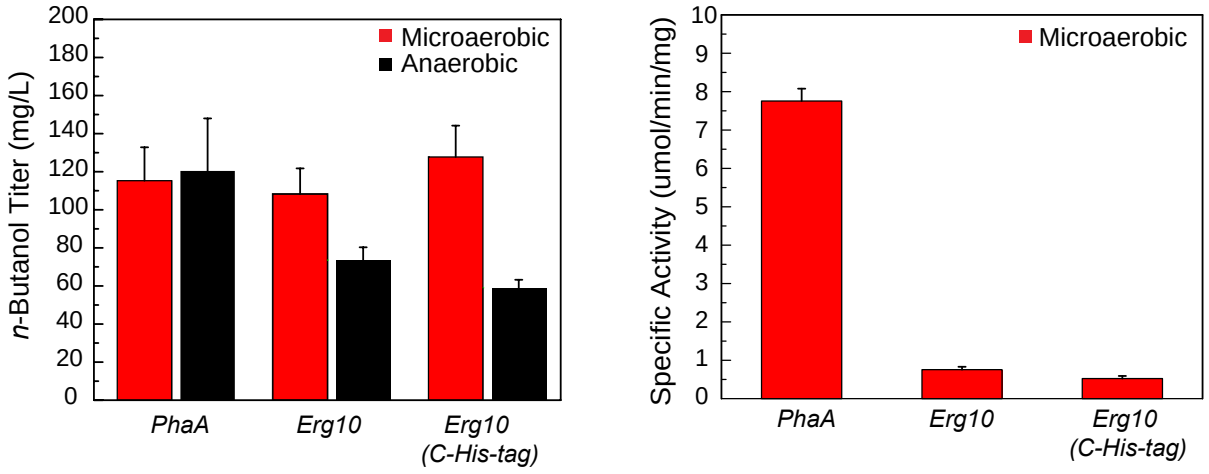


Figure 4.12. Production of *n*-butanol production with different thiolases. BY4741*adh1*-Δ pESCLeu2d-ter.adhE2 (#795), pRS416-EgPNO (#1214) with various upstream pathways were used as the production hosts. pESCHis-Bu2 (#800) carried phaA from *Ralstoni eutropha*. pESCHis-Erg10.hbd.crt (#1383) and pESCHis-Erg10His$_{10}$.hbd.crt (#1384), both carried Erg10 from *Schizosaccharomyces pombe.* Cultures were grown in defined drop out media with 2% (*w/v*) galactose for 3 d (n = 3).

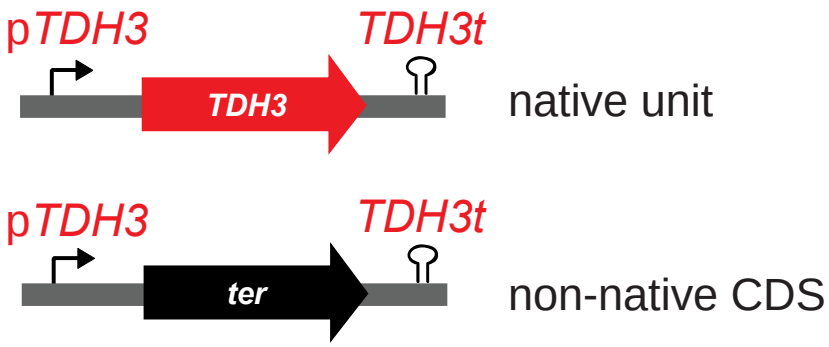
We also examined the Erg10 thiolase from *Schizosaccharomyces pombe* as another eukaryotic gene. Regardless the origin of the thiolase, there was no different in *n*-butanol titer under microaerobic condition; titer dropped slightly under anaerobic condition from 120 ± 30 mg L^{-1} to 75 ± 5 mg L^{-1} (*Figure. 4.12*). Since cell lysate activity assays showed that the strains expressing PhaA contained 8-fold greater thiolase activity than those expression Erg10, we conclude that the thiolase step does not serve as a significant bottleneck.

RT- PCR shows that pathway transcript levels are high. We wanted to determine if issues with transcript abundance was contributing to the low *n*-butanol production titer. The highly expressed endogenous protein, TDH3, from glycolysis was selected for comparison. We constructed a plasmid where Ter was driven by the *pTDH3* and *TDH3t* on a low-copy plasmid with the CEN ARS origin of replication. In addition, Ter was codon optimized by the codon table generated by the glycolytic enzymes from *S. cerevisiae* only (*Figure. 4.13A*). Thus, we had a system we could compare the transcript abundance that was encoded by the native unit vs. the non-native coding sequence within a similar context. The two strains were grown under *n*-butanol production conditions to extract RNA for target transcript quantification. Real-time PCR data of mRNA after 12 h of growth showed there was no significant difference between the TDH3 and Ter transcript level (*Figure. 4.13B*).

RNA sequencing to characterize global changes in response to *n*-butanol pathway expression. In addition to the targeted transcript quantification, we conducted an RNA-Seq experiment, which allowed us to profile transcripts in the *n*-butanol pathway (*Figure 4.14A*) as well as the global transcriptome landscape changes with and without the *n*-butanol pathway (*Figure 4.15*). The RNA-Seq experiment included a comparison of three different strains: (1) host with no plasmids (BY47471*adh1*-Δ), (2) host with empty vector controls (BY4741*adh1*-Δ pESCLeu2d pESCHis pESC Ura; #68-#69-#70), and (3) host with the *n*-butanol pathway (BY4741*adh1*Δ pESCHis-Bu2 pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer pESCUra-(Pcons)PDCzm.eutE; #800-#1454#903). RNA-Seq data showed that all pathway transcripts expression levels are high compared to TDH3. Indeed all transcriptions were more abundant (2- to 10-fold) with the exception of *pdc*, which showed a similar expression level as TDH3 (*Figure 4.14B*). This experiment is consistent with the results from RT-PCR (*Figure 4.13*). Together, they suggest that transcript abundance is not the basis for the low production titer and that issues appear to arise post-transcriptionally.

At the transcriptome level, genes that are differentially expressed between these three sets of strains (host only, host with empty vectors, and host with the *n*-butanol pathway) mainly fall into the following five categories based on the Gene Onotology enrichment analysis. They are amino acid transport/metabolism, metabolic processes, transport, phosphate metabolism, and DNA transcription. These genes are shared when the analysis was conducted between all three groups, the host only and empty vectors, the host only and the host with the *n*-butanol pathway, and the host with empty vector vs. the host with the *n*-butanol pathway (*Figure 4. 15*). However, genes involved in protein folding, proteolysis and translation were differentially expressed when the cells carried the *n*-butanol pathway, which were not observed between the host and host with empty vectors group comparison (*Figure 4.15C*). This suggested that overexpressing the *n*-butanol pathway is causing protein folding stress response and translation burden to the cell. Although the

A



B

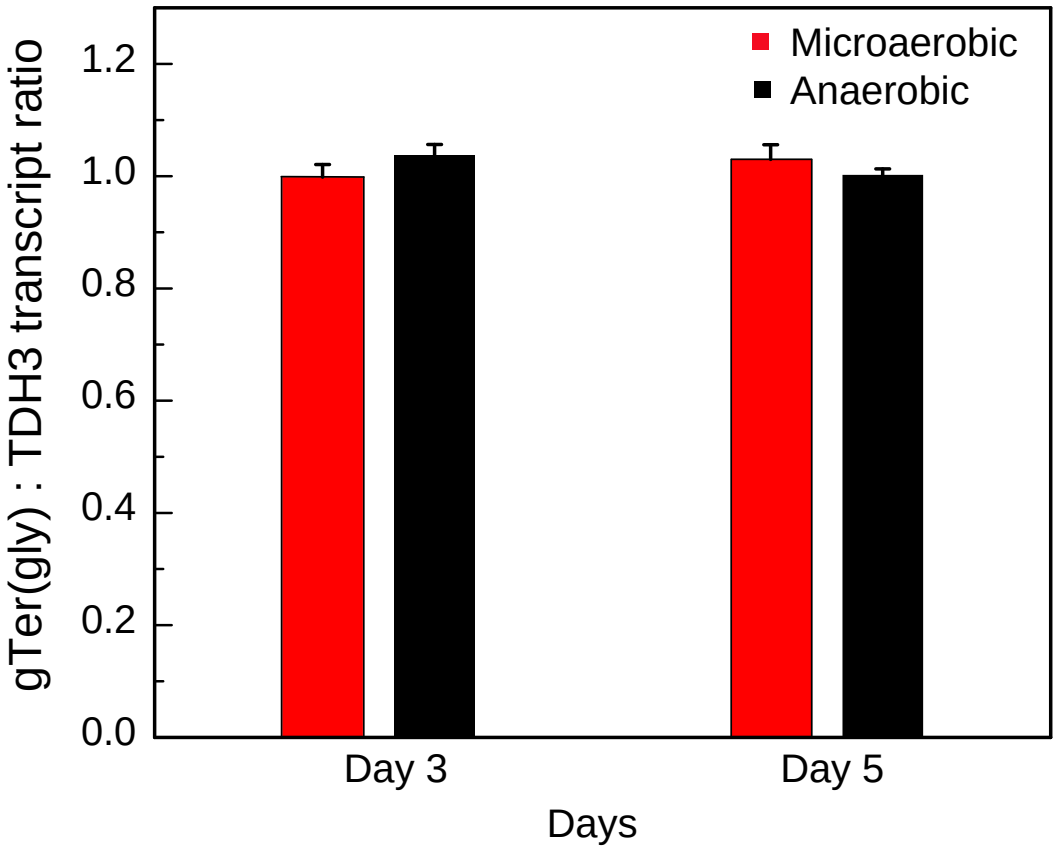
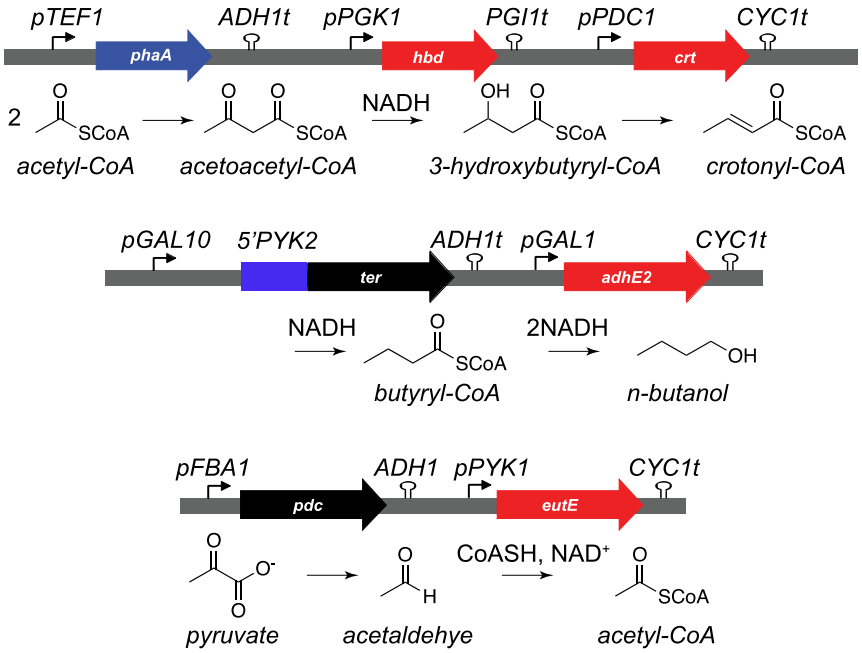


Figure. 4.13. Examining the abundance of the TdTer transcript compared to TDH3. (A) Construct design consisting of TdTer codon-optimized based on glycolytic gene usage flanked with the TDH3 promoter and terminator inserted into the pRS316 plasmid (pRS316-TDH3p.sTdTer(gly).TDH3t, #1800). TDH3 is expressed endogenously from the chromosomal copy. (B) Comparison of transcript abundance. pRS316-(TDH3p)sTdTer(gly)TDH3t (#1800) was transformed to BY4741*adh1*-Δ and grown in defined drop out media with 2% (*w/v*) galactose under both microaerobic (3 d) and anaerobic (5 d) conditions (n = 3). mRNAs were isolated and quantified by RT-PCR. All the samples were normalized to the ACT1 transcript.

A



B

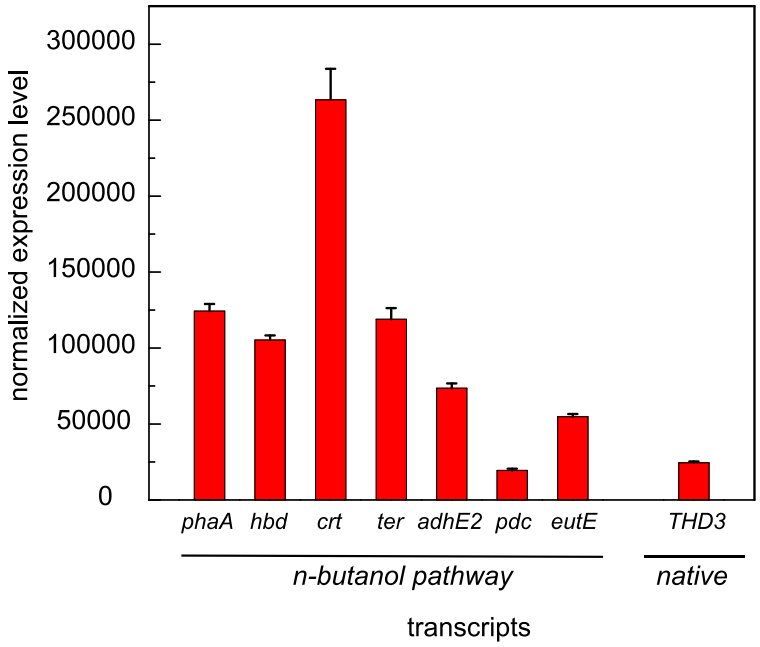
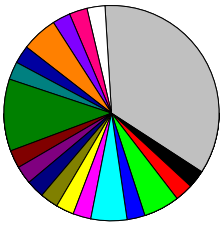


Figure 4.14. RNA sequencing to compare changes in the transcriptome with and without the *n*-butanol pathway. (A) Design of the *n*-butanol pathway strain for RNA-seq. These plasmids (pESCHis-Bu2 (#800), pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454), and pESCURA-P(cons)PDCzm.eutE (#903)) were transformed into the BY4741*adh1*-Δ host All these plasmids contained 2-micron origin of replication. (B) Normalized transcript expression level from the RNA-Seq data (n = 3). RNA-Seq data was processedby the Qiagen CLC Genomics Workbench.

A

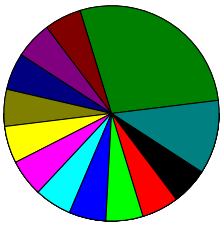
Host only vs. host with empty vectors (Up-regulated)



GO	Percentage
lipid metabolic process	2.70
alcohol metabolic process	2.70
amino acid transmembrane transporter	5.41
ATP catabolic process	2.70
cellular amino acid metabolic process	5.41
citrulline metabolic process	2.70
de novo pyrimidine nucleobase biosynthetic process	2.70
ethanol catabolic process	2.70
glycerol metabolic process	2.70
intracellular protein transport	2.70
meiotic nuclear division	2.70
metabolic process	10.81
ornithine carbamoyltransferase involved in arginine biosynthesis	2.70
protein localization to bud neck	2.70
regulation of transcription, DNA-templated	5.41
septin ring assembly	2.70
transport	2.70
tRNA wobble uridine modification	2.70
unknown	35.41

Total number of genes: 37

Host only vs. host with empty vectors (Down-regulated)

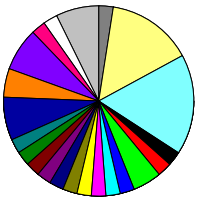


GO	Percentage
de novo pyrimidine nucleobase biosynthetic process	5.56
agglutination involved in conjugation with cellular fusion	5.56
amino acid catabolic process to alcohol via Ehrlich pathway	5.56
cellular response to DNA damage stimulus	5.56
glyoxylate cycle	5.56
histindine biosynthetic process	5.56
metabolic process	5.56
negative regulation of protein kinase activity	5.56
pheromone-dependent signal transduction invovled in conjugation with cellular fusion	5.56
phosphate-containing compound metabolic process	5.56
polyphosphate metabolic process	5.56
transport	27.78
unknown	11.11

Total number of genes: 18

B

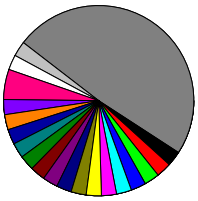
Host only vs. host with the *n*-butanol pathway (Up-regulated)



GO	Percentage
amino acid catabolic process to alcohol via Ehrlich pathway	2.44
de novo pyrimidine nucleobase biosynthetic process	2.44
agglutination involved in conjugation with cellular fusion	4.88
amino acid transmembrane transport	2.44
aromatic amino acid family catabolic process to alcohol via Ehrlich pathwya	2.44
carbohydrate metabolic process	2.44
cellular response to DNA damage stimulus	2.44
DNA replication-dependent nucleosome assembly	2.44
glycerol metabolic process	2.44
glyoxylate cycle	2.44
histidine biosynthetic process	2.44
mitochondrial electron transport, ubiquinol to cyctochrome c	2.44
negative regulation of protein kinase activity	2.44
pheromone-dependent signal transduction involved in conjugation with cellular fusion	7.32
phosphate-containing compound metabolic process	4.88
polyphosphate metabolic process	7.32
potassium ion transmembrane transport	2.44
ribosomal small subunit assembly	2.44
sulfur amino acid metabolic process	7.32
transcription, DNA-templated	2.44
transport	14.63
unknown	17.07

Total number of genes: 41

Host only vs. host with the *n*-butanol pathway (Down-regulated)

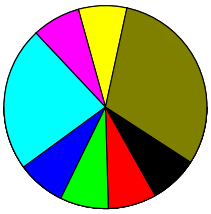


GO	Percentage
transport	2.56
adenine catabolic process	2.56
amino acid catabolic process to alcohol via Ehrlich pathway	2.56
amino acid transmembrane transport	2.56
ATP catabolic process	2.56
biotin biosynthetic process	2.56
cellular amino acid metabolic process	2.56
citrulline metabolic process	2.56
G1/S transition of mitotic cell cycle	2.56
glyoxylate cycle	2.56
metabolic process	2.56
nuclear-transcribed mRNA catabolic process, nonsense mediated decay	2.56
protein folding	2.56
protein phosphorylation	2.56
proteolysis	2.56
regulation of pH	2.56
regulation of transcription, DNA-templated	5.13
response to unfolded protein	2.56
thiamine metabolic process	2.56
unknown	48.72

Total number of genes: 39

C

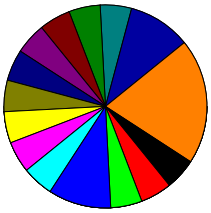
Host empty vectors vs. host with the *n*-butanol pathway (Up-regulated)



GO	Percentage
aromatic aminod acid family catabolic process to alcohol via Ehrlich pathway	7.69
DNA replication-dependent nucleosome assembly	7.69
ethanol catabolic process	7.69
methionine metabolic process	7.69
phosphate-containing compound metabolic process	23.08
protein folding	7.69
transcription, DNA templated	7.69
unknown	30.77

Total number of genes: 19

Host empty vectors vs. host with the *n*-butanol pathway (Down-regulated)



GO	Percentage
translation	5
adenylate cyclase-modulating G-protein coupled receptor signaling pathway	5
amino acid catabolic process to alcohol via Ehrilch pathway	5
biotin biosynthetic process	10
cellular response to DNA damage stimulus	5
GPI anchor biosynthetic process	5
lipid metabolic process	5
mitochondrial genome mainenance	5
nuclear-transcribed mRNA catabolic process, non-sense mediated decay	5
protein phosphorylation	5
proteolysis	5
response to unfolded protein	5
transcription, DNA-templated	5
transport	10
unknown	20

Total number of genes: 20

Figure 4.15. RNA-Seq profiles of host only, host with empty vectors, and the n-butanol pathway. Gene Ontology (GO) for genes differentially expressed between host only and host with empty vectors (A), host only and host with the n-butanol pathway (B), and host with empty vectors and host with the n-butanol pathway (C). GO analysis were performed using the CLC Genome Workbench software. GO category represented by genes that are up-regulated and down-regulated 24 h growth after inoculation at 30 ºC in the microaerobic production conditions as described in the method. The percentage represents the number of genes within each GO divided by the total number of differentially regulated genes.

number of genes in the protein folding and translation categories was low compared to other biological processes, exploring these genes might provide unique insights on post-transcriptional regulation on heterologous protein expression.

The TdTer transcript is 5′-capped. Although our data have shown that transcript abundance is not a contributor to the low production titer, RNA processing is very complex in eukaryote systems (*Figure. 4.2*), including mRNA transport, modification (5′-capping and 3′-tailing), and translation efficiency that directly affects protein synthesis. Furthermore, it has been known that transcript abundance does not correlate well with protein abundance. Taking the published RNA-Seq and proteomic data from *S. cerevisiae*, we reanalyzed the data focusing on glycolytic and ribosomal genes. The linear correlation coefficient between the transcript and protein abundance ranges from 0.26 to 0.59, indicating a poor fit (*Figure. 4.16*). We therefore decided to examine some the post-transcriptional events that ultimately control functional protein expression.

It has been well documented that translation initiation is limiting step for translation, with cap-dependent translation initiation serving as the canonical mechanism in eukaryotes. We wanted to assess if highly abundant mRNAs coded by heterologous genes were indeed capped as other translation initiation mechanisms could be introduced such as cap-independent or internal ribosome entry site-mediated [*39, 40*]. We identified enzymes selectively digest RNAs with specific modifications and adapted a method to assess 5′-capping [*13*]. We digested all un-capped RNAs via series enzymatic reactions and subsequently performed RT-PCR using specific primers to detect the mRNA. Briefly, we treated RNAs from cells that expressed Ter with Antarctic phosphatase, follow by T4 polynucleotide kinase, and XRN-1, a 5' monophosphate specific exoribonuclease. After these enzymatic treatments, the RNAs were extracted for cDNA synthesis followed by the RT-PCR analysis (*Figure. 4.17*). We saw amplification after all these treatments which suggested that Ter was capped and should be competent to undergo cap-dependent translation (*Figure. 4.18*), In the future, we would also carry out the control to show the converse that digestion of RNAs with a cap-removal enzyme (Tobacco acid pyrophosphatase) followed by XRN-1 exoribonuclease leads to the expected disappearance of TdTer from the mRNA pool [*13*].

Ter transcript has lower translation efficiency compare to TDH3 and global translation is problematic. The level of functional protein expression is determined by the translation efficiency of the transcript. We wanted to compare the translation profile for the TdTer and TDH3 transcripts by performing polysome profiling. RNAs were extracted from the polysome fractions and RT-PCR was contducted to quantify the Ter and TDH3 transcript abundance in each fraction. RT-PCR showed that overall the TDH3 transcript level showed a 1.5-2-fold greater abundance than the Ter transcript in this experiment, which is not significant. Indeed, replicates have shown that the abundance for both transcripts is quite similar (*Figure. 4.13*). However, we did observe more significant changes in the polysome fractions, where TDH3 showed a 4-fold greater abundance compared to the TdTer transcript. This suggests that TDH3 has a slightly higher translation efficiency than TdTer (*Figure. 4.19A*). In addition, we examined a polysome profile with the entire *n*-butanol pathway to investigate global translation. Comparing with the empty vector control, cells that carried the *n*-butanol showed global translation is significantly reduced. This observation suggests that cells are under stress, as indicated by the smaller 80S and polysome

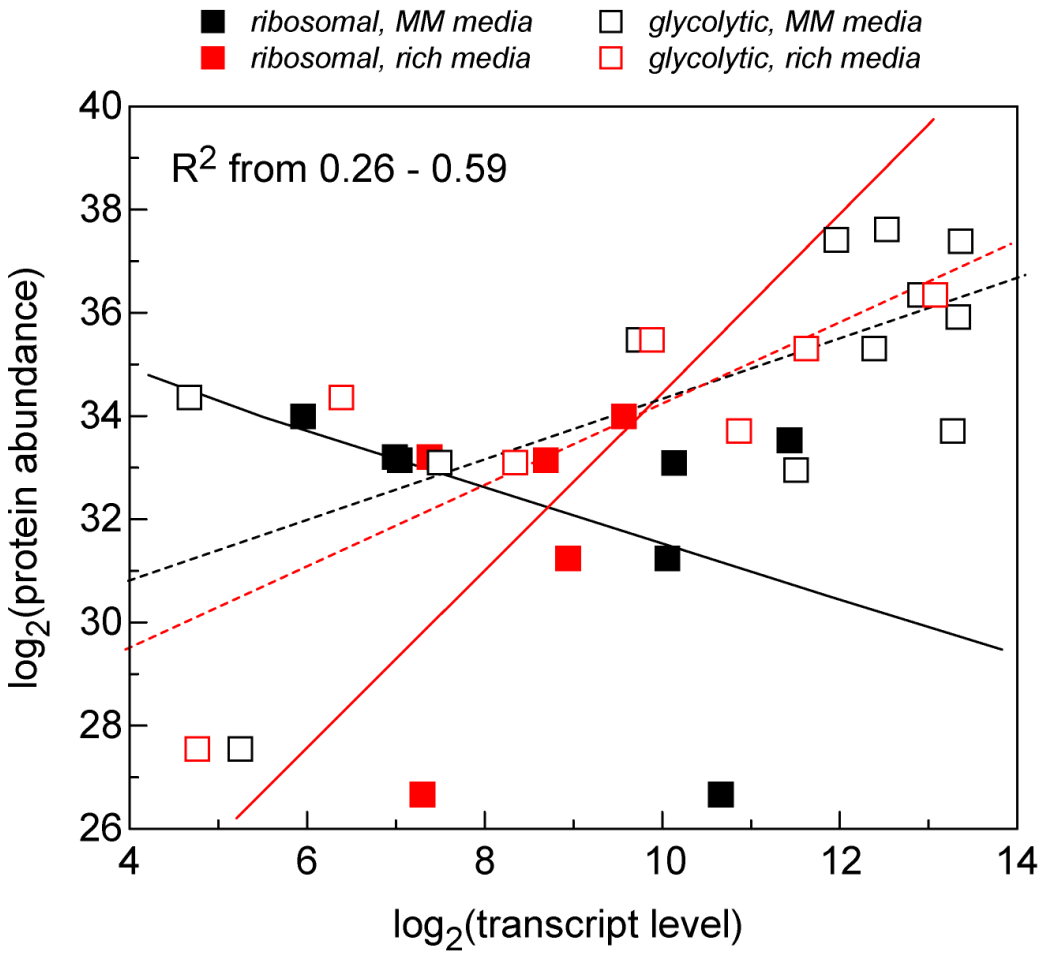


Figure. 4.16. Poor correlation between protein levels and transcript levels in *S. cerevisiae* under different media conditions. Both glycolytic and ribosomal genes showed poor correlation between protein and transcript levels under rich and minimal media. Protein abundance was extracted from De Godoy *et. al*.[23]. Transcripts abundance under rich media was extracted from Nagalakshmi *et. al*.[25] and transcripts from minimal media were extracted from Lin *et. al*.[28].

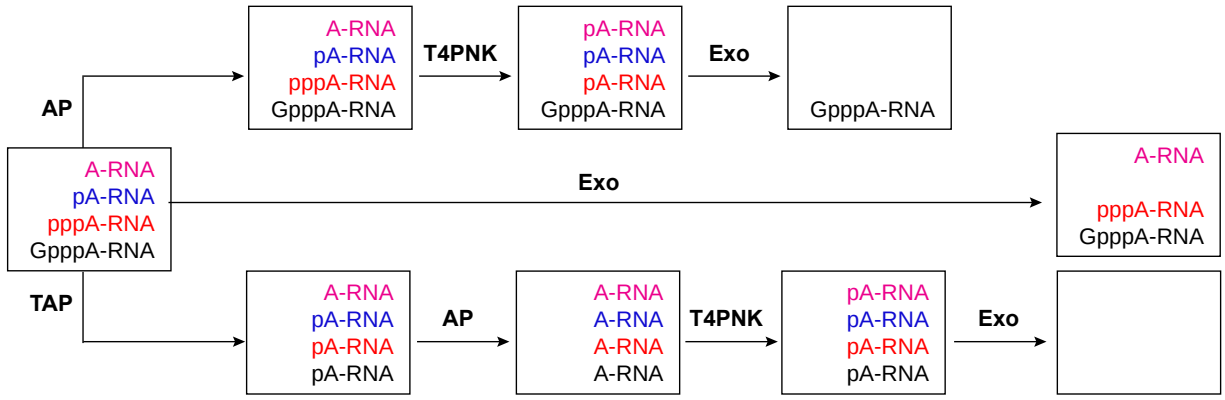


Figure. 4.17. 5′-cap assays for transcripts. Total RNA was extracted and subjected to enzymatic hydrolysis by Antarctic phosphatase to remove all the phosphate end modifications on uncapped transcripts. Samples were then treated with T4 PNK to add a 5'-phosphate to uncapped transcripts before XRN-I Exo digestion. PCR amplification of the remaining pool should yield product for 5'-capped substrates that are excluded in these reactions.

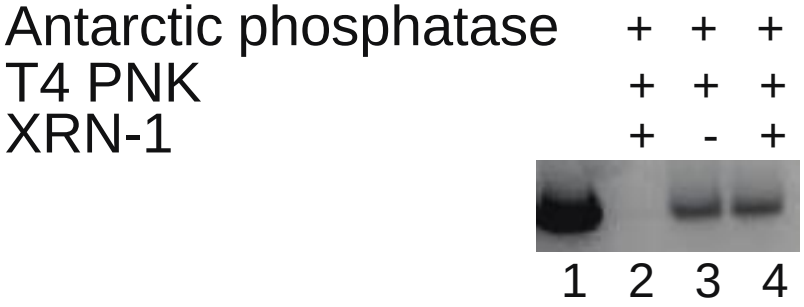


Figure. 4.18. Gel analysis of TdTer transcript 5′-cap assay. BY4741*adh1*-Δ with the pTDH3_gTdTer_TDH3 plasmid *was* grown in defined media with 2% *w/v* galactose under microaerobic conditions and grown for 24 h. RNA was subjected to the 5'-cap assay and used in a RT-PCR quantification after purification. PCR products were run on a 1% agarose gel and stained with ethidium bromide for qualitative analysis. (Lane 1) Plasmid contained TdTer gene was used as a template for control for RT-PCR as a positive control. (Lane 2) mRNA was isolated from cell culture transformed with empty vector control as a negative control. (Lane 3) mRNA isolated from culture containing the pTDH3_gTdTer_TDH3 plasmid and treated with XRN-1. (Lane 4) mRNA isolated from culture containing the pTDH3_gTdTer_TDH3 plasmid without XRN-1 treatment.

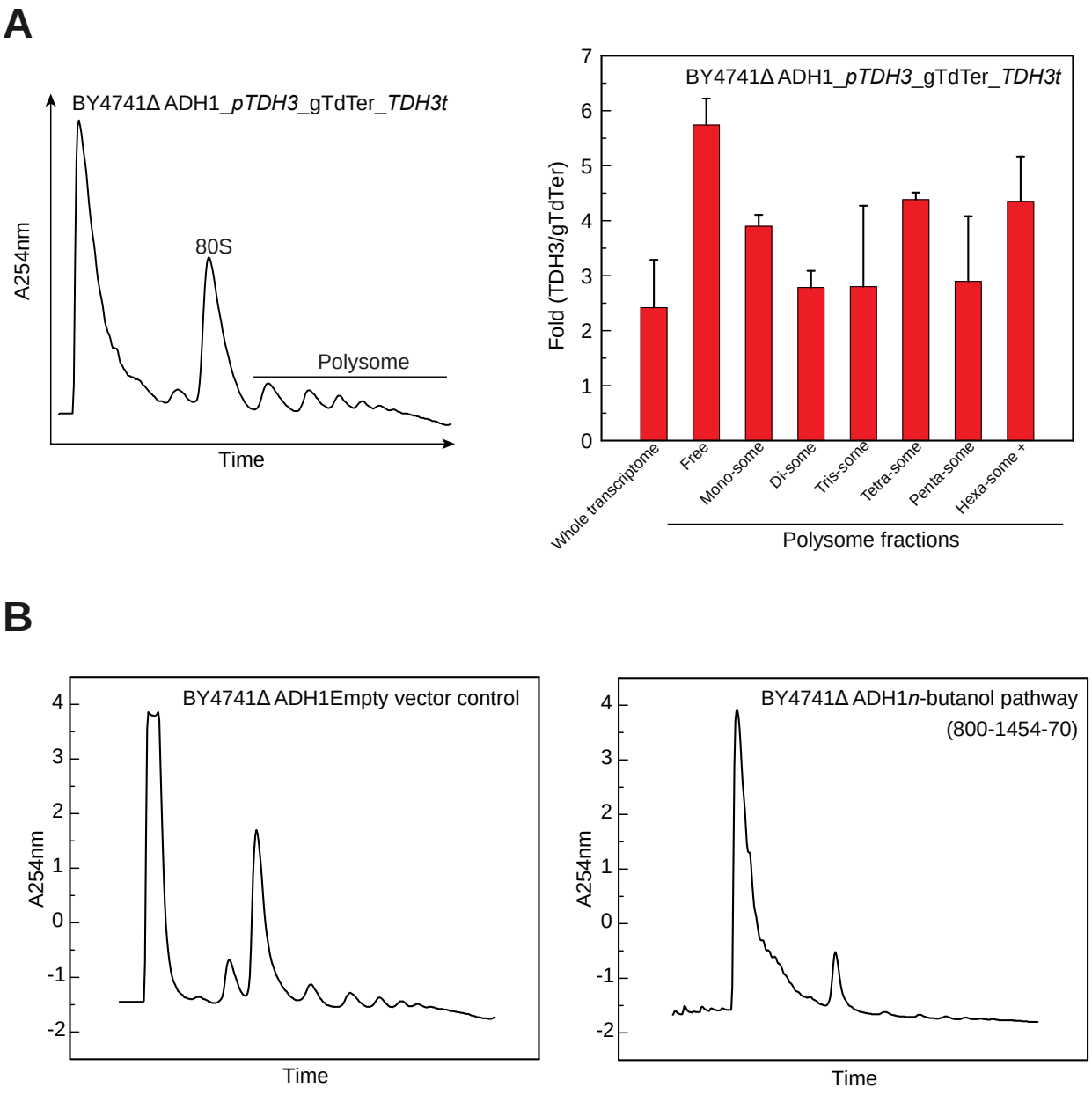


Figure. 4.19. Polysome profiles for cells expressing TdTer or the *n*-butanol pathway compared to an empty vector control. *S. cerevisiae* cultures were grown in defined media with 2% *w/v* galactose under microaerobic conditions and grown for 24 h. Cycloheximide was added before harvesting and lysates were prepared and subjected to polysome analysis using a 10-50% *w/v* sucrose (A) BY4741*adh1*-Δ pRS316-TDH3p.sTdTer(gly).TDH3t (#1800). (left) Polysome profile. (right) Relative abundance of TDH3 and TdTer transcripts. RNAs from different fractions from the gradient were isolated and used as template for real-time PCR. (B) Comparison of polysome profiles of strains containing empty plasmids and the *n*-butanol pathway. (left) BY4741Δ*adh1* pESCHis (#68) pESCLeu2d (#69) pESCUra (#70). (right) BY4741*adh1*-Δ pESCHis-Bu2 (#800) pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454) and pESC-Ura (#70).

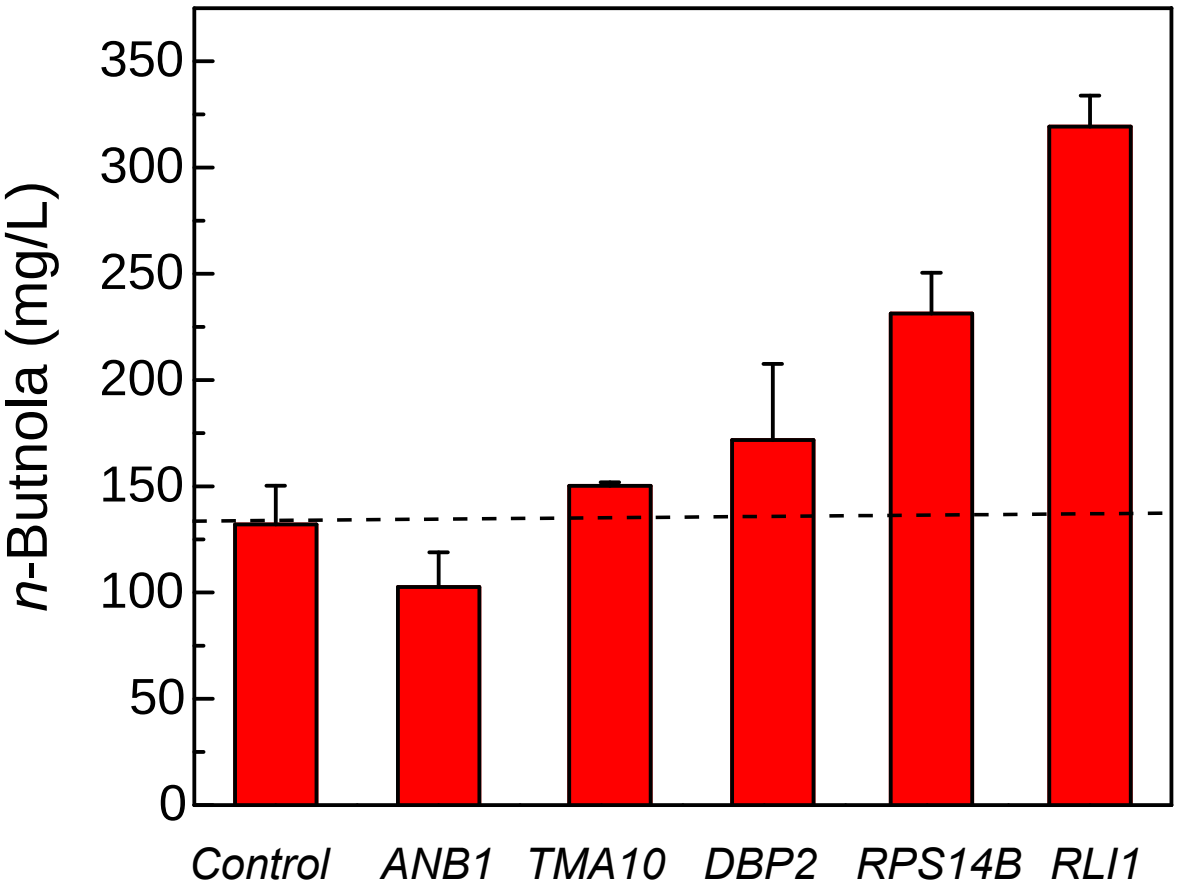


Figure. 4.20. *n*-Butanol production with co-expression of candidates from RNA-Seq data. BY4741*adh1*-Δ pESCHis-Bu2 (#800) pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454) with pESCUra plasmids for co-expression of upregulated genes from RNA sequencing experiments (pESC-Ura #70, empty plasmid; pESCUra-ANB1, #2590; pESCUra-RPS14B, #2591; pESCUra-TMA10; #2592; pESCUra-DBP2, #2599; pESCUra-RLI1, #2600). Production was conducted in defined media with 2% *w/v* galactose under microaerobic conditions for 3 d (n = 3).

peaks [*41, 42*] (*Figure. 4.19B*). Exploring factors that stimulate global translation would alleviate the translation challenge, which might ultimately improve production titer. Interestingly, our RNA-Seq data showed that genes involved ribosome biogenesis and protein translation were differentially expressed (*Figure. 4.15*, *Appendix 4.7*). We therefore co-expressed these factors, TMA10 (pathway: empty vector – 2.63 fold; pathway: host only- 5.25 fold), DBP2 (pathway: empty vector – 3 fold; pathway : host only- 2.08 fold), RPS14B (pathway: empty vector – 1.3 fold; pathway : host only- 3.9 fold), and ANB1 (pathway: empty vector -2.58-fold) with the *n*-butanol pathway to test their effect on product titers. In addition to this list, we also added RLI1 to the screening list as it has been reported RLI1 assists translation re-initiation [*43*] . Interestingly *RPS14B* and *RLI1* both increased *n*-butanol production titer compared to the empty vector control by 2.25- and 1.86-fold, respectively (*Figure. 4.20*).

Expressing the upstream pathway (PhaA-Hbd-Crt) on a CEN ARS plasmid lowered *n*-butanol titer. The RNA-Seq data suggested that all transcripts from the *n*-butanol pathway were relatively high compared to the highly expressed endogenous transcript *TDH3* (*Figure. 4.14B*). We hypothesized that expressing the pathway on the high copy number of plasmid might lead to a metabolic burden and that lowering the expression level could alleviate the stress. This hypothesis is consistent with the polysome profiles that showed down-regulation of global translation with expression of the *n*-butanol pathway (*Figure 4.19B*). We decided to overexpress the upstream portion of the *n*-butanol pathway (PhaA-Hbd-Crt) on a lower copy number plasmid with the CEN ARS origin of replication since PhaA showed low dependence on enzyme activity. This construct was tested with three plasmid variants carrying the downstream portion of the pathway (Ter-AdhE2) were preserved on a high-copy plasmid as they are known bottlenecks in this pathway. Overall, all three strains showed that high-copy number of the upstream pathway is still required to achieve maximal *n*-butanol titer. (*Figure. 4.21*).

Protein degradation is eliminated in protease knockout strains. We expressed and purified Ter and AdhE2 to raise antibody for these two proteins in order to directly measure protein abundance in cell lysate by Western blot (*Figure 4.22*). Western blot showed ~50% of the Ter protein was in the insoluble fraction and the majority of the AdhE2 protein was in the insoluble fraction. In addition, both Ter and AdhE2 were heavily degraded (*Figure. 4.23BC*). To address the degradation issue, we decided to examine the expression of Ter in two protease knockout strains BJ1991 and BJ5457, where vacuole proteases *PEP4* and *PRB1* were knocked out. Interestingly, Ter degradation was fully abolished in these two protease knockout strains (*Figure. 4.23B*) and the enzymatic activity of Ter also improved 5.4-fold (*Figure. 4.24*). However, when *n*-butanol production was tested in BY4741*adh1*-Δ *pep4*-Δ *pbr1*-Δ background host, there was no distinguishable difference from the BY4741*adh1*-Δ parent strain. These results suggest that even though protein degradation is eliminated, other factors contribute to the low product titer (*Figure. 4.24*). Western blot with the TDH3 antibody suggested that the vacuolar protein degradation is not specific to TdTer as TDH3 degradation is also inhibited (*Figure. 4.23B*). These data may indicate that reducing vacuolar degradation may not improve cytosolic availability of the protein.

Heat shock proteins program the *n*-butanol production profile. We wanted to explore if other stress and protein degradation pathways may play a role in heterologous pathway expression levels. Since protein quality control appears to be a contributing factor, it is possible that other heat shock proteins could alleviate the degradation and protein solubility problems. Thus, we screened TdTer expression in hosts where genes that encode for heat shock proteins or the ubiquitination pathway

were deleted (*Table 4.1*). First, we transformed the Ter construct (pTDH3_gTdTer_TDH3t (#1800) with the CEN ARS origin of replication in these knockout hosts and monitored TdTer expression by Western blot. Interestingly, in addition to *PBR1* and *PEP4* knockouts, knocking out genes involved in the ubiquitination pathway, *RKR1* and *HDR1* appeared to alleviate Ter degradation. Knocking out *STE13* and *YDJ1* almost completely abolished Ter expression, suggesting that they play a critical role on Ter expression (*Figure. 4.25A*). Consistently, the *SSA1* knockout strain showed a significant defect when grown on galactose. Since both *YDJ1* and *SSA1* are on the same protein folding pathway, we hypothesized that their overexpression could improve protein expression. While, no significant changes in TdTer expression were observed by Western blot co-expression of *SSA1* improved *n*-butanol production from 150 ± 2 mg L^{-1} to 260 ± 25 mg L^{-1} (*Figure. 4.25B*). With this promising result in hand, we decided to increase the expression level of SSA1 by placing it on a high-copy 2 micron plasmid, resulting in an increase of *n*-butanol from 140 ± 12 mg L^{-1} to 540 ± 10 mg L^{-1} (*Figure. 4.25C*).

Since these knock-out strains yielded a different expression profile for Ter, we decided to screen *n*-butanol production in these hosts as a quick and indirect readout for functional protein expression level for the enzymes in the entire *n*-butanol pathway. We did observe a dynamic range of production titer for *n*-butanol, where multiple strains showed almost two-fold improvement in production titer compared to the parent strain (*Figure. 4.26*). Next, we knocked out the major alcohol dehydrogenase, *ADH1* from the heat-shock and chaperone knockout strains with the goal to improve *n*-butanol further as it greatly improves *n*-butanol titer (*Figure. 4.3*). Unfortunately, none these strains gave an improved production profile (*Figure. 4.26*), suggesting that they may not be as effective at higher product yields.

Over expressing the PDH bypass and knocking out *GCN5* increased *n*-butanol production. Yeast has gone through a long history of evolution on ethanol fermentation. Ethanol fermentation is the major fermentation pathway that depletes carbon input. Deletion of the major alcohol dehydrogenase isozyme, *ADH1* improved *n*-butanol production (Dr. Michiei Sho). In addition to endogenous fermentation pathways competition, limited cytosolic acetyl-CoA presents a great challenge for increasing the *n*-butanol production. We addressed the availability of cytosolic acetyl-CoA challenge through two different approaches. First, we overexpressed the PDH bypass pathway to drive the flux from pyruvate to cytosolic acetyl-CoA. The PDH bypass pathway includes two enzymes, the pyruvate decarboxylase (*pdc)* that converts pyruvate to an aldehyde, which is subsequently ligated with a CoA by the *eutE* to product acetyl-CoA (*Figure. 4. 27*). With the over expression of the bypass pathway, the *n*-butanol titer increased from 180 ± 5 mg L^{-1} to 360 ± 15 mg L^{-1}. The second strategy to increase cytosolic acetyl-CoA is to minimize the expense of acetyl-CoA. Besides being the central building block, acetyl-CoA is also a precursor for post-transcriptional modification. Namely, acetyl-CoA is the substrate for histone acetylation, which is executed by the acetyl transferase, *GCN5*. Indeed, knocking out *GCN5* further improved *n*-butanol from 360 ± 15 mg L^{-1} to 550 ± 10 mg L^{-1} (*Figure. 4.28*).

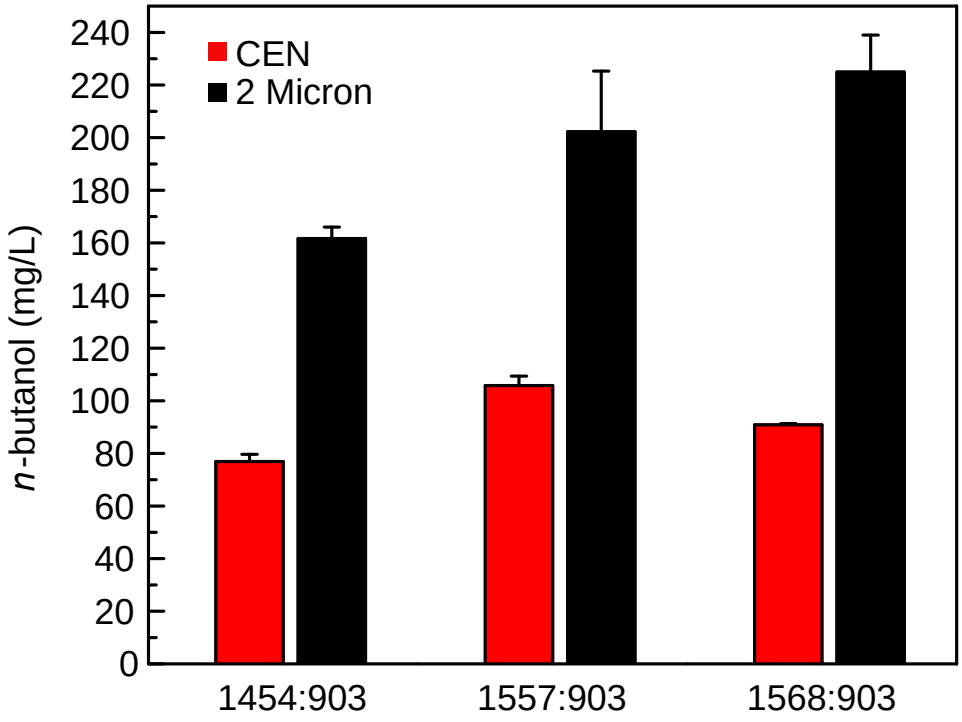


Figure 4.21. *n*-Butanol production with a reduced copy number plasmid for PhaA-Hbd-Crt. BY4741*adh1*-Δ pESCUra.(Pcons)PDCzm.eutE (#903) containing varied *n*-butanol plasmids were compared. Cultures were grown in defined media with 2% (*w/v*) galactose under microaerobic condition for 3 d (n = 3). Plasmids for TdTer.AdhE2 expression are: pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer, #1454; pESCLeu2d-AdhE2.TDH3p(5'UTR-PYK2)sTdTer(gly), #1557; pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)sTdTer, #1568. Comparison is made between PhaA-Hbd-Crt on low-copy (pRS413-Bu2 #932, red) and high (pESC.His-Bu2 #800, black) plasmids.

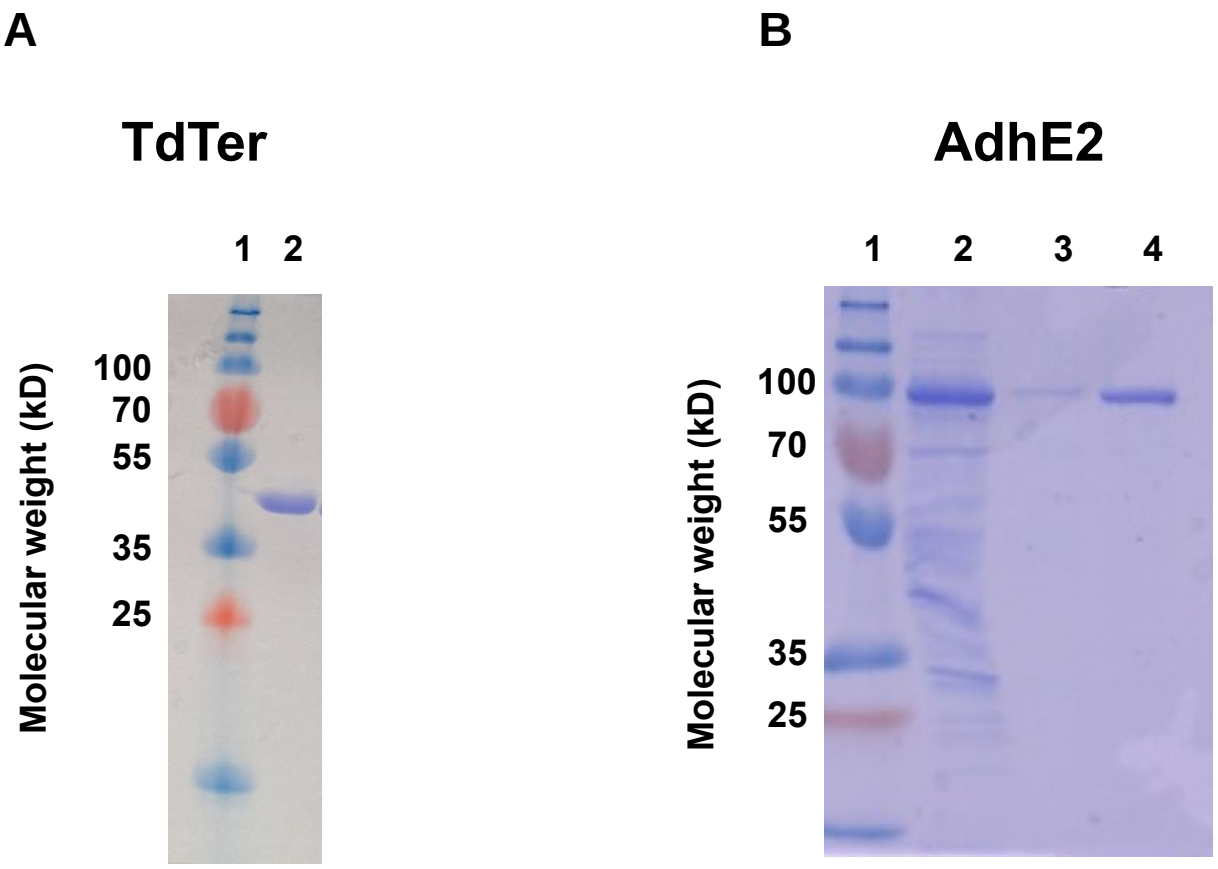


Figure 4.22. SDS-PAGE of TdTer and AdhE2 protein purification. (A) TdTer (44 kD): (Lane 1) Ladder, (Lane 2) His-TdTer. (B) Strep-AdhE2 (94 kD): (Lane 1) Ladder, (Lane 2) Cell lysate, (Lane 3) Elution 1, (Lane 4) Elution 2.

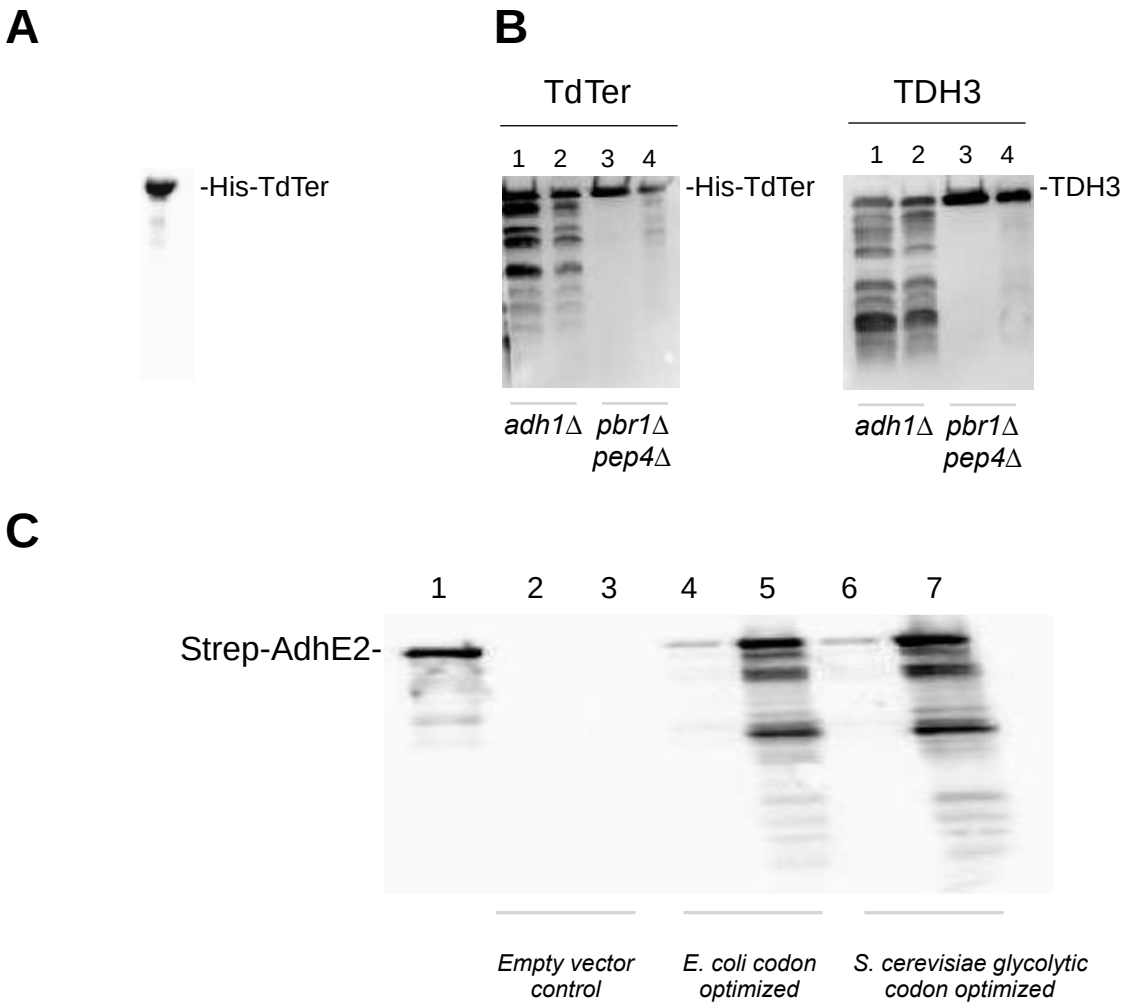


Figure 4.23. Western blots for TdTer, AdhE2, and TDH3. (A) Characterization of the Ter antibodies with purified His-Ter. Ter and AdhE2 antibodies were raised for to analyze the expression profile of the Ter and AdhE2 proteins by immunoblotting. (B) Western blot comparing the expression of Ter and TDH3 in the BY4741*adh1*-Δ and BJ1991Δ*pbr1*Δ*pep4* hosts. Both hosts were transformed with the pTDH3_gTdTer_TDH3t plasdmid (#1800) for the expression of TdTer. Lane 1 (soluble fraction) and 2 (insoluble fraction) represent the expression pattern in BY4741*adh1*-Δ host. Lane 3 (soluble fraction) and 4 (insoluble fraction) represent the expression profile in BJ1991Δ*pbr1*Δ*pep4*. (C) Characterization of AdhE2 expression profiled in the BY4741*adh1*-Δ host. Two different coding sequences for AdhE2 were examined. The host was transformed with the *n*-butanol pathway pESCHis-Bu2 (#800), pESCUra-P(cons)PDCzm.eutE (#903), and either pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454) for the *E. coli* codon optimized AdhE2 or pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)sTdTer (#1568) for the *S. cerevisiae* glycolytic codon optimized AdhE2 and grew under the standard microaerobic conditions for 3 d (n = 3). Biomass was then harvested and lysed for western blot analysis. There was no significant difference in protein expression pattern between the *E. coli* codon optimized and the *S. cerevisiae* glycolytic codon optimized version. Most of the AdhE2 protein were in the insoluble fraction.

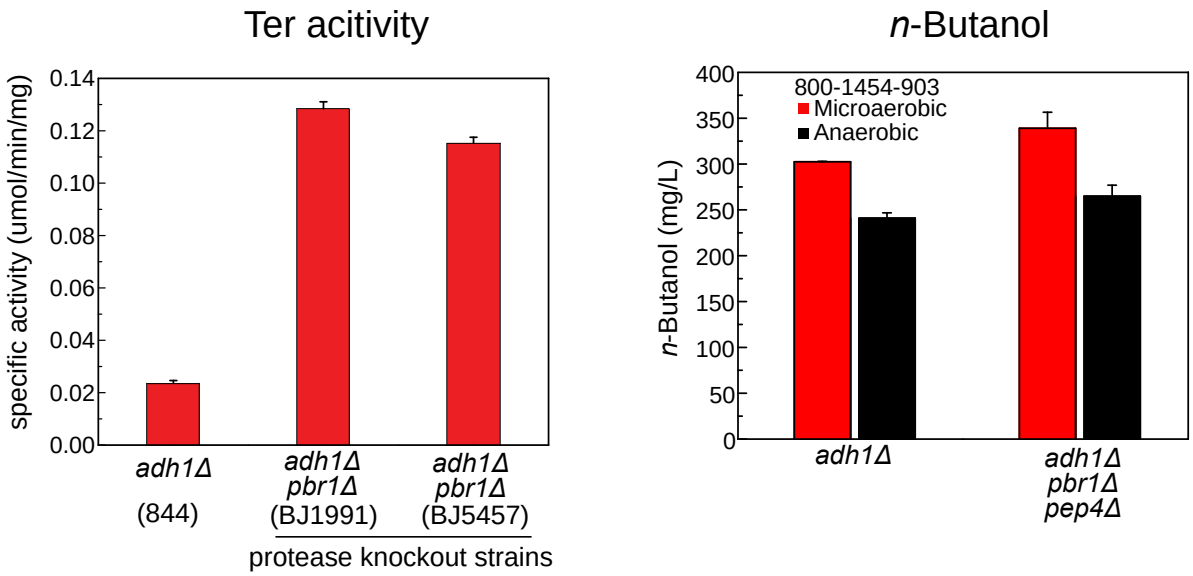
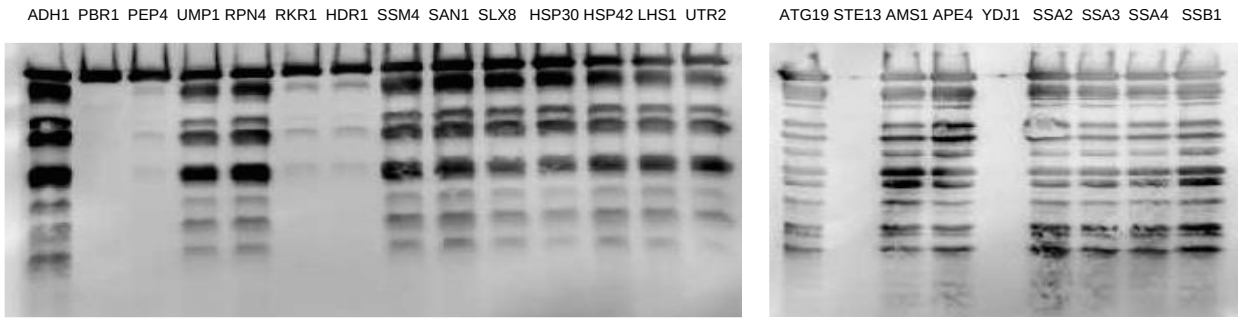


Figure 4.24. Ter activity and *n*-butanol production in vacuole protease knockout hosts. BY4741*adh1*-Δ, BJ1991 *pbr1*-Δ *pep4*-Δ, and BY5457 *pbr1*-Δ *pep4*-Δ were transformed with the pTDH3_gTdTer_TDH3t plasmid (#1800) for the expression of Ter. Cultures were grown under standard microaerobic production conditions for 3 d (n = 3). Biomass was harvested and lysed for Ter activity assays, which showed thatTer activity increased by 5 fold in both vacuole protease knockout strains (BJ1991 and BJ5457). under the single *adh1* knockout and the triple knockout host, where *adh1*, *pbr1*, and *pep4* were deleted. The same host strains were transformed with pESCHis-Bu2 (#800), pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454), and pESCUra.P(cons)PDCzm.eutE (#903) to examine *n*-butanol production. Cultures were grown under the defined media with 2% w/v galactose under both anaerobic and aerobic conditions (n = 3). There was no significant difference in production titer between the single and triple knockout host under both microaerobic and anaerobic conditions.

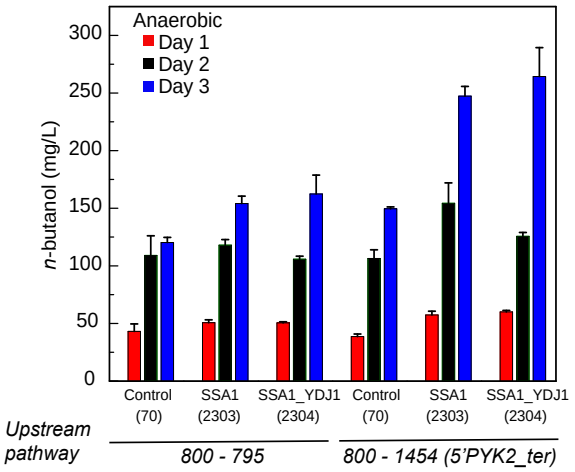
Table 4.1. Selected knockouts for Ter expression screening. Genes involved in ubiquitination and heat shock response were chosen.

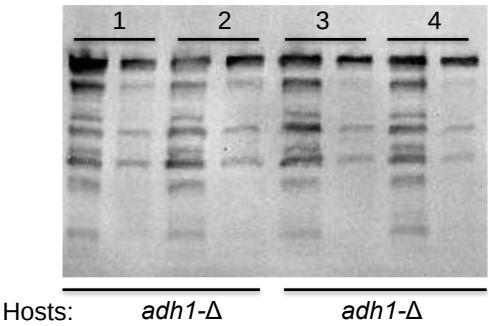
Gene	Functions
PBR1	Vacuolar proteinase B (yscB) with H3 N-terminal endopeptidase activity; serine protease of the subtilisin family;
PEP4	Vacuolar aspartyl protease (proteinase A); required for posttranslational precursor maturation of vacuolar proteinases;
UMP1	Chaperone required for correct maturation of the 20S proteasome;
RPN4	Transcription factor that stimulates expression of proteasome genes;
RKR1	RING domain E3 ubiquitin ligase; involved in ubiquitin-mediated degradation of non-stop proteins; component of ribosome-bound RQC (ribosome quality control) complex required for degradation of polypeptides arising from stalled translation; degrades products of mRNAs lacking a termination codon regardless of a poly(A) tail; functional connections to chromatin modification
HDR1	Ubiquitin-protein ligase; functions in ER retention of misfolded proteins; required for ER-associated degradation (ERAD) of misfolded proteins; genetically linked to the unfolded protein response (UPR); regulated through association with Hrd3p; contains an H2 ring finger; likely plays a general role in targeting proteins that persistently associate with and potentially obstruct the ER-localized translocon
SSM4	Ubiquitin-protein ligase involved in ER-associated protein degradation; located in the ER/nuclear envelope; ssm4 mutation suppresses mRNA instability caused by an rna14 mutation
SAN1	Ubiquitin-protein ligase; involved in proteasome-dependent degradation of aberrant nuclear proteins; targets substrates with regions of exposed hydrophobicity containing 5 or more contiguous hydrophobic residues; contains intrinsically disordered regions that contribute to substrate recognition; prefers a window of exposed hydrophobicity that causes a particular level of protein insolubility, suggesting that San1p evolved to target highly aggregation-prone proteins
SLX8	Subunit of Slx5-Slx8 SUMO-targeted ubiquitin ligase (STUbL) complex; stimulated by prior attachment of SUMO to the substrate; contains a C-terminal RING domain; forms nuclear foci upon DNA replication stress; null mutants are aneuploid, have a metaphase delay, and spindle defects including: mispositioned spindles, fish hook spindles, and aberrant spindle kinetics; required for maintenance of genome integrity like human ortholog RNF4
HSP30	Negative regulator of the H(+)-ATPase Pma1p; stress-responsive protein; hydrophobic plasma membrane localized; induced by heat shock, ethanol treatment, weak organic acid, glucose limitation, and entry into stationary phase
HSP42	Small heat shock protein (sHSP) with chaperone activity; forms barrel-shaped oligomers that suppress unfolded protein aggregation; involved in cytoskeleton reorganization after heat shock; protein abundance increases and forms cytoplasmic foci in response to DNA replication stress
LHS1	Molecular chaperone of the endoplasmic reticulum lumen
UTR2	Chitin transglycosylase; functions in the transfer of chitin to beta(1-6) and beta(1-3) glucans in the cell wall; similar to and functionally redundant with Crh1; glycosylphosphatidylinositol (GPI)-anchored protein localized to bud neck
ATG19	Receptor protein for the cytoplasm-to-vacuole targeting (Cvt) pathway;
STE3	Receptor for a factor pheromone; couples to MAP kinase cascade to mediate pheromone response;
AMS1	Vacuolar alpha mannosidase; involved in free oligosaccharide (fOS) degradation; delivered to the vacuole in a novel pathway separate from the secretory pathway
APE4	Cytoplasmic aspartyl aminopeptidase with possible vacuole function; Cvt pathway cargo protein; cleaves unblocked N-terminal acidic amino acids from peptide substrates; forms a 12-subunit homo-oligomer; M18 metalloprotease family
YDJ1	Type I HSP40 co-chaperone; involved in regulation of HSP90 and HSP70 functions; acts as an adaptor that helps Rsp5p recognize cytosolic misfolded proteins for ubiquitylation after heat shock; critical for determining cell size at Start as a function of growth rate; involved in protein translocation across membranes; member of the DnaJ family
SSA1	ATPase involved in protein folding and NLS-directed nuclear transport
SSA2	ATP-binding protein
SSA3	ATPase involved in protein folding and the response to stress
SSA4	Heat shock protein that is highly induced upon stress
SSB1	Cytoplasmic ATPase that is a ribosome-associated molecular chaperone

A


ADH1 PBR1 PEP4 UMP1 RPN4 RKR1 HDR1 SSM4 SAN1 SLX8 HSP30 HSP42 LHS1 UTR2
ATG19 STE13 AMS1 APE4 YDJ1 SSA2 SSA3 SSA4 SSB1

B


Anaerobic
Day 1
Day 2
Day 3
n-butanol (mg/L)
300
250
200
150
100
50
0
Control (70)
SSA1 (2303)
SSA1_YDJ1 (2304)
Control (70)
SSA1 (2303)
SSA1_YDJ1 (2304)
Upstream pathway
800 - 795
800 - 1454 (5'PYK2_ter)


1
2
3
4
Hosts: adh1-Δ adh1-Δ
Chaperones: - + SSA1p & YDJ1p

C

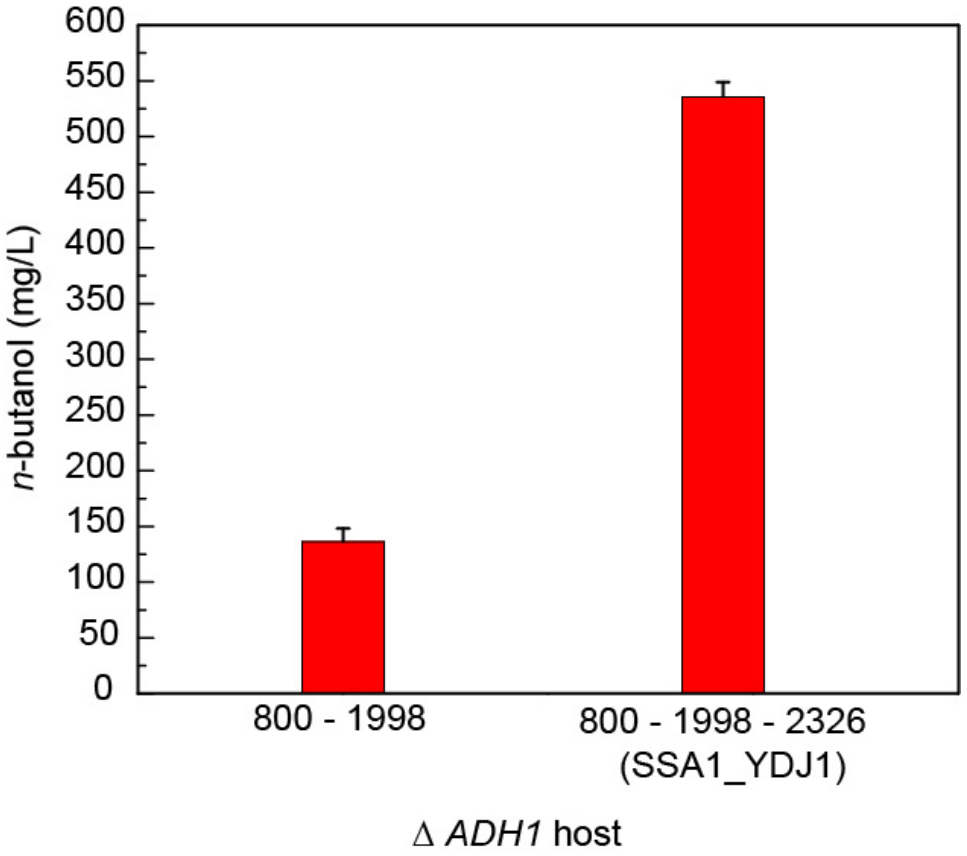


Figure 4.25. Analysis of the effect of protein quality control gene knockouts on Ter expression and chaperone co-expression of n-butanol production. (A) Western blot comparing Ter expression in hosts with different genes involved in protein quality control knocked out. All knockout hosts were derived from the BY4742 parent strain. Hosts were transformed with the pTDH3_gTdTer_TDH3t (#1800) for Ter expression. Cultures were grown under defined media with 2% *w/v* galactose for 3 d under microaerobic conditions (n = 3). Biomass were then harvested and lyted for Western blot analysis. Most of strains gave a relative similar Ter expression profile as compared to the Δ*adh1* knockout production host, where Ter is heavily degraded. Ter degradation was diminished with the deletion of *pbr1, pep4, rkr1*, and *hdr1*. Deletion of *ste3* and *ydj1* greatly diminished the expression of TdTer, suggesting the importance of these elements on Ter expression. (B) (Left) *n*-Butanol production with overexpression of YDJ1p and SSA1p. BY4741*adh1*-Δ pESC.His-Bu2 (#800) was used as the host with either pESCLeu2d-ter.adhE2 (#795) or pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454) downstream. This strain was tested with co-expression of SSA1 (pRS316_TDH3_SSA1_TDH3, #2303) or SSA1 and YDJ1 (pRS316_SSA1_YDJ1, #2304) on a plasmid with the CEN ARS origin. Cultures were grown in defined drop out media with 2% (*w/v*) galactose under anaerobic conditions. Samples were harvested every 24 h up to 3 d to measure production titer (n = 3). Overexpression SSA1 gave a higher titer when the production was conducted with pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454). (Right) Characterizing TdTer expression with the co-expression of SSA1P and YDJ1p from the production experiment. There was not significant difference on expression profile with and without the expression of SSA1p and YDJ1P chaperones. (C) n-Butanol production with SSA1 and YDJ1 co-expressed on a high copy number plasmid. BY4741*adh1*-Δ pESC.His-Bu2 (#800) pVYY1.5.1 (#1998) was co-transformed with and empty vector control (pESCLeu2d, #70) or the plasmid carrying SSA1 and YDJ1 (pESCLeu2d_YDJ1_SSA1, #2326). Production was performed under anaerobic conditions for 5 d of growth (n = 3).

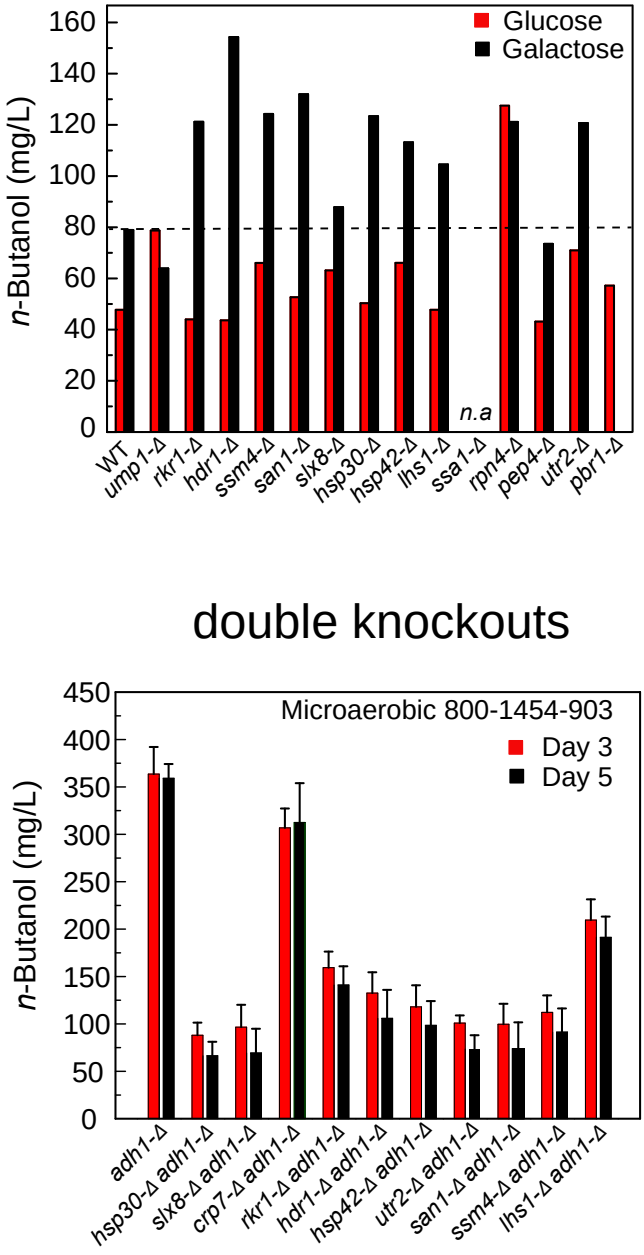


Figure 4.26. *n*-Butanol production with single and double knockout hosts. All single knockout hosts contain deletions in genes involved in either the ubiquitination pathway or heat shock response. Double knockout hosts have the major alcohol dehydrogenase (*adh1*) deleted in addition to the original knockout. All hosts were derived from the BY4742 parent strain. Hosts were transformed with the *n*-butanol pathway (pESCHis-Bu2, #800; pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer, #1454; pESCUra-P(cons)PDCzm.eutE, #903). For single knockouts, cultures were grown in defined drop out media under microaerobic conditions for 3 d with either 2% (w/v) glucose (red) or galactose (black) (n = 3). About 60% of these single knockouts gave a higher production titer as compared to the parent strain under galactose condition. For double knockouts, cultures were grown in defined drop out media with 2% (w/v) galactose under microaerobic conditions for either 3 or 5 d (n = 3). Product titer dropped by 2- to 4-fold in the double knockout hosts as compared to the Δ*adh1* control.

Developing a genetic selection to improve *n*-butanol production. Our optimization data suggests that rational design approaches to greatly improving the *n*-butanol titer in yeast are challenging. Given the success in achieving large increases in yield in *E. coli* by adaptive evolution (*Chapter 2*), we were interested in implementing a similar design in *S. cerevisiae*. A major challenge in this area is that S. cerevisiae has been evolved over a long period for ethanol production and has many redundant routes for fermentation of ethanol as well as the secondary product, glycerol. As such, all major fermentation pathways needed to be eliminated to replace ethanol and glycerol pathways with the *n*-butanol pathway as the only route for redox balance and ATP generation. To do so, all five major alcohol dehydrogenases (ADH1, ADH5, ADH6, ADH4, ADH3) and two glycerol-3- phosphate dehydrogenases (GPD1 and GPD2) were deleted to prevent production of ethanol and glycerol, resulting a septuple knockout strain (BY4741Δ*7*) (*Figure. 4.*27). This strain grew very slowly due to the tendency of *S. cerevisiae* to grow via fermentative pathways even under aerobic conditions. Various *n*-butanol pathway variants, with different ALDH-ADH pairs, were integrated in BY4741Δ7 host for adaptive evolution in rich media (YPG). Even with the *n*-butanol pathway, BY4741Δ*7* still showed a significant growth defect. However, after only three passage of cultures, they began to exhibit a highly-improved growth phenotype, reaching OD_{600} = 4-5 after 24 to 48 h growth in YPG media under anaerobic conditions (*Figure. 4.29).* Upon the observation of improved growth phenotype, we harvested cultures for *n*-butanol production analysis. Our preliminary data shows that the improved growth rate seems to correlate with a concomitant increase *n*-butanol titer as compared to the initial culture of approximately 3-fold (*Figure. 4. 29*). This finding shows promise for the isolation of higher productively *n*-butanol strains with longer time frame of evolution.

4.4. Conclusion

In this Chapter, we describe the construction of an *n*-butanol pathway for *S. cerevisiae*. Using the same pathway enzymes, we found that initial production titers were approximately 400-fold lower than the equivalent pathway in *E. coli*, suggesting that major challenges exist in heterologous protein expression or building block availability. We have identified the *trans*-enoyl-CoA reductase (Ter) (*Figure 4.1*) as the pathway bottleneck as increases in *n*-butanol titer were found to correlate well with increased Ter specific activity. We chose to optimize TdTer expression and use the TdTer transcript as a target for basic studies to understand the fate of heterologous transcripts. First, multiple approaches were explored to improve TdTer expression with a result of increasing *n*-butanol titer. Promoter screening showed that strong promoter discovered from the anaerobic fermentation, pCCW12 gave the highest production titer (480 ± 5 mg L^{-1}). Consistently with published literature, codon-optimizing TdTer with codon table generated by glycolytic genes only from *S. cerevisiae* slightly improved titer. Revisiting published translation efficiency and proteomic data from *S. cerevisiae* showed that both glycolytic and ribosomal genes have high translation efficiency and are highly abundant. UTR sequences from both glycolytic and ribosomal genes were identified from the published RNA-Seq data. Introducing those UTRs to the TdTer

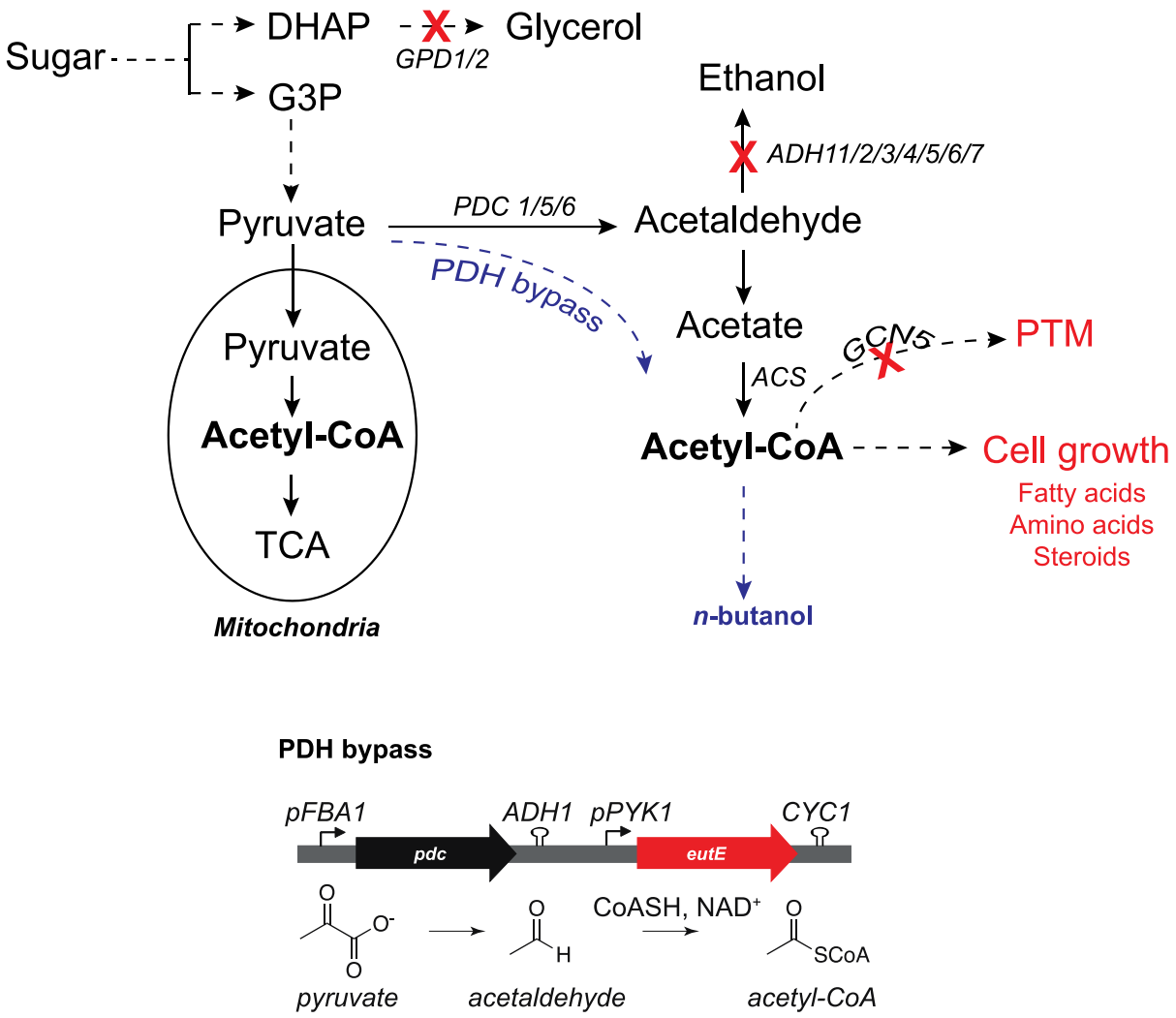
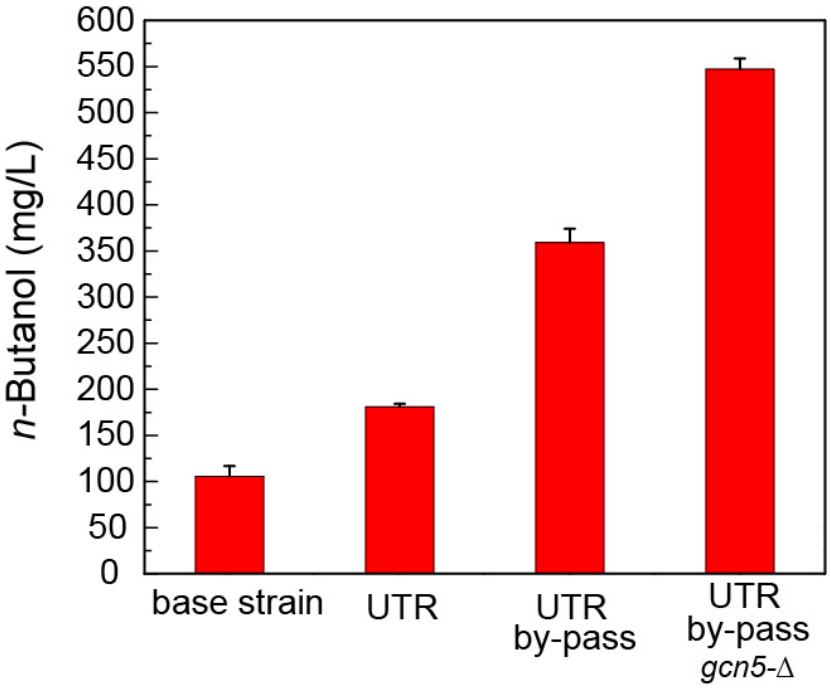


Figure 4.27. Approaches to improve cytosolic acetyl-CoA pool in *S. cerevisiae*. Knocking out alcohol dehydrogenases and glycerol phosphate dehydrogenases reduce carbon flux going to ethanol and glycerol production. Acetyl transferase (*GCN5*) was also knocked out to diminish the usage of acetyl-CoA for posttranscriptional modification. The pyruvate dehydrogenase (PDH) bypass pathway was included to drive the flux from pyruvate to cytosolic acetyl-CoA.

A



B

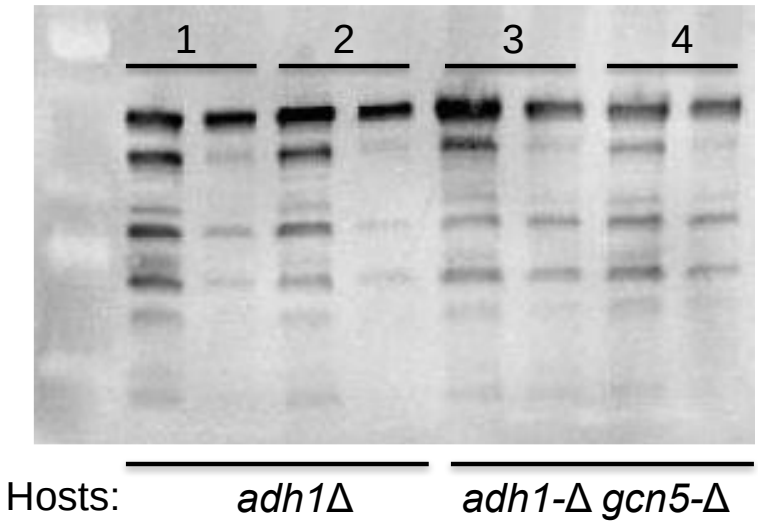


Figure 4.28. Analysis of the effect of *gcn5* deletion. (A) *n*-Butanol production combining TdTer UTR optimization, the PDHc bypass, and deletion of *gcn5*. Cultures were grown in defined drop out media with 2% (*w/v*) galactose under microaerobic conditions for 3 d (n = 3). The following plasmids were used. Base strain: pESCHis-Bu2 (#800), pESCLeu2d-ter-adhE2 (#795), and pESCUra (#70); UTR: pESCHis-Bu2 (#800), pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454), and pESCUra (#70); Both UTR and bypass: pESCHis-Bu2 (#800), pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer (#1454), and pESCUra-P(cons)PDCzm.eutE (#903). Control host: BY4741*adh1*-Δ. *GCN5* knockout host: BY4741*adh1*-Δ *gcn5*-Δ. (B) Production cultures were harvested, lysed, and analyzed by Western blotting with the Ter antibodies. Lane 1 (soluble fraction) and 2 (insoluble fraction) are duplicates from the BY4741*adh1*-Δ host. Lane 3 (soluble fraction) and 4 (insoluble fraction) are duplicates from the BY4741*adh1*-Δ *gcn5*-Δ host.

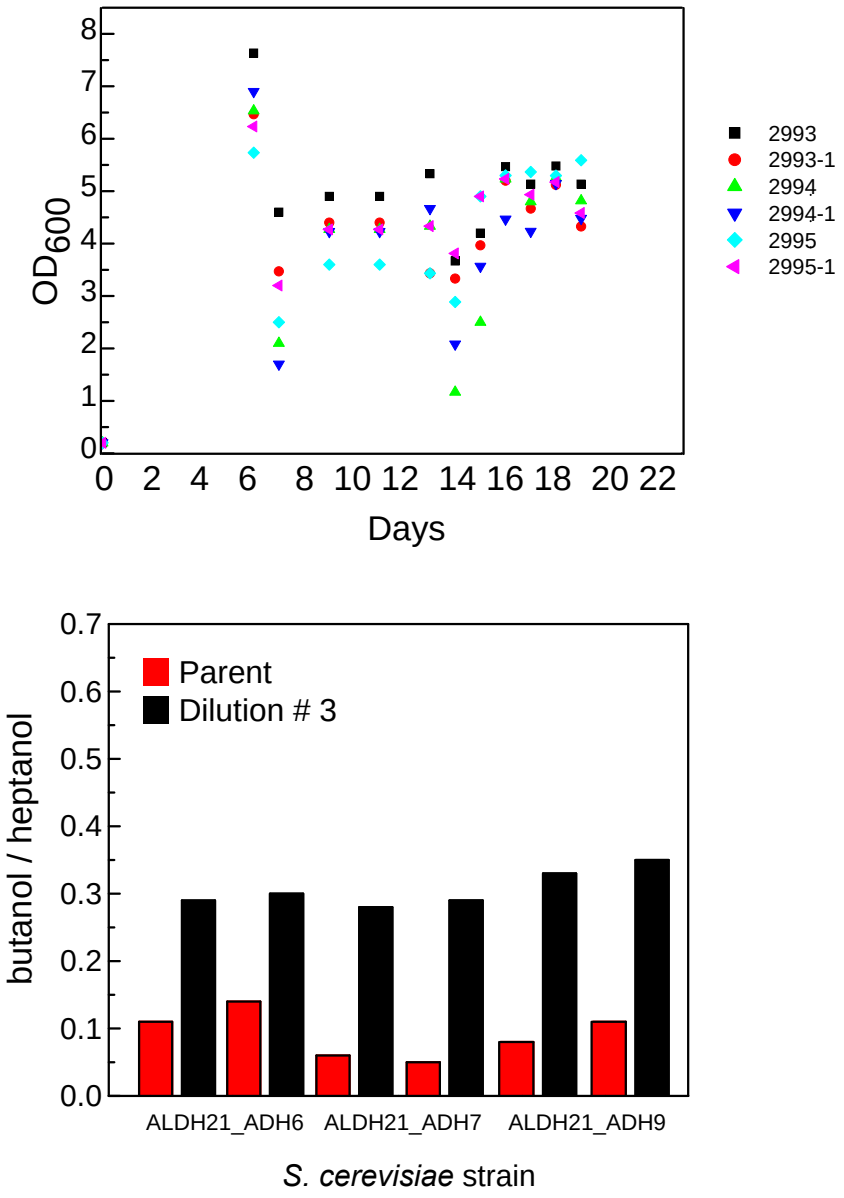


Figure. 4.29. Cell growth and *n*-butanol profiles for the adaptive evolution culture with the BY4741Δ7 host. BY4741 *adh1*-Δ *adh5*-Δ *adh6*-Δ *adh4*Δ::eutE *adh3*Δ::pdc *gpd1*-Δ *gpd2*-Δ *yprc*15Δ::Pha_hbd_Crt *yprc*3Δ::Ter_ADLHx_ADHx. x represents different alcohol and aldehyde dehydrogenases. Cultures were grown in YPG with 2% *w/v* galactose under anaerobic conditions. OD_{600} were measured before every dilution. Supernatant from cultures after 3 dilutions were extracted with toluene and analyzed by GC-MS for *n*-butanol production. Cultures showed improved growth phenotype after 3 dilutions and the evolved culture showed improved *n*-butanol production titer compared to the initial culture.

expression cassette improved overall *n*-butanol titer by 10-fold. Additionally, other factors have been reported to affect protein expression were studied, which included selection marker and terminators. Overall, introducing UTRs to the bottleneck step gave the greatest improvement on production titer, achieved 340 ± 10 mg L^{-1}.

From a basic science perspective, we have examined heterologous protein expression at many levels and stages of the mRNA life scale. Both targeted transcript quantification and global transcriptome experiments showed that transcript abundance was not limited compared with the highly-expressed endogenous transcript, TDH3, indicating the problems arise post-transcriptionally. Our preliminary data also suggested that TdTer was 5′-capped and should be able to initiate translation through the typical cap-dependent mechanism. However, targeted transcript quantification from the polysome fraction showed that TdTer had a lower translation efficiency compared with TDH3. Furthermore, polysome profiling showed global translation was down regulated when cells carried the entire *n*-butanol pathway, possibility related to cell stress. Consistent with this observation, RNA-seq experiments show that transcriptions involved in ribosome biogenesis, translation, and protein quality control are differentially regulated. Indeed, overexpressing genes involved in ribosome biogenesis and translation improved *n*-butanol titer modestly.

Lastly, Western blot analysis and enzyme assays were used to examine heterologous expression at the protein level. Western analysis showed that both TdTer and AdhE2 were highly degraded and insoluble. Screening of protease and heat shock protein knockouts revealed protein degradation was alleviated with certain knockouts. In addition, overexpressing proteins in the protein folding pathway (*SSA1p* and *YDJ1p*) improved *n*-butanol production titer by 4-fold (540 ± 10 mg L^{-1}).

We took multiple routes to address the challenge of limited cytosolic acetyl-CoA pool. First, the bypass pathway was overexpressed to drive the carbon flux from pyruvate to cytosolic acetyl-CoA. Second, the acetyl-CoA transferase, *GCN5*, which uses acetyl-CoA as a donor for histone modification was knockout with the goal to improve acetyl-CoA. Both approaches had showed an improved *n*-butanol production titer. Taking all together, introducing the UTR to Ter, overexpressing the bypass pathway, and deleting *GCN5*, we achieved the production titer of n-butanol to 550 ± 10 mg L^{-1}.

Finally, we constructed the BY4741Δ7 host with the *n*-butanol pathway integrated in the genome in order to test the possibility of using adaptive evolution to improve product titers. Our initial selection experiment showed both improved cell growth phenotype and *n*-butanol titer. This suggested adaptive evolution could be a promising approach to improve *n*-butanol production profile.

4.5. References

1. Liu, L.; Redden, H.; Alper, H. S. Frontiers of Yeast Metabolic Engineering: Diversifying beyond Ethanol and Saccharomyces. *Current Opinion in Biotechnology*. 2013, pp 1023–1030.
2. Nevoigt, E. Progress in Metabolic Engineering of *Saccharomyces Cerevisiae*. *Microbiol. Mol. Biol. Rev.* **2008**, *72*, 379–412.
3. Wen, M.; Bond-Watts, B. B.; Chang, M. C. Production of Advanced Biofuels in Engineered E. Coli. *Curr Opin Chem Biol* **2013**, *17*, 472–479.
4. Global 1-Butene Market By Application (LLDPE, HDPE, Valeraldehyde, and Others), By Region (North America, Europe, Asia-Pacific, and Rest of the World) - Demand-Supply, Price, Analysis and Forecast To 2021 http://www.micromarketmonitor.com/market-report/1-butene-reports-6686141708.html (accessed Dec 27, 2017).
5. Bond-Watts, B. B.; Bellerose, R. J.; Chang, M. C. Enzyme Mechanism as a Kinetic Control Element for Designing Synthetic Biofuel Pathways. *Nat Chem Biol* **2011**, *7*, 222–227.
6. Ryan, O. W.; Poddar, S.; Cate, J. H. D. Crispr–Cas9 Genome Engineering in *Saccharomyces Cerevisiae* Cells. *Cold Spring Harb. Protoc.* **2016**, *2016*, 525–533.
7. Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6*, 343–345.
8. Blaisse, Michael R, Chang, M. C. Y. Exploring Carbon-Carbon Bond Formation in Nature for the Production of Biorenewable Fuels and Chemicals, UC Berkeley, 2016.
9. Davis, Matthew A. ; Chang, M. C. Y. Exploring in Vivo Biochemistry with C4 Fuel and Commodity Chemical Pathways, UC Berkeley, 2015.
10. Ryan, O. W.; Skerker, J. M.; Maurer, M. J.; Li, X.; Tsai, J. C.; Poddar, S.; Lee, M. E.; DeLoache, W.; Dueber, J. E.; Arkin, A. P.; et al. Selection of Chromosomal DNA Libraries Using a Multiplex CRISPR System. *Elife* **2014**, *3* (August2014), 1–15.
11. Benchling (https://benchling.com/).
12. Davis, J. T.; Moore, R. N.; Imperiali, B.; Pratt, A. J.; Kobayashi, K.; Masamune, S.; Sinskey, A. J.; Walsh, C. T.; Fukui, T.; Tomita, K. Biosynthetic Thiolase from Zoogloea Ramigera. I. Preliminary Characterization and Analysis of Proton Transfer Reaction. *J. Biol. Chem.* **1987**, *262*, 82–89.
13. Li, J.; Rahmeh, A.; Morelli, M.; Whelan, S. P. J. A Conserved Motif in Region V of the Large Polymerase Proteins of Nonsegmented Negative-Sense RNA Viruses That Is Essential for MRNA Capping. *J. Virol.* **2008**, *82*, 775–784.
14. Bolstad, B. M.; Irizarry, R. a; Astrand, M.; Speed, T. P. A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Variance and Bias. *Bioinformatics* **2003**, *19*, 185–193.
15. Clarkson, B. K.; Gilbert, W. V.; Doudna, J. A. Functional Overlap between EIF4G Isoforms in *Saccharomyces Cerevisiae*. *PLoS One* **2010**, *5*.
16. Ashe, M. P.; De Long, S. K.; Sachs, A. B. Glucose Depletion Rapidly Inhibits Translation Initiation in Yeast. *Mol. Biol. Cell* **2000**, *11*, 833–848.
17. Daran-Lapujade, P.; Rossell, S.; van Gulik, W. M.; Luttik, M. A. H.; de Groot, M. J. L.;

Slijper, M.; Heck, A. J. R.; Daran, J.-M.; de Winde, J. H.; Westerhoff, H. V.; et al. The Fluxes through Glycolytic Enzymes in *Saccharomyces Cerevisiae* Are Predominantly Regulated at Posttranscriptional Levels. *Proc. Natl. Acad. Sci.* **2007**, *104*, 15753–15758.

18. Blazeck, J.; Alper, H. S. Promoter Engineering: Recent Advances in Controlling Transcription at the Most Fundamental Level. *Biotechnology Journal*. 2013, pp 46–58.
19. Dvir, S.; Velten, L.; Sharon, E.; Zeevi, D.; Carey, L. B.; Weinberger, A.; Segal, E. Deciphering the Rules by Which 5 ’-UTR Sequences Affect Protein Expression in Yeast. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E2792–E2801.
20. Redden, H.; Morse, N.; Alper, H. S. The Synthetic Biology Toolbox for Tuning Gene Expression in Yeast. *FEMS Yeast Research*. 2015.
21. Mignone, F.; Gissi, C.; Liuni, S.; Pesole, G. Untranslated Regions of MRNAs. *Genome Biol.* **2002**, *3*, REVIEWS0004.
22. Gallie, D. R. A Tale of Two Termini: A Functional Interaction between the Termini of an MRNA Is a Prerequisite for Efficient Translation Initiation. *Gene*. 1998, pp 1–11.
23. De Godoy, L. M. F.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Fröhlich, F.; Walther, T. C.; Mann, M. Comprehensive Mass-Spectrometry-Based Proteome Quantification of Haploid versus Diploid Yeast. *Nature* **2008**, *455*, 1251–1254.
24. Ingolia, N. T.; Ghaemmaghami, S.; Newman, J. R. S.; Weissman, J. S. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science.* **2009**, *324*, 218–223.
25. Nagalakshmi, U.; Wang, Z.; Waern, K.; Shou, C.; Raha, D.; Gerstein, M.; Snyder, M. The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science (80-.).* **2008**, *320*, 1344–1349.
26. Lee, M. E.; Aswani, A.; Han, A. S.; Tomlin, C. J.; Dueber, J. E. Expression-Level Optimization of a Multi-Enzyme Pathway in the Absence of a High-Throughput Assay. *Nucleic Acids Res.* **2013**, *41*, 10668–10678.
27. Partow, S.; Siewers, V.; Bjørn, S.; Nielsen, J.; Maury, J. Characterization of Different Promoters for Designing a New Expression Vector in *Saccharomyces Cerevisiae*. *Yeast* **2010**, *27*, 955–964.
28. Lin, Y.; Chomvong, K.; Acosta-Sampson, L.; Estrela, R.; Galazka, J. M.; Kim, S. R.; Jin, Y. S.; Cate, J. H. Leveraging Transcription Factors to Speed Cellobiose Fermentation by *Saccharomyces Cerevisiae*. *Biotechnol. Biofuels* **2014**, *7*.
29. Hanson, G.; Coller, J. Translation and Protein Quality Control: Codon Optimality, Bias and Usage in Translation and MRNA Decay. *Nature Reviews Molecular Cell Biology*. 2018, pp 20–30.
30. Elena, C.; Ravasi, P.; Castelli, M. E.; Peir??, S.; Menzella, H. G. Expression of Codon Optimized Genes in Microbial Systems: Current Industrial Applications and Perspectives. *Frontiers in Microbiology*. 2014.
31. Curran, K. A.; Leavitt, J. M.; Karim, A. S.; Alper, H. S. Metabolic Engineering of Muconic Acid Production in *Saccharomyces Cerevisiae*. *Metab. Eng.* **2013**, *15*, 55–66.
32. Agashe, D.; Martinez-Gomez, N. C.; Drummond, D. A.; Marx, C. J. Good Codons, Bad Transcript: Large Reductions in Gene Expression and Fitness Arising from Synonymous Mutations in a Key Enzyme. *Mol. Biol. Evol.* **2013**, *30*, 549–560.

33. Ingolia, N. T.; Brar, G. A.; Rouskin, S.; Mcgeachy, A. M.; Weissman, J. S. The Ribosome Profiling Strategy for Monitoring Translation in Vivo by Deep Sequencing of Ribosome-Protected MRNA Fragments.

34. Lanza, A. M.; Curran, K. A.; Rey, L. G.; Alper, H. S. A Condition-Specific Codon Optimization Approach for Improved Heterologous Gene Expression in *Saccharomyces Cerevisiae*. *BMC Syst. Biol.* **2014**, *8*.

35. Wiedemann, B.; Boles, E. Codon-Optimized Bacterial Genes Improve L-Arabinose Fermentation in Recombinant *Saccharomyces Cerevisiae*. *Appl. Environ. Microbiol.* **2008**, *74*, 2043–2050.

36. Curran, K. A.; Karim, A. S.; Gupta, A.; Alper, H. S. Use of Expression-Enhancing Terminators in *Saccharomyces Cerevisiae* to Increase MRNA Half-Life and Improve Gene Expression Control for Metabolic Engineering Applications. *Metab. Eng.* **2013**, *19*, 88–97.

37. Xia, P. F.; Zhang, G. C.; Liu, J. J.; Kwak, S.; Tsai, C. S.; Kong, I. I.; Sung, B. H.; Sohn, J. H.; Wang, S. G.; Jin, Y. S. GroE Chaperonins Assisted Functional Expression of Bacterial Enzymes in *Saccharomyces Cerevisiae*. *Biotechnol. Bioeng.* **2016**, *113*, 2149–2155.

38. Chen, X.; Li, S.; Liu, L. Engineering Redox Balance through Cofactor Systems. *Trends Biotechnol* **2014**, *32*, 337–343.

39. Hellen, C. U. T.; Sarnow, P. Internal Ribosome Entry Sites in Eukaryotic MRNA Molecules. *Genes and Development*. 2001, pp 1593–1612.

40. Lee, A. S.-Y.; Burdeinick-Kerr, R.; Whelan, S. P. J. A Ribosome-Specialized Translation Initiation Pathway Is Required for Cap-Dependent Translation of Vesicular Stomatitis Virus MRNAs. *Proc. Natl. Acad. Sci.* **2013**, *110*, 324–329.

41. Shenton, D.; Smirnova, J. B.; Selley, J. N.; Carroll, K.; Hubbard, S. J.; Pavitt, G. D.; Ashe, M. P.; Grant, C. M. Global Translational Responses to Oxidative Stress Impact upon Multiple Levels of Protein Synthesis. *J. Biol. Chem.* **2006**, *281*, 29011–29021.

42. Kang, H. A.; Hershey, J. W. B. Effect of Initiation Factor EIF-5A Depletion on Protein Synthesis and Proliferation of *Saccharomyces Cerevisiae*. *J. Biol. Chem.* **1994**, *269*, 3934–3940.

43. Young, D. J.; Guydosh, N. R.; Zhang, F.; Hinnebusch, A. G.; Green, R. Rli1/ABCE1 Recycles Terminating Ribosomes and Controls Translation Reinitiation in 3′UTRs In Vivo. *Cell* **2015**, *162*, 872–884.

Appendix 1: *Complete list of constructs*

No.	Plasmid
1328	pESC_Leu_adhE2_EgTer (YCO)
1383	pESC_His_Erg10_hbd_crt (No His Tag)
1384	pESC_His_Erg10_hbd_crt (C terminal Hisx10)
1385	pESC_Leu_AdhE2(YCO)_TdTer
1386	pESC_Leu_adhE2(YCO)_EgTer
1387	pESC_Leu_adhE2(YCO)_EgTer(YCO)
1413	pESC_Leu_AdhE2_5'UTRTPI1_TdTer
1414	pESC_Leu_AdhE2_5'UTR_TDH2(YJR009C))TdTer
1415	pESC_Leu_AdhE2_5'UTR_FBA1(YKL060C)TdTer
1416	pESC_Leu_AdhE2_5'UTR_GPM1(YKL152C)TdTer
1417	pESC_Leu_AdhE2_5'UTR_(YLR075W))TdTer
1418	pESC_Leu_AdhE2_5'UTR_(YHL001W)TdTer
1419	pESC_Leu_AdhE2_5'UTR_(YJL177W)TdTer
1424	pESC_Leu_AdhE2_TdTer_3'UTR FBA1
1425	pESC_Leu_AdhE2_TdTer_3'UTR (YJL177W)
1426	pESC_Leu_AdhE2_5'UTR FBA_TdTer_3'UTR FBA1#1426
1427	pESC_Leu_AdhE2_5'UTR FBA_TdTer_3'UTR YJL177W#1427
1428	pESC_Leu_adhE2_MECR1
1429	pESC_Leu_adhE2_Hisx10MECR1_#1429
1453	pESC_Leu_AdhE2_5'UTR_TDH1_TdTer
1454	pESC_Leu_AdhE2_5'UTR_PYK2_TdTer
1455	pESC_Leu_AdhE2_5'UTR_PGI1TdTer
1456	pESC_Leu_AdhE2_5'UTR_PFK1_TdTer
1457	pESC_Leu_AdhE2_5'UTR_PFK2_TdTer
1458	pESC_Leu_AdhE2_5'UTR_ENO1_TdTer
1459	pESC_Leu_AdhE2_5'UTR_ENO2_TdTer
1460	pESC_Leu_AdhE2_5'UTR_CDC19_TdTer
1461	pRS313
1462	pRS314
1463	pRS315
1464	pESC_Leu_AdhE2_5'UTR_TDH3_TdTer
1465	peSC_Leu_adhE2_MECR1 (No MP No His))
1471	pESC_Leu_adhE2_Hisx10MECR1 (No MP)
1472	pESC_Leu_AdhE2_5'UTR FBA_MECR1_3'UTR FBA1
1473	pESC_Leu_AdhE2_5'UTR_(YJL177W)MECR1
1474	pESC_Leu_AdhE2_5'UTR_(YHL001W)MECR1
1475	pESC_Leu_AdhE2_5'UTR_TDH2(YJR009C))MECR1
1525	pESC_Leu_AdhE2_CCW12_5'UTR_PYK2_TdTer
1534	pESC_Leu_AdhE2_TDH3_5'UTR_PYK2_TdTer
1551	pESC_LeuAdhE2_5'PYK2_TdTer(S.c. gly)
1552	pESC_LeuAdhE2_5'PYK2_TdTer(S.c)
1556	pESC_Leu_AdhE2_CCW12_5'UTR_PYK2_TdTer(gS.c)
1557	pESC_Leu_AdhE2_TDH3_5'UTR_PYK2_TdTer (S.c glycolytic gene codon optimized)
1558	pESC_Leu_AdhE2_CCW12_5'UTR_PYK2_TdTer(S.c codon optimized))
1559	pESC_Leu_AdhE2_TDH3_5'UTR_PYK2_TdTer (S.c codon optimized)

1568	pESC_Leu_AdhE2(gS.c)_5'UTR_PYK2_TdTer
1619	pESC_Leu_1556_TdTer_CDC19
1620	pESC_Leu_1556_TdTer_PGI1
1621	pESC_Leu_1556_TdTer_PGK1
1622	pESC_Leu_1556_TdTer_ENO1
1623	pESC_Leu_1556_TdTer_ENO2
1624	pESC_Leu_1556_TdTer_TDH2
1625	pESC_Leu_1556_TdTer_GPM1
1626	pESC_Leu_1556_TdTer_PFK2
1631	pESC_Leu_1556_TdTer_TPI1
1632	pESC_Leu_1556_TdTer_FBA1
1633	pESC_Leu_1556_TdTer_PYK2
1667	pESC_Leu_1556_TdTer_TDH1 1667
1779	pSNR52_HO1
1782	pCAS_Rgt2
1799	pVYY1.0.0_2
1800	pRS316_TDH3_gTdTerTDH3
1801	pCas_ADH1
1821	pVYY1.1.0
1822	pVYY1.2.0
1823	pVYY1.3.0
1824	pVYY1.4.0
1825	pVYY1.5.0
1826	pVYY1.6.0
1827	pVYY1.8.0
1828	pVYY1.C.0
1832	pRS426-BT
1833	pRS316-BT-BEST
1846	pRS316_BT-BEST
1848	pVYY1.7.0
1849	pVYY2.1.0
1850	pVYY2.2.0
1851	pVYY2.3.0
1852	pVYY2.4.0
1853	pVYY2.5.0
1854	pVYY2.6.0
1855	pVYY2.8.0
1856	pVYY2.C.0
1858	pVYY2.7.0
1879	pVYY1.0.0.5
1880	pVYY3.1.0
1881	pVYY3.2.0
1882	pET31B_T7_S2
1930	pVYY3.C1.0
1931	pVYY3.C2.0
1943	pCAS_Pphe_BSAI
1972	pVYY1.2.1
1973	pVYY1.4.1

1974	pVYY1.6.1
1975	pVYY1.7.1
1976	pVYY1.8.1
1977	pVYY1.C.1
1978	pVYY3.C3.0
1982	pVYY3.C4.0
1997	pVYY1.1.1
1998	pVYY1.5.1
1999	pVYY3.C5.0_Broccoli
2000	pVYY3.C6.0_dBroccoli
2001	pVYY1.0.1_1
2002	pVYY1.3.1
2046	pCAS_Pphe-BsaI_NAT
2047	pCAS_Pphe-_NAT_PBR1(g2)
2048	pCAS_Pphe-_NAT_PEP4(g1)
2049	pET16b-His-Ter (E.coli)
2050	pVYY_His_Ter
2185	pRS315_GroEL
2186	pRS316_TDH3p__TDH3t
2187	pESC_Leu_adhE2_DnaJ
2188	pESC_Leu_DnaJ_DnaK
2192	NONE
2198	pESC_Leu_GroEL_GroES
2199	pESC-Leu_YDJ1
2200	pCAS_Pphe-_NAT_Adh1
2201	NONE
2214	pCAS_Pphe-_NAT_g2Adh1
2215	pCAS_Pphe-_NAT_g3Adh1
2236	pCAS_Pphe-_NAT_g4ADH1
2303	pRS316_TDH3_SSA1_TDH3
2304	pRS316_SSA1_YDJ1
2307	pCAS_Pphe-_NAT_g1GPD1
2308	pCAS_Pphe-_NAT_g2GPD1
2326	pESC-Leu_YDJ1_SSA1
2327	pESC-Leu_SSA1
2328	pESC-Ura-SSA1
2329	pESC-Ura-SSA1_YDJ1
2353	pESC-URA_TIF51Ap_TIF51A
2354	pESC-URA_TIF51Ap_TIF51A_Gal_TIF51B_TIF51Bt
2355	pESC_URA.P(cons)PDCzm. eutE_SSA1
2356	pESC_Leu_adhE2_MT_Ter
2357	pESC_Leu_MT_adhE2_MT_Ter
2358	pESC_HIS_Bu2_MT-PhA
2390	pESC_HIS_Bu2_MT-PhA_MT-hbd_MT-Crt
2391	pESC_Leu_CCW12Ter_TDH3ALD5_FBA1ADH_2
2401	pESC_Leu_5'PYK2_AdhE2_5'UTR_PYK2_TdTer
2413	pESC_Ura_903_Prime_SSA1_YDJ1
2414	pRS316_TDH3p_TdTer_eGFP_TDH3t

2415	pRS316_TDH3p_5'PYK2_TdTer_eGFP_TDH3t
2498	pUC-UAS1B16-Leum
2499	pUC-UAS1B20-Leum
2500	pUC-UAS1B28-Leum
2501	pUC-UAS1B16-TEF(504)
2502	pUC-UAS1B16-TEF(272)
2515	pCAS_Pphe-_NAT_g1ADH5
2516	pCAS_Pphe-_NAT_g2ADH5
2517	pCAS_Pphe-_NAT_g3ADH5
2518	pCAS_Pphe-_NAT_g1ADH6
2519	pCAS_Pphe-_NAT_g2ADH6
2520	pCAS_Pphe-_NAT_g3ADH6
2521	pCAS_Pphe-_NAT_g1GCY1
2522	pCAS_Pphe-_NAT_g2GCY1
2523	pCAS_Pphe-_NAT_g3GCY1
2556	pESC_Leu. (5'UTR)Tdter_Aldh5_ADH2
2557	pESC_Leu. (5'UTR)Tdter_Aldh5_ADH8
2558	pESC_Leu. (5'UTR)Tdter_Aldh5_ADH22
2559	pESC_Leu. (5'UTR)Tdter. Aldh6.Adh2.
2560	pESC_Leu. (5'UTR)Tdter. Aldh6.Adh8
2561	pESC_Leu. (5'UTR)Tdter. Aldh6.Adh22.
2562	pESC_Leu. (5'UTR)Tdter. Aldh7.Adh2
2563	pESC_Leu. (5'UTR)Tdter. Aldh7.Adh8
2564	pESC_Leu. (5'UTR)Tdter. Aldh7.Adh22
2565	pESC_Leu. (5'UTR)Tdter. Aldh10.Adh2
2566	pESC_Leu. (5'UTR)Tdter. Aldh10.Adh8
2567	pESC_Leu. (5'UTR)Tdter. Aldh10.Adh22
2568	pESC_Leu. (5'UTR)Tdter. Aldh12.Adh2
2569	pESC_Leu. (5'UTR)Tdter. Aldh12.Adh8
2570	pESC_Leu. (5'UTR)Tdter. Aldh12.Adh22
2571	pESC_Leu_HSP30p_AdhE2_HSP26p_uPYK2_TdTer
2578	pESC-URA_HSP26pTdTer_HSP30_AdhE2
2589	None _ S288C 1n LYP1::GH1-1 TRP1::CDT1 N209S F262Y
2590	pESC_URA_ANB1
2591	pESC_URA_RPS14B
2592	pESC_URA_TMA10
2599	pESC_URA_DBP2
2600	pESC_URA_RLI1
2601	pCAS_Pphe-_NAT_g5ADH1
2602	pCAS_Pphe-_NAT_g6ADH1
2603	pCAS_Pphe-_NAT_g1GPD2
2604	pCAS_Pphe-_NAT_g1DHH1
2605	pCAS_Pphe-_NAT_g2DHH1
2606	pCAS_Pphe-_NAT_g1COS12_ORF
2607	pCAS_Pphe-_NAT_g1LEU2
2608	pCAS_Pphe-_NAT_g1HIS3
2648	pESC_Ura_903_Prime_RPS14B
2656	pESC_Ura_Bypass_CYC1

2662	pCas_TetR
2663	pTargetF_g1PhaA
2664	pTargetF_g3PhaA
2672	pCAS_Pphe-_NAT_g2LEU2
2701	None
2746	pTargetF_g4PhaA_g1Km_g3Cb
2759	pESC_Leu. (5'UTR)Tdter. Aldh21.Adh2
2760	pESC_Ura_903_Prime_RLI1_v2
2782	pCAS_Pphe-_NAT_g1ADH4
2783	pCAS_Pphe-_NAT_g1ADH3
2784	pCRISPR_gibson_1guide_2409pcnB
2786	pCRISPR_gibson_1Guide
2792	pCRISPR_Tet
2794	pCRISPR_gibson_1guide_2406_rpoC
2796	pESC_Leu. (5'UTR)Tdter. Aldh21.Adh3
2797	pESC_Leu. (5'UTR)Tdter. Aldh21.Adh4
2798	pESC_Leu. (5'UTR)Tdter. Aldh21.Adh5
2799	pESC_Leu. (5'UTR)Tdter. Aldh21.Adh6
2800	pESC_Leu. (5'UTR)Tdter. Aldh21.Adh7
2801	pESC_Leu. (5'UTR)Tdter. Aldh21.Adh9
2802	pESC_Leu. (5'UTR)Tdter. Aldh21.Adh10
2803	pESC_Leu. (5'UTR)Tdter. Aldh21.Adh12
2804	pESC_Leu. (5'UTR)Tdter. Aldh21.Adh13
2805	pESC_Leu. (5'UTR)Tdter. Aldh21.Adh14
2811	pKD46-Cas9-RecA-Cure_Sp
2935	pCRISPR_Tet_g1Km
2936	pCRISPR_Tet_g3Cb
2937	pCRISPR_Tet_g1Cm
2938	pCRISPR_gibson_1guide_2403g2NADP

Appendix 2: *Strains, plasmids, oligonucleotides, sequences, and genome sequencing results for Chapter 2*

Appendix 2.1: Strains

E. coli DH10B was used for DNA construction. *E. coli* DH1 (ATCC 39936), DH1Δ5, BW25113Δ5-T1R, DH1Δ5_2406_pcnB(R149L), DH1Δ5_2406_rpoC(M466L), DH1Δ5_2406_pcnB(R149L)_rpoC(M466L) were used for production and evolution experiments.

Organism	Name	Description	Source
E. coli	DH10B	F- endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ-	Invitrogen
E. coli	DH1Δ*5*	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC*	Dr. Miao Wen
E. coli	BW25113Δ*5*-T1R	BW25113 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC* ∆*fhuA*, P1 transduced *fhuA*:Km^R from 1637 parent to 1435 then recycled Km marker	Dr. Matthew Davis
E. coli	DH1Δ5_2406_pcnB(R149L)	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC pcnB*(R19L)	This study
E. coli	DH1Δ5_2406_rpoC(M466L)	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC rpoC*(M466L)	This study
E. coli	DH1Δ5_2406_pcnB(R149L)_ rpoC(M466L)	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC pcnB*(R19L) *rpoC*(M466L)	This study

Appendix 2.2: Plasmids

The pCRISPR-Gibson1 plasmids were constructed to clone constructs with specific guide sequence to target *E. coli* genome for introduction of point mutants. The parent plasmid, pCRISPR-Gibson1 (#2786), was generated from pCRISPR (Addgene 42875) to introduce cut sites between sgRNA promoter and the sgRNA to facilitate the use of Gibson assembly to introduce guide sequences for the target DNA. All guide sequences were generated using the Benchling CRISPR tool (*see Appendix 2.3* for guide sequences).

pCRISPR-PcnB2409 (#2784) was constructed by insertion of the annealed oligonucleotides, P1155 and P1156, and inserted into the XbaI-HindIII site of pCRISPR-Gibson1 using the Gibson protocol.

pCRISPR-RpoC2406 (#2794) was constructed by insertion of the annealed oligonucleotides, P1232 and P1233, and inserted into the XbaI-HindIII site of pCRISPR-Gibson1 using the Gibson protocol.

Appendix 2.3: Oligonucleotides

Oligos used for plasmids and strains construction and strain constructions. All guide sequences for CRISPR-Cas9 genome editing are highlighted in grey. Repair fragments that were used are listed in the bottom of this table. The "*" indicates the phosphorothioate bond modification.

Name	Sequence
P1151_pCRISPR_gib_guideF	ataccgctcgccgcagccgaacgccctaggtctagggcggcggatttgtc
P1141*_pCRISPR_gibson_2R	gctgttttgaatggtcccaaaacCCGCGGAAGCTTgttttagagctatgctgttttgaatggtc
P1141_pCRISPR_gibson_3F	gctgttttgaatggtcccaaaacCCGCGGAAGCTTgttttagagctatgctgttttgaatggtc
P1142pCRISPR_gibson_3R	attcaaaacagcatagctctaaaacTCTAGAgttttgggaccattcaaaacagc
P1138_pCRISPR_gibson_1F	atgctgttttgaatggtcccaaaacTCTAGAgttttagagctatgctgttttgaatggtc
P1152_pCRISPR_gib_guideR	gaggccctttcgtcttcacctcgagtccctatcagtgatagagattgacatcc
P1156_pCRISPR_2409_pcnB_R	aaacagcatagctctaaaacCTACGCTGTAATACAGGCTGgttttgggaccattcaaaac
P1155_pCRISPR_2409_pcnB_F	gttttgaatggtcccaaaacCAGCCTGTATTACAGCGTAGgttttagagctatgctgttt
P1233_g2rpoC_R	aaacagcatagctctaaaacCGGCGAACGGCGAACCAATCgttttgggaccattcaaaac
P1232_g2rpoC_F	gttttgaatggtcccaaaacGATTGGTTCGCCGTTCGCCGgttttagagctatgctgttt
P1227_2406_pcnB RF_R	A*CGTAATCACGGACGGTAAAATCtGCTACGCTGTAATACAGGCTGTTGATAGTGAAA TCGCGGaGCTGGGCGTCTTCTTCGATGGAGCCGAAAATGT*T
P1226_2406_pcnB RF_F	A*ACATTTTCGGCTCCATCGAAGAAGACGCCCAGCtCCGCGATTTCACTATCAACAGC CTGTATTACAGCGTAGCaGATTTTACCGTCCGTGATTACG*T
P1231_2406_rpoC_RF_R	T*CCTGAGACGGAACGATGATTGGTTCGCCGTTCGCCGGtGACAGGATGTTGTTGGT AGACATCATCAGCGCACGCGCTTCCAGCTGGGCTTCCAGCGTCAGCGGTACGTGAA CAGCCAgCTGGTCACCATCGAA*G
P1230_2406_rpoC_RF_F	C*TTCGATGGTGACCAGcTGGCTGTTCACGTACCGCTGACGCTGGAAGCCCAGCTG GAAGCGCGTGCGCTGATGATGTCTACCAACAACATCCTGTCaCCGGCGAACGGCGA ACCAATCATCGTTCCGTCTCAGG*A

Appendix 2.4: DNA probes for rRNA depletion for RNA-Seq library preparation

Name	Sequence
23S-3	CACTTATCTCTTCCGCATTTAGCTACCGGGCAGTGCCATTGGCATGACAACCCGAACACCAGTGATGCGTCCACTCCGGT
23S-4	CCTCTCGTACTAGGAGCAGCCCCCCTCAGTTCTCCAGCGCCCACGGCAGATAGGGACCGAACTGTCTCACGACGTTCTAA
23S-5	ACCCAGCTCGCGTACCACTTTAAATGGCGAACAGCCATACCCTTGGGACCTACTTCAGCCCCAGGATGTGATGAGCCGAC
23S-6	ATCGAGGTGCCAAACACCGCCGTCGATATGAACTCTTGGGCGGTATCAGCCTGTTATCCCCGGAGTACCTTTTATCCGTT
23S-7	GAGCGATGGCCCTTCCATTCAGAACCACCGGATCACTATGACCTGCTTTCGCACCTGCTCGCGCCGTCACGCTCGCAGTC
23S-8	AAGCTGGCTTATGCCATTGCACTAACCTCCTGATGTCCGACCAGGATTAGCCAACCTTCGTGCTCCTCCGTTACTCTTTA
23S-9	GGAGGAGACCGCCCCAGTCAAACTACCCACCAGACACTGTCCGCAACCCGGATTACGGGTCAACGTTAGAACATCAAACA
23S-10	TTAAAGGGTGGTATTTCAAGGTCGGCTCCATGCAGACTGGCGTCCACACTTCAAAGCCTCCCACCTATCCTACACATCAA
23S-11	GGCTCAATGTTCAGTGTCAAGCTATAGTAAAGGTTCACGGGGTCTTTCCGTCTTGCCGCGGGTACACTGCATCTTCACAG
23S-12	CGAGTTCAATTTCACTGAGTCTCGGGTGGAGACAGCCTGGCCATCATTACGCCATTCGTGCAGGTCGGAACTTACCCGAC
23S-13	AAGGAATTTCGCTACCTTAGGACCGTTATAGTTACGGCCGCCGTTTACCGGGGCTTCGATCAAGAGCTTCGCTTGCGCTA
23S-14	ACCCCATCAATTAACCTTCCGGCACCGGGCAGGCGTCACACCGTATACGTCCACTTTCGTGTTTGCACAGTGCTGTGTTT
23S-15	TTAATAAACAGTTGCAGCCAGCTGGTATCTTCGACTGATTTCAGCTCCACGAGCAAGTCGCTTCACCTACATATCAGCGT
23S-16	GCCTTCTCCCGAAGTTACGGCACCATTTTGCCTAGTTCCTTCACCCGAGTTCTCTCAAGCGCCTTGGTATTCTCTACCTG
23S-17	ACCACCTGTGTCGGTTTGGGGTACGATTTGATGTTACCTGATGCTTAGAGGCTTTTCCTGGAAGCAGGGCATTTGTTGCT
23S-18	TCAGCACCGTAGTGCCTCGTCATCACGCCTCAGCCTTGATTTTCCGGATTTGCCTGGAAAATCAGCCTACACGCTTAAAC
23S-19	CGGGACAACCGTCGCCCGGCCAACATAGCCTTCTCCGTCCCCCCTTCGCAGTAACACCAAGTACAGGAATATTAACCTGT
23S-20	TTCCCATCGACTACGCCTTTCGGCCTCGCCTTAGGGGTCGACTCACCCTGCCCCGATTAACGTTGGACAGGAACCCTTGG
23S-21	TCTTCCGGCGAGCGGGCTTTTCACCCGCTTTATCGTTACTTATGTCAGCATTCGCACTTCTGATACCTCCAGCATACCTC
23S-22	ACAGTACACCTTCACAGGCTTACAGAACGCTCCCCTACCCAACAACGCATAAGCGTCGCTGCCGCAGCTTCGGTGCATGG
23S-23	TTTAGCCCCGTTACATCTTCCGCGCAGGCCGACTCGACCAGTGAGCTATTACGCTTTCTTTAAATGATGGCTGCTTCTAA
23S-24	GCCAACATCCTGGCTGTCTGGGCCTTCCCACATCGTTTCCCACTTAACCATGACTTTGGGACCTTAGCTGGCGGTCTGGG
23S-25	TTGTTTCCCTCTTCACGACGGACGTTAGCACCCGCCGTGTGTCTCCCGTGATAACATTCTCCGGTATTCGCAGTTTGCAT
23S-26	CGGGTTGGTAAGTCGGGATGACCCCCTTGCCGAAACAGTGCTCTACCCCCGGAGATGAGTTCACGAGGCGCTACCTAAAT
23S-27	AGCTTTCGGGGAGAACCAGCTATCTCCCGGTTTGATTGGCCTTTCACCCCCAGCCACAAGTCATCCGCTAATTTTTCAAC
23S-28	ATTAGTCGGTTCGGTCCTCCAGTTAGTGTTACCCAACCTTCAACCTGCCCATGGCTAGATCACCGGGTTTCGGGTCTATA
23S-29	CCCTGCAACTTAACGCCCAGTTAAGACTCGGTTTCCCTTCGGCTCCCCTATTCGGTTAACCTTGCTACAGAATATAAGTC
23S-30	GCTGACCCATTATACAAAAGGTACGCAGTCACACGCCTAAGCATGCTCCCACTGCTTGTACGTACACGGTTTCAGGTTCT
23S-31	TTTTCACTCCCCTCGCCGGGGTTCTTTTCGCCTTTCCCTCACGGTACTGGTTCACTATCGGTCAGTCAGGAGTATTTAGC
23S-32	CTTGGAGGATGGTCCCCCCATATTCAGACAGGATACCACGTGTCCCGCCCTACTCATCGAGCTCACAGCATGTGCATTTT
23S-33	TGTGTACGGGGCTGTCACCCTGTATCGCGCGCCTTTCCAGACGCTTCCACTAACACACACACTGATTCAGGCTCTGGGCT
23S-34	CCTCCCCGTTCGCTCGCCGCTACTGGGGGAATCTCGGTTGATTTCTTTTCCTCGGGGTACTTAGATGTTTCAGTTCCCCC
23S-35	GGTTCGCCTCATTAACCTATGGATTCAGTTAATGATAGTGTGTCGAAACACACTGGGTTTCCCCATTCGGAAATCGCCGG
23S-36	TTATAACGGTTCATATCACCTTACCGACGCTTATCGCAGATTAGCACGTCCTTCATCGCCTCTGACTGCCAGGGCATCCA
23S-37	CCGTGTACGCTTAGTCGCTTAA
16S-1	TAAGGAGGTGATCCAACCGCAGGTTCCCCTACGGTTACCTTGTTACGACTTCACCCCAGTCATGAATCACAAAGTGGTAA
16S-2	GCGCCCTCCCGAAGGTTAAGCTACCTACTTCTTTTGCAACCCACTCCCATGGTGTGACGGGCGGTGTGTACAAGGCCCGG
16S-3	GAACGTATTCACCGTGGCATTCTGATCCACGATTACTAGCGATTCCGACTTCATGGAGTCGAGTTGCAGACTCCAATCCG
16S-4	GACTACGACGCACTTTATGAGGTCCGCTTGCTCTCGCGAGGTCGCTTCTCTTTGTATGCGCCATTGTAGCACGTGTGTAG
16S-5	CCCTGGTCGTAAGGGCCATGATGACTTGACGTCATCCCCACCTTCCTCCAGTTTATCACTGGCAGTCTCCTTTGAGTTCC
16S-6	CGGCCGGACCGCTGGCAACAAAAGATAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATTTCACAACACGAGCTGACG
16S-7	ACAGCCATGCAGCACCTGTCTCACAGTTCCCGAAGGCACCAATCCATCTCTGGAAAGTTCTGTGGATGTCAAGACCAGGT
16S-8	AAGGTTCTTCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCATTTGAGTTTTAAC
16S-9	CTTGCGGCCGTACTCCCCAGGCGGTCGACTTAACGCGTTAGCTCCGGTAGCCACGCCTCAAGGGCACAACCTCCAAGTCG
16S-10	ACATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGCACCTGAGCGTCAGTCTTCGT
16S-11	CCAGGGGGCCGCCTTCGCCACCGGTATTCCTCCAGATCTCTACGCATTTCACCGCTACACCTGGAATTCTACCCCCCTCT
16S-12	ACGAGACTCAAGCTTGCCAGTATCAGATGCAGTTCCCAGGTTGAGCCCGGGGATTTCACATCTGACTTAACAAACCGCCT
16S-13	GCGTGCGCTTTACGCCCAGTAATTCCGATTAACGCTTGCACCCTCCGTATTACCGCGGCTGCTGGCACGGAGTTAGCCGG
16S-14	TGCTTCTTCTGCGGGTAACGTCAATGAGCAAAGGTATTAACTTTACTCCCTTCCTCCCCGCTGAAAGTACTTTACAACCC
16S-15	GAAGGCCTTCTTCATACACGCGGCATGGCTGCATCAGGCTTGCGCCCATTGTGCAATATTCCCCACTGCTGCCTCCCGTA
16S-16	GGAGTCTGGACCGTGTCTCAGTTCCAGTGTGGCTGGTCATCCTCTCAGACCAGCTAGGGATCGTCGCCTAGGTGAGCCGT
16S-17	TACCCCACCTACTAGCTAATCCCATCTGGGCACATCCGATGGCAAGAGGCCCGAAGGTCCCCCTCTTTGGTCTTGCGACG
16S-18	TTATGCGGTATTAGCTACCGTTTCCAGTAGTTATCCCCCTCCATCAGGCAGTTTCCCAGACATTACTCACCCGTCCGCCA
16S-19	CTCGTCAGCAAAGAAGCAAGCTTCTTCCTGTTACCGTTCGACTTGCATGTGTTAGGCCTGCCGCCAGCGTTCAATCTGAG
16S-20	CCATGATCAAACTCTTCAATTTAAA
5S-1	ATGCCTGGCAGTTCCCTACTCTCGCATGGGGAGACCCCACACTACCATCGGCGCTACGGCGTTTCACTTCTGAGTTCGGC
5S-2	ATGGGGTCAGGTGGGACCACCGCGCTACGGCCGCCAGGCA
23S-1	AAGGTTAAGCCTCACGGTTCATTAGTACCGGTTAGCTCAACGCATCGCTGCGCTTACACACCCGGCCTATCAACGTCGTC
23S-2	GTCTTCAACGTTCCTTCAGGACTCTCAAGGAGTCAGGGAGAACTCATCTCGGGGCAAGTTTCGTGCTTAGATGCTTTCAG

Appendix 2.5: ALDH sequences

ALDH1

ATGAATAAAG ACACCCTGAT TCCAACTACC AAAGATCTGA AGCTGAAAAC TAATGTCGAA
AACATCAATT TGAAGAACTA CAAAGATAAC AGCTCGTGTT TTGGCGTGTT CGAAAACGTT
GAGAACGCGA TCAATTCCGC CGTTCACGCA CAGAAGATTC TGAGCCTGCA CTACACCAAA
GAGCAGCGTG AGAAGATCAT TACGGAAATC CGCAAAGCGG CGCTGGAGAA TAAAGAGGTG
CTGGCTACCA TGATTCTGGA AGAAACCCAC ATGGGTCGTT ATGAGGACAA AATCCTGAAG
CACGAGCTGG TCGCTAAGTA CACCCCTGGC ACCGAGGACC TGACGACCAC GGCATGGAGC
GGTGATAACG GTTTGACGGT CGTCGAAATG AGCCCGTATG GCGTCATTGG TGCAATTACC
CCTAGCACCA ATCCGACCGA AACTGTGATC TGCAACTCTA TCGGTATGAT TGCTGCGGGC
AATGCCGTCG TTTTCAATGG CCATCCGGGT GCGAAGAAGT GCGTTGCGTT CGCTATTGAA
ATGATCAACA AGGCGATCAT CTCCTGTGGT GGTCCGGAGA ACTTGGTGAC CACGATCAAA
AATCCGACGA TGGAGAGCCT GGATGCGATC ATTAAACATC CGCTGATTAA ACTGTTGTGC
GGTACGGGCG GTCCGGGTAT GGTTAAAACG CTGCTGAATA GCGGCAAAAA GGCAATCGGT
GCCGGTGCGG GCAACCCGCC AGTTATCGTA GACGACACGG CGGACATTGA AAAGGCCGGT
AAGAGCATTA TTGAGGGTTG TTCTTTTGAC AACAATCTGC CGTGTATTGC GGAGAAAGAA
GTGTTTGTTT TCGAGAATGT GGCGGATGAC CTGATTAGCA ACATGCTGAA AAACAATGCA
GTTATCATCA ACGAGGATCA AGTCTCCAAG CTGATCGATC TGGTGTTGCA GAAAAACAAC
GAAACCCAAG AGTACTTCAT TAACAAGAAG TGGGTTGGTA AGGATGCAAA GCTGTTTAGC
GACGAGATTG ACGTGGAAAG CCCGAGCAAT ATCAAATGCA TCGTGTGCGA GGTCAATGCA
AATCACCCGT TCGTTATGAC CGAACTGATG ATGCCGATCC TGCCGATTGT TCGCGTGAAA
GATATCGATG AGGCGGTCAA ATACACTAAG ATCGCGGAGC AGAATCGTAA ACATAGCGCG
TACATCTATA GCAAGAACAT CGACAACTTG AATCGTTTCG AACGTGAGAT CGACACCACG
ATTTTTGTGA AAAACGCAAA GAGCTTCGCC GGTGTGGGCT ATGAAGCCGA AGGCTTTACC
ACCTTTACCA TTGCGGGCAG CACGGGCGAG GGTATTACCT CTGCACGTAA TTTCACCCGT
CAACGCCGCT GCGTTCTGGC CGGTTAA

ALDH2

ATGAATGACA TCGAAATCGC CCAAGCCGTA AGCACTATTC TGAGCAAGTT CACTAAAGCA
ACGCCTGACG AGGCTCCGGC GACCTCGGAA GCCGCACGTG TCGATGGTCT GGATGAGATT
GTGGCAAAAG CCTTGGCCCA GCACAGCAGC GTGCGCGATG CTTCTGCGAT TAGCCAAGTT
GCGAAAGTTG CCAACGCTTC TACCGGTGCG TTCGATACGA TGGACGAGGC GATCTCCGCA
GCGGTTTTGG CACAGGTCCA ATATCGTCAT TGTTCTATGC AGGATCGCGC AAGCTTTATC
AATGGTATTC GCGACGTGTT CCTGCAAGAG GACGTGCTGT GTGCCCTGAG CCGCATGGCG
GTGGAAGAAA CCGGTATGGG TAACTACGAA GATAAGCTGA TCAAAAATCG CGTGGCCGCA
CTGAAAACGC CGGGTATTGA GGATCTGACG ACCAGCGCGG TTAGCGGCGA CGGTGGCCTG
ACGCTGATTG AATACAGCGC GTTCGGCGTC ATTGGCAGCA TCACCCCAAC CACGAACCCG
ACGGAAACGA TCATCAACAA TTCTATCGGC ATGCTGGCAG CGGGCAATAC CGTCGTCTTT
AGCCCGCACC CGCGTTCCCG CAAGGTTTCC CTGTACGCGG TGGAATTGAT CAACAATAAA
CTGGCGCAGC TGGGTGCACC GGCCAACATG GTAGTGACCG TGACCAAGCC GAGCATCGAC
AACACCAATG TTCTGATTAA TGATCCGCGT ATTAACATGC TGGTAGCAAC CGGCGGTCCG
GCGATTGTTA AGACCGTTAT GAGCAGCGGT AAAAAGGCGA TCGGTGCGGG TGCTGGTAAC
CCGCCTGCGG TTGTGGATGA AACGGCGGAC ATTGAGAAGG CTGCGCGTGA TATCATTAAA
GGTTGCAGCT TCGACAACAA TCTGCCATGT GTCGCAGAAA AAGAGGTCAT CGTTGTCAAT
CAGGTTGCTG ATTACCTGAT CCATTGCATG AAGAAAAGCG GTGCCTATCT GCTGTGCGAC
AAGAAACTGA GCCAGCAACT GCAGAGCCTG GTCTTGAACG AGAAGGGTAC TGGCCCGAAT
ACCGCGTTCG TGGGCAAAGA CGCACGTTAC ATCCTGCAGC AACTGGGCAT CCAGGTTGGC
GACGACATTA AGGTCATTTT GATCGAAGCG GAGAAAACCC ACCCGTTTGT TGTTCACGAG
CTGATGATGC CGGTCTTGCC GGTTGTGCGT GTGGACAATG TGGATGAGGC GATTGAGCTG
GCAGTGAAGG TGGAGCATGG TAACCGCCAC ACGGCGGTCA TGCACTCCAC CAACGTTGAG
AAGTTGACCA AGATGGCGCG TCTGATTCAA ACGACCATCT TTGTCAAAAA TGGTCCGTCG
TATGCGGGCC TGGGCGTTGG TGGTGAGGGT CATGCGACCT TTACCATTGC TGGCCCGACG
GGTGAAGGTC TGACCAGCGC CCGTAGCTTC GCACGTCGTC GTCGTTGCGT GATGGTCGAG
GCGCTGAACA TTCGCTAA

ALDH3

ATGATTAAGG ACACTCTCGT AAGCATCACC AAGGATCTGA AATTGAAAAC GAATGTAGAG
AACGCCAATC TGAAGAACTA CAAGGACGAT TCGAGCTGCT TCGGTGTTTT TGAAAATGTG
GAGAATGCTA TTAGCAATGC GGTGCATGCG CAGAAAATCC TGTCCCTGCA TTACACCAAA
GAGCAACGCG AAAAGATCAT CACTGAGATT CGTAAGGCCG CACTGGAGAA TAAAGAGATC
CTGGCGACCA TGATTCTGGA AGAAACCCAC ATGGGTCGTT ACGAGGATAA GATTCTGAAG
CACGAATTGG TTGCCAAGTA CACTCCGGGT ACCGAAGATC TGACCACCAC GGCGTGGAGC
GGTGATAACG GTCTGACCGT TGTCGAGATG AGCCCGTATG GTGTTATCGG TGCCATTACC
CCTTCTACGA ATCCGACGGA AACCGTGATC TGCAACAGCA TCGGCATGAT TGCGGCAGGC
AATACCGTGG TGTTCAATGG CCATCCGGGT GCCAAGAAGT GTGTCGCGTT TGCAGTTGAG
ATGATTAACA AAGCAATCAT TTCTTGTGGT GGCCCGGAAA ACCTGGTTAC CACCATCAAG
AACCCGACGA TGGACAGCTT GGACGCAATT ATCAAACACC CGTCCATTAA ACTGCTGTGC
GGTACGGGTG GCCCAGGCAT GGTCAAGACG TTGCTGAACA GCGGTAAAAA GGCGATTGGT
GCGGGTGCCG GCAATCCGCC GGTCATTGTG GACGACACGG CTGACATCGA GAAAGCGGGC
AAAAGCATCA TTGAAGGCTG CAGCTTCGAC AACAATCTGC CGTGCATCGC GGAGAAAGAG
GTTTTTGTTT TTGAGAACGT CGCAGACGAT CTGATTTCGA ACATGCTGAA GAATAATGCG
GTCATTATCA ATGAGGACCA GGTTAGCAAA TTGATCGATC TGGTCCTGCA GAAGAACAAC
GAGACTCAAG AATATAGCAT TAACAAAAAG TGGGTGGGTA AAGATGCGAA GCTGTTTCTG
GACGAGATTG ATGTGGAGTC TCCGAGCAGC GTTAAGTGTA TCATCTGCGA AGTGTCCGCT
CGCCACCCGT TCGTCATGAC CGAGCTGATG ATGCCGATCC TGCCAATTGT GCGTGTGAAA
GATATTGACG AAGCAATCGA GTACGCTAAA ATCGCAGAAC AAAATCGCAA ACACAGCGCA
TATATCTATA GCAAAAACAT CGACAACCTG AACCGTTTCG AACGCGAAAT TGATACCACC
ATTTTCGTCA AGAACGCTAA AAGCTTTGCG GGTGTTGGTT ACGAGGCCGA AGGCTTTACC
ACGTTCACCA TTGCGGGCAG CACGGGCGAG GGTATCACGT CCGCGCGTAA TTTCACCCGT
CAGCGTCGTT GTGTTCTGGC GGGTTAA

ALDH4

ATGTCATTTG ATATCAACAA TGCACAAGGC GTATTTGAAA CGGTAGAAGC AGCAATTGAA
GCCACCCACA AAGCCCAGGT GGAGTTCTAT GCGAACTCCA CTAAAGAGGG CCGTGAGGCG
ATCCTGACCG CTATCCGTGG CGCCGTGTTG GCGAAAGCGG AAGATTTCGC CAAAATGGTT
CGCGAAGAAA CCAAGCTGGG CCGTGTCGAG GATAAGATCG CGAAACATCA ACTGACCGCA
GCCAAGACCC CGGGTACCGA GGTCCTGGAA ACGAAGGTTT GGAGCGGTGA CAACGGTATC
AGCCTGGAAG AGCGTGCGCC GTACGGTGTC ATCGGCGCTG TCACCCCGGT TACGAATCCG
ACGGAAACGA TCGTCAACAA CGCAATTAGC ATGCTGGCGA GCGGCAACGC GGTGACGTTC
AATGTGCATC CATCCTCGAA AGTTGTGAGC GCAGTTATGA TCGACATGAT TAACAAAACG
ATTGTTGCTG CGGGTGGTCC GGCGAACCTG GTGACTATGG TTAAAGAACC AACGCTGGAA
ACGCTGAACG AAATCGCGAA AAGCCCGCTG GTGAATATGT TGGTCGGTAC GGGCGGTCCG
GGCCTGGTGA AGGCGATTCT GCAATCTGGC AAGAAAGGTG TCGGTGCGGG TGCGGGTAAT
CCGCCGGTGA TTGTCGATGC ATCTGCTAAT CTGGACCTGG CTGCAGCGGG TGTATACGGC
GGTGCCAGCT TCGACAATAA CCTGTTGTGT ATTGGCGAGA AAGAGGTGTT CGTTGAGGAT
AGCGTCGCGG ACGAGTTTCT GGCTAAGCTG GAAGCGACCG GTGCCTATGT TCTGAGCGCA
GAAGAGGCGG AGAAGTTGAC CGCTCAGATC CTGACGATGG ACGAGATCGA CGGTGCGAAA
CCGTGTACCG CACAGGAAAT TGCGCGTGTG TGGCACCCGG TCAAGCAGCA CGTTGGTCAA
GATGCGGGTG AGATCCTGAA GTCCATCGGT GTCGAGAGCG AAACCCGTCT GGCGGTGATG
GTTGTGGAGA ATGATCATCC TCTGGTTCAC GTCGAGCAGA TGATGCCGGT GCTGCCGGTT
GTGCGTTGCG CGAATATTGA CGAGGCGATC GAGCGCGCAG TTGCGGCCGA GCGTGGCAAC
AAGCACAGCG CGTGCATCTA CAGCGGCAAC ATTGAGAATG TTACCAAGTT CGGTCGTGCA
ATTAACACCA CCATCTTTGC CCACAACGGT CCGACCTTGA GCGGTGTCGG CTACAATGCA
GAAGGTACCA GCACCTTTAC CATTGCAGGC CCGACTGGTG AGGGTATTAC CAATGCGTAT
AGCTTCACCC GCGCACGTCG CTTTGCCATT GCCCAGGGCG GTCTGCGCAT TGTTTAA

ALDH5

ATGTCCGTAA ACGAGAAGAT GGTCCAAGAT ATTGTACAAG AAGTCGTAGC TAAAATGCAA
ATTAGCTCCG ACGTCAGCGG CAAGAAGGGC GTTTTTAGCG ATATGAATGA AGCAATCGAG

GCGAGCAAAA AGGCACAGAA AATCGTGGCT AAAATGAGCA TGGACCAACG CGAAGCCATT
ATCAGCAAGA TCCGTGAGAA GATTAAAGAG AATGCGGAAA TTCTGGCGCG TATGGGTGTT
GAAGAAACCG GCATGGGTAA TGTTGGCCAC AAAATTCTGA AGCATCAGCT GGTTGCGGAA
AAGACCCCGG GTACCGAGGA CATCACGACG ACGGCTTGGT CTGGTGATCG TGGTTTGACT
TTGATCGAAA TGGGCCCGTT CGGCGTTATC GGCGCGATCA CCCCGTGCAC TAACCCGTCT
GAAACCGTGC TGTGTAATAC GATCGGTATG CTGGCGGGTG GTAACACCGT TGTCTTTAAC
CCACATCCAG CCGCCATCAA GACCAGCATC TATGCGGTGA ATCTGCTGAA CGAGGCATCC
GTCGAGGTTG GTGGTCCGGA GAATATTGCG GTGACCGTCG AGCACCCGAC GATGGAAACC
TCGGATATCA TGATGAAGCA CAAGGACATC CATCTGATTG CGGCTACGGG CGGTCCGGGC
GTTGTGACCG CCGTCCTGAG CAGCGGTAAA CGCGGTATTG GTGCGGGTGC TGGCAACCCG
CCTGCGTTGG TCGACGAAAC GGCCGACATT CGCAAGGCCG CAGAGGATAT TGTGAACGGT
TGTACCTTCG ACAATAATCT GCCGTGCATT GCGGAGAAAG AAATTGTGGC AGTGGATTCG
ATCGCAGATG AGCTGTTGCA CTACATGGTG AGCGAGCAGG GCTGTTACAT GATCAGCAAA
GAAGAGCAGG ACGCGCTGAC CGAAGTTGTT CTGAAAGGCG GTCGTCTGAA TCGCAAATGC
GTGGGCCGTG ACGCGAAAAC GTTGCTGGGT ATGATTGGTA TCACGGTTCC GGACAATATT
CGTTGCATCA CGTTTGAGGG TCCGAAAGAG CATCCGCTGA TCGCGGAAGA ACTGATGATG
CCGATTCTGG GCGTGGTTCG TGCGAAAGAT TTTGATGATG CAGTGGAGCA GGCAGTGTGG
CTGGAGCACG GTAACCGCCA CAGCGCGCAC ATTCATAGCA AGAACGTTGA CAACATCACC
AAATACGCAA AAGCCATTGA CACCGCGATT CTGGTCAAGA ACGGTCCGAG CTATGCAGCA
CTGGGCTTCG GTGGTGAGGG CTATTGCACC TTCACCATCG CCAGCCGTAC CGGCGAGGGT
CTGACTAGCG CGAGCACGTT CACCAAGCGT CGCCGTTGTG TCATGACCGA TTCTCTGTGC
ATTCGTTAA

ALDH6

ATGAAAGAGG GTGTAATTCG CTTGGACATG GACATTAAGG TAATTGAACA GTTGGTAGAA
CAAGCGCTGA AAGAGATTAA GGCTGAGCAA CCTCTGAAAT TCACCGCTCC GAAACTGGAA
CGTTACGGCG TGTTCAAGAC GATGGACGAG GCGATCGCTG CGTCTGAAGA GGCACAGAAA
AAGCTGCTGT TCTCCAAAAT CAGCGATCGT CAGAAGTACG TTGATGTGAT TCGTAGCACC
ATCATTAAGC GCGAGAACCT GGAACTGATC AGCCGCCTGT CTGTTGAAGA GACTGAAATT
GGTGACTACG AACACAAATT GATCAAAAAT CGTCTGGCAG CGGAAAAGAC GCCAGGCACG
GAAGATCTGC TGACCGAGGC CATTACGGGT GATAACGGCT TGACCCTGGT TGAGTATTGC
CCGTTCGGTG TGATTGGTGC GATTACCCCG ACCACCAATC CAACCGAAAC GATCATCAAT
AACAGCATCA GCATGATTGC GGGTGGCAAC ACGGTCGTCT TTAGCCCGCA TCCGCGTGCA
AAGAAGGTGA GCCAGATGAC CGTCAAGATG CTGAACAAAG CACTGATTGA CAACGGCGCA
CCGCCGAATC TGATCACTAT GGTGGAAGAG CCGTCTATTG AGAACACGAA CAAAATGATC
GACAATCCGT CCGTTCGCCT GCTGGTTGCT ACCGGTGGCC CGAGCATCGT CAAGAAAGTC
CTGTCCAGCG GCAAGAAAGC CATCGGTGCC GGTGCGGGTA ATCCGCCAGT CGTTGTCGAC
GAGACTGCCG ACATTGATAA GGCGGCCAAA GATATTGTGG ATGGTTGTAG CTTTGACAAC
AATGTGCCGT GCATTGCAGA GAAAGAAGTC TTTGCGGTTG ACTCGATTTG CGACTACCTG
ATCCACCACA TGAAAGAGAA TGGCGCGTAT CAGATCACGG ACCCTATGTT GCTGGAGCAA
CTGGTTGCGC TGGTTACGAC CGAAAAGGGC GGTCCGAAAA CCAGCTTCGT GGGCAAGAGC
GCTCGTTATA TCCTGGATAA GCTGGGTATC ACGGTCGATG CGTCCGTCCG TGTGATTATC
ATGGAAGTGC CGAAGGATCA CCTGTTGGTG CAAGAAGAGA TGATGATGCC GATCCTGCCG
GTGGTCCGTG TTAGCGATGT GGATACCGCA ATCGAGTACG CACACCAGGC GGAGCATGGT
AATCGCCATA CCGCGATGAT GCACAGCAAA AACGTTGAGA AACTGAGCAA AATGGCCAAG
ATTATGGAAA CCACGATCTT TGTTAAGAAC GCGCCGAGCT ATGCGGGCAT TGGTGTTGGT
GGTGAGGGCT ACACCACCTT CACTATCGCA GGCCCGACCG GTGAGGGTCT GACCAGCCCG
CGTACCTTCT GTCGTAAGCG CAAATGTGTT ATGACGGACG CCTTTAGCAT TCGTTAA

ALDH7

ATGGAACGCA ACTTGTCGGT ACTCTCGCAA ACTAATGACT TGAAAATCAC TAAACGCACG
GAAGGTGATA AAAGCAATAA CAAAGAAAGC TATCTGGGTG TGTTTAAGAA GGTCGAAAAT
GCGATCACCA AAGCCATTTA CGCGCAGAAG AAACTGTCTC TGTATTACAC CAAAGAGGAC
CGCGAGCGTA TCATTAAGAG CATTCGTAAG GCCACCTTGG AAAACAAAGA GATCCTGGCC
AAGATGATCG TGGATGAAAC GCACATGGGC CGTTATGAGG ACAAGATCCT GAAGCACGAG
TTGGTGGCGA AATACACGCC TGGTACCGAG GACCTGATCA CGACCGCGTG GAGCGGCGAT

CAAGGTCTGA CGCTGGTCGA AATGAGCCCG TACGGCGTTA TTGGCGCGAT TACGCCGAGC
ACCAATCCTA CTGAAACCGT GATCTGCAAC AGCATTGGTA TGATTGCAGC TGGCGATTCC
GTCGTGTTTA ATGGTCATCC GGGTGCCAAG AAATGTGTTG CGTTTGCAGT CGACATGATT
AACAAAGCTG TTATCCGTGA GGGCGGTCCG GAGAACCTGG TGACCACGGT GGAGAACCCG
ACGATGGAGA GCCTGAATGT CATTATGAAG CACCCGTACA TCAAGCTGCT GTGTGGCACC
GGTGGTCCGG GTTTGATTAA GACCCTGCTG AACTCCGGTA AGAAAGCGAT TGGCGCAGGC
GCTGGTAATC CGCCGGTTAT TGTTGATGAT TCCGCCGACA TCGACAAAGC GGCAAAGAAC
ATCATTGAGG GTTGCAGCTT CGACAACAAT CTGCCGTGTA TCGCGGAAAA AGAGGTTTTT
GTGTTCGAGA ATGTCGCGAA TGATCTGATT CAGAACATGA TCAAGAATAA CGCAGTGCTG
ATTAATGAAA ACCAAGTCAG CAAACTGCTG GATCTGGTTC TGCTGGAGCG CAAGGATGAA
ACCCTGGAGT ATGCGATTAA CAAGAAATGG GTGGGTAAGG ATGCGAAACT GTTTCTGGAC
AAAATCGGCA TTAAGGCTAG CGATAACGTT CGTTGCATCA TCTGCGAAGT TGACGCGAAC
CACCCGTTCG TTATGACCGA ATTGATGATG CCGATTCTGC CAATTGTCCG TGTTAAGGAC
GTCGACGAGG CGATTGAATG TGCGAAAACC GCAGAGCAGC GTAAACGCCA TTCTGCATAT
ATGTACAGCA AGAATATTGA CAATCTGAAT CGTTTTGAAA AAGAGATCGA TACGACGATC
TTCGTGAAGA ATGCCAAAAG CTTCGCGGGT GTGGGTTTCG GTGCAGAAGG CTTTACGACC
TTCACCATCG CTGGCCCGAC CGGTGAGGGC ATCACCAGCG CACGTAACTT CACCCGTCAG
CGCCGTTGCG TTCTGGCCGG TTAA

ALDH8

ATGAATAACA ATCTGTTTGT AAGCCCTGAA ACGAAAGACT TGAAACTGCG CACTAATGTT
GAGAACTTGA AATTCAAAGG TTGTGAGGGT GGCTCCACCT ACATCGGCGT GTTTGAGAAT
GCAGAAACCG CGATCGACGA GGCGGTTAAC GCGCAAAAGC GTCTGAGCCT GTACTACACC
AAAGAACAGC GTGAGAAGAT TATCACGGAA ATTCGTAAAG TTACCCTGAA GAATAAAGAG
ATTCTGGCAC AAATGATTCT GGAAGAAACG CACATGGGTC GTTATGAAGA TAAGATCCTG
AAGCACGAGC TGGTCGCGAA GTATACGCCG GGTACCGAGG ACCTGGCAAC CACCGCGTGG
TCCGGTGACA ACGGCCTGAC TGTCGTGGAG ATGTCTCCGT ACGGTGTTAT TGGTGCGATC
ACCCCGTCGA CCAATCCGAC CGAAACGATC ATCTGCAACA GCATCGGTAT GATCGCGAGC
GGTAACGCGG TTGTTTTCAA CGGCCATCCA GGCGCAAAGA AATGTGTGGC GTTCGCGGTT
GATATGATTA ACCGCGCGAT TATCAGCTGC GGTGGTCCGC GCAATCTGGT GACCGCGATC
AAGAACCCGA CGATGGAGAG CTTGGATGCC ATCATCAAGC ACCCGGCGAT TAAGCTGTTG
TGCGGTACGG GCGGTCCGGG TATGGTTAAA ACCCTGCTGA GCAGCGGTAA GAAGAGCATT
GGTGCAGGCG CTGGTAATCC ACCGGTCATT GTGGATGACA CCGCCGACAT CGAGAAGGCT
GGTAAGAGCA TTATCGAGGG TTGTAGCTTC GACAATAATC TGCCGTGCAT CGCGGAAAAA
GAGGTGTTCG TTTTTGAAAA CGTCGCAGAC GACCTGATTA AGAATATGCT GAAAAACAAT
GCAGTCATTA TCAATAAAGA CCAAGTTAGC CGCCTGGTGA ATCTGGTCCT GCAGAAGAAC
AATGAAACCA GCGAATATAC CATCAACAAG AAATGGGTCG GCAAAGACGC AAAGCTGTTC
TTGGATGAGA TTGATGTCGA GTCTAGCTCC GATGTTCGCT GCATTATCTG CGAAGTGGAT
GCCGACCACC CGTTCGTCAT GACCGAACTG ATGATGCCGA TCCTGCCGAT TGTGCGTGTG
AAAGATATTG ATGAGGCCAT CAAATATGCC AAAATTGCCG AGCAGAACCG TAAACATAGC
GCGTACATCT ATAGCAAAAA CATTGAGAAC CTGAATCGTT TTGAAAAAGA GATTGATACG
ACCATTTTTG TGAAGAACGC GAAGTCGTTT GCAGGCGTCG GCTACGGCGC TGAGGGTTTC
ACGACTTTTA CCATTGCTGG CTGTACGGGC GAGGGCATCA CGAGCGCCCG TAACTTCACC
CGTCAGCGTC GCTGTGTGTT TGTTGGTTAA

ALDH9

ATGAATGACT TTAACATGAT CGATATCGAG AGCATTGTCA AAAACATTGT AAAAGAATTG
ACCGGTAACG AGAAGGGCCA GGGTGCGATC ACGACCGCGA CCGCTCCGAA AGAAGCCAAT
CCGCTGGTTG ACATTGAGAA AAAGATTATG GGTTTTATGA ATACCCCGAC CATGCCTGTG
GGTGAGTACG GCGTGTTCGA GGACATCAAC GACGCGATCG AACAAGCATG GCTGGCCGAG
CAGGAGTATC GTAAAGTTGG CCTGGATAAG CGTACGGAGA TTATCGAGGC TTTCAAGGCA
GAAGTGCGCA AAAATGTCGA AGAGATCTCC CGTCGTACCT TTGAAGAAAC GGGTATGGGC
CGTTATGAGG ATAAGATCCT GAAAAACAAC CTGGCCTTGG ATAAGACGCC GGGTGTGGAA
GATCTGGAAG CGGGTGTGAA AACGGGCGAT GGTGGTCTGA CCCTGTATGA GATGTCGCCG
TTCGGTGTCA TTGGTGCGAT CGCTCCGAGC ACCAATCCGA CGGAAACTAT TATCAATAAT
GGCATTAGCA TGCTGGCGGG TGGTAACACC GTCGTGTTCA GCCCGCATCC AGGTGCGAAA

GACGTCAGCG TGTTTATCGT TCAACTGATT AACAAAGCGA TCGAGCGTAT CAACGGTCCG
AAGAACCTGA TCGTTACGGT GAAGAACCCG AACATCGAAA GCACCAACAT TATGTTGGCG
CATCCGAAGG TGAATATGAT TTGCGCGACC GGCGGTCCGG GCATCGTTAA GGTTGCTCTG
AGCTCTGGCA AGAAGGCGAT TGGTGCCGGT GCGGGCAATC CGCCGGTGGT GGTGGACGAA
ACCGCAGACA TCGAGAAAGC GGCAGTTGAC ATTATCGACG GCTGTAGCTT CGACAATAAT
CTGCCGTGTA TCTGCGAGAA AGAGGTCATT GTTGTTGACA AGGTTGCGGA CTACCTGAAA
ACGTGTATGA GCAAGTATTG CGCACTGGAG ATTACGGACA AGAACATGTT GGCACAGCTG
GAGAAGCTGG TGCTGACCGA AAATGGCACG ATCAACAAAC AATTTGTCGG CAAGAACGCA
GATTACATTA TGAGCAAATT GGGTGTCAAT ATCGATCCGA GCATTCGCGT CATCTTTGCA
GAGGTGGAAG CGAATCACCC GTTCGCCGTC GAAGAGCTGA TGATGCCTAT TCTGCCGGTC
ATCCGTGTTC GCAACGTTGA TGAGGCCATC GATCTGGGTG TAGAGCTGGA ACATGGTAAT
CGTCACACCG CGATCATGCA CAGCAAACAC ATTGATAATC TGTCCAAGTT TGCCAAAGCG
GTTCAGACCA CGATTTTCGT CAAAAACGCG CCATCCTACG CAGGCATTGG TTACGGCGCA
GAAGGCCACG GTACCTTCAC CATTGCCGGT CCGACTGGTG AGGGCCTGAC CAGCGCTCGC
ACCTTCACTC GCAAACGTCG TTGCGTTATG GTTGACAACT TTTCTATTAA GTAA

ALDH10

ATGGAATTGG AAAGCAACGA ATTGAGCGTG ATTATTGAGA AGGTACTGAA AGAAATGAAC
AAGAAAGAGT TTGGTAAGAA AGAGAGCGAC GGTATTTTCG ATACGATGGA CGAGGCCGTT
GAGGCGTCTT ACGAGGCACA GAAGAAATAC AGCTCGTACT CCCTGGAGCA GCGCGAGAAG
CTGATTCAAG CAATGCGTAA AGCGATCATG GATAATGCGA TGGAAGTCGC TAATCTGTGT
GTGAAAGAAA GCGGTATGGG TCGTGTCGAC CACAAATACT TGAAACTGAA ATTGATTGTT
GAAAAGACGC AAGGTACGGA AATCCTGCGT CCGGAAGTTT ACACCGGTGA CAACGGCCTG
ACCCTGATTG AACATGGTGC TTTCGGTGTT ATCGGTGCCA TTACGCCGAG CACCAATCCG
GCAGCGACCG TCGCGTGCAA CTCCATCTGC ATGCTGGCGG GTGGTAATAC TGTGGTTTTT
AGCCCGCACC CAGGTGCGCT GAATAGCTGC TTGACCATGA TCCGCATTCT GAATAAAGCA
ATCAAAGAGG CCGGTGGTCC GGAGAACCTG ATTACCAGCG TGAAAGCACC TAGCATTGAG
AATACCAATA TCATGATTAA CCACAAGCGT ATTCGCCTGG TCGTGGCTAC CGGCGGTCCG
GGCATTGTGA AACTGGTGCT GTCCAGCGGC AAGAAGGCGA TCGGTGCGGG TGCCGGCAAT
CCGCCGGTTG TTGTGGATGA AACCGCCGAC ATTCCGAAGG CGGCACGTGA CATCATTGCC
GGCTGCAGCT TTGACAATAA TCTGCCGTGC ATCGCAGAGA AAGAAGCAAT TGTCGTCGAG
AGCGTTTACG AAGAATTGAT TAAAGAGTTC AAGAAAAACC GCGTCGTTTA CGAGCTGACG
GACGAAGAGG CCGAAAAACT GGTTGGCAAG GTCCTGAACT ATGATGAGAA GAACAAGAAG
TATAGCATCA ACAAAAAGTT CGTCGGTAAA GATGCGAAAT ATCTGCTGGA GAGCATCGGC
AAGGATGCGG GCACGGGTGT TGAGTGTCTG ATTTATCGTG CGGAGAATAG CCACCCGTTC
GTCCAAGAAG AGCTGATGAT GCCGATCCTG CCGATCGTCA AGGTTAAGAA CGTGGACGAA
GCGATCGAAA CCGCAGTGGA AGATGAGCAT GGCAATCGTC ATACGGCGAT GATGCACAGC
AAAAACGTTG TGAACCTGAC GAAGATGGCG CGTGCGATCG ATACCACTAT CTTCGTGAAA
AACGCACCGT CTTATGCGGG TATCGGCTTT GGTGGCGAGG GTCACACCAC CTTTACCATT
GCTGGCCCAA CCGGTGAGGG CATCACCAAC GCCGTTACCT TCACGCGCCA GCGTCGTTGT
ACGATGGTGG ACTCTTTTCG CATCGTGTAA

ALDH11

ATGGAGATCG GCGCAAAAGA AATTGAGTTA ATCGTAAGAG AAGTTTTGGC AGGCATTGAA
TCTCGTGGCC CGAAGCTGAG CTACATTCCG GCCCAAAGCG ACAACGGTGT TTTTGAGCGC
GTGGAAGATG CCATTGGTGC GGCGCATACC GCGCAACGCG AATGGGTCGA GCATTACCGT
GTTGAGGATC GCCGTCGCAT CATCGAGGCA ATCCGTATGA CGGCAAAGAG CCACGCGAAA
ACCTTGGCGA AGCTGGTGTG GGAAGAAACG GGCATGGGTC GCTTTGAGGA TAAGATTCAG
AAGCACATGG CAGTCATCGA GAAAACGCCA GGCGTTGAGT GCCTGACCAC GGACGCAATT
TCCGGCGACG AGGGTCTGAT GATCGAAGAG TACGCTCCGT TTGGTGTTAT TGGTGCGATC
ACCCCGTCCA CGAACCCAAC CGAAACCATC ATTAACAATA CTATCAGCAT GATTGCGGGT
GGCAATGCGG TGGTGTTCAA CGTTCACCCT GGTGCGAAGA AATGTTGCGC GCACTGTCTG
AAGCTGCTGC ATCAAGCTAT CGTCGAGAAC GGTGGCCCTG CCAACCTGAT TACCATGCAG
AAAGAGCCGA CTATGGAAGC TGTGACCAAG ATGACCTCTG ACCCGCGTAT CCGTCTGATG
GTCGGTACGG GTGGTATGCC GATGGTCAAT GCGTTGCTGC GTTCGGGCAA GAAAACGATC
GGTGCAGGCG CTGGTAATCC GCCGGTTATT GTGGATGATT CCGCGGACGT GAGCCTGGCA

GCGCGTGAGA TTTATCGCGG TGCCAGCTTC GATAACAATA TTCTGTGCCT GGCGGAAAAA
GAGGTTTTTG TGATGGAGAA AGCTGCGGAT GAACTGGTTA ACAACCTGGT GAAAGAAGGC
GCATATCTGC TGAATCCGAT GGAGCTGAAT GAGATTTTGA AATTCGCAAT GATCGAAAAG
AACGGCAGCT GCGAGGTCAA CAAGAAGTGG GTCGGCAAGG ACGCCGGTCT GTTTCTGGAA
GCCATTGGCG TCAGCGGCCA CAAAGACGTT CGTCTGCTGA TTTGTGAAAC CGACCGCAAT
CACCCGTTCG TCATGGTTGA GCAGCTGATG CCGATTCTGC CGATCGTCCG TCTGCGCACC
TTCGAAGAGT GCGTGGAGAG CGCGGTGGCA GCGGAAAGCG GCAATCGTCA CACGGCGAGC
ATGTTCAGCC GCAATGTGGA GAATATGACC CGTTTCGGTA AAGTTATCGA GACTACCATT
TTCACCAAAA ACGGTAGCAC GTTGAAAGGT GTTGGTATCG GTGGTGAGGG TCATACCACC
ATGACCATCG CGGGTCCGAC GGGTGAAGGT CTGACCTGTG CCCGTAGCTT TACGCGTCGT
CGTCGCTGCA TGCTGGCCGA GGGCGGTTTG CGTATCATTT AA

ALDH12

ATGGACGCAC AAAAGATTGA AAAACTGGTA CGCAAGATTT TGGAAGAGAT GGAAGAGAAA
AAGAAACCGG CCGAGACTGA GTGTGAATGG GGTATCTTTG ACCACATGAA CCAGGCGATT
GAAGCGGCGG AAATTGCGCA AAAAGAGCTG GTTCAACTGA GCCTGGGTCA GCGTGGCAAA
CTGATTGAAG CAATTCGTAA GGCTGCGAAA GAGAACGCGG AGAAGTTCGC GCGCATGGCA
GTCGATGAGA CTGGTATGGG CAAATACGAG GACAAAATCG TCAAAAATCT GCTGGCTGCC
GAAAAGACCC CGGGTATCGA AGATCTGCGC ACCGAGGTGT TTAGCGGTGA CGACGGCTTG
ACGTTGGTGG AGCTGAGCCC GTACGGCGTG ATCGGCGCTA TCACCCCGAC CACCAACCCG
ACCGAAACCA TCATTTGTAA TTCCATTGGT ATGATCGCGG CAGGCAACGC AGTCGTCTTT
TCCCCGCACC CGCGTGCGAA GAACACCTCT CTGTACGCAA TTAAGATTTT CAATCAGGCG
ATCGTTGAGG CGGGTGGTCC GAAGAACCTG ATTACCACGG TAGCAAACCC GAGCATTGAA
CAAGCCGAGA TCATGATGAA GCACAAAACG ATCAAAATGC TGGTTGCTAC CGGTGGTCCG
GGTGTGGTGA AGGCGGTTCT GAGCAGCGGT AAGAAGGCCA TCGGCGCTGG TGCGGGTAAT
CCGCCTGTGG TTGTTGACGA AACTGCGGAT ATTGAGAAGG CAGCCAAAGA CATCATCGCA
GGCTGCTCGT TCGATAACAA TTTGCCGTGC GTTGCCGAGA AAGAGGTGAT TGCAGTGGAA
AGCATCGCAG ATCGTCTGAT CGACTATATG AAAAAGCACG GTGCGTATGA GATTACCAAT
AAAGAGCAGA TCCAGCAACT GACCGATCTG GTTGTCGAGA ACGGCCATGC CAACAAAGAG
TTCGTCGGTA AAGACGCCGC GTACATCCTG AAGCATATCG GTATCAATGT TCCGCCGGAT
ACCCGTGTGG CCATTATGGA AGTGGATGGC AAACACCCAC TGGTTACGGT TGAGCTGATG
ATGCCGATCC TGCCAATTGT GCGTGTCAAA AATGTTGACC AGGCAATCGA ACTGGCGGTC
GAAGTTGAGC ACGGCTTCCG TCATACGGCG ATTATGCATA GCAAGAACGT TGATCACCTG
ACGAAATTCG CAAAGGCGAT CCAGACGACC ATTTTTGTGA AGAATGCTCC TAGCTATGCG
GGCATTGGTG TGGGCGGTGA AGGTTACGCT ACCTTTACCA TCGCGGGTCC GACGGGTGAG
GGCCTGACGA GCGCGAAGGA TTTCGCGCGT AAGCGCAAAT GCGTCCTGGT CGACGCCTTG
TCTATTCGCT AA

ALDH13

ATGAACAAGG ATACGACGAT TAGCGAAACC GAGAACTTGA AATTTAAAAC GAACATTAAG
AATGCTGACC TGAAGAATTA CGAGAATAGC ACGAGCTATT CCGGCGTTTT TGAAGATGTC
GAGGTGGCGA TCAACAAGGC CATCACCGCG CAGAAAGAGT TCAGCCTGTA CTATACGAAA
GAGCAGCGCG AGAAAATCCT GACTGAGATT CGTAAAGCGA CCCTGAAAAA CAAAAAGATT
CTGGCGAAGA TGATTCTGGA CGAAACCCAC ATGGGCCGCT ATGAGGATAA GATCTTGAAG
CATGAACTGG TTGCAAAATA CACCCCGGGT ATTGAGGATC TGACTACCAC CGCTTGGTCC
GGCGACAATG GCCTGACCGT TGTTGAAATG GCGCCGTACG GTGTGATTGG TGCAATTACG
CCTAGCACCA ACCCGACGGA AACCGTTATC TGCAATAGCA TCGGTATGAT CGCAGCGGGC
AATGCAGTGG TTTTCAATGG TCACCCGAGC GCAAAGAAGT GTGTGGCCTT TGCTGTCGAT
ATGATCAATA AAGCAATCGT CAGCTGTGGT GGCCCGAAAA ACCTGATTAC CGCGGTGAAA
AACCCGACGA TGGAGAGCTT GGATGCGATT ATCAAGCATC CGGAAATCAA ACTGCTGTGT
GGTACCGGTG GCCCAGGTAT GGTGAAAACC CTGTTGAACA GCGGCAAGAA AGCCATCGGT
GCGGGTGCCG GTAATCCGCC GGTGATTGTC GACGATACCG CGGATATCGA AAAGGCGGGT
AAAAACATCA TTGAGGGTTG CTCGTTCGAC AATAATCTGC CATGCATCGC CGAAAAAGAG
GTTTTTGTCT TTGACAACGT TGCCGACAAT CTGATTGATA ACATGTTGAA GAATAACGCT
GTGATCATCA ATAAGGACAA AATCACCAAG CTGCTGAATC TGATCCTGCA GAAAAACAAT
GAAACGCAGG AGTATAACAT TAACAAGAAG TGGGTCGGCA AAGACGCGAA GCTGTTCCTG

AATGAGATTG ACGTTGAGGC GCCGAGCAGC GTTCGTTGCA TTATCTGTGA GGTGGAACCG
GATCACCCGT TCGTGATGAC CGAGCTGATG ATGCCGATCC TGCCGATTGT CCGTGTTAAG
AACATTGACG ACGCGATCCA ATACGCAAAG ATCGCGGAAC AATCTCGCAA ACACAGCGCG
TACATTTACT CCAAAAACAT CGATAATCTG AATCGTTTTG AAAAAGAGAT TGACACCACG
ATTTTCGTCA AGAACGCAAA GTCTTTCGCG GGTGTGGGCT ATAACGCAGA AGGTTTCACG
ACCTTCACTA TTGCGGGTTG CACGGGCGAG GGTATTACCA GCGCTCGTAA CTTTACCCGT
CAACGCCGTT GCGTCCTGGC CGGCTAA

ALDH14

ATGGAATTTG AGGTAAACAA CATTGAAGAA ATTGTGGAAC TGATTATGAA GAAGATGGCA
GAGTCTAACA TCAGCACGGC GGGTAATTCC AAAAATGGTG TGTTCGACAA TGTGGACGAG
GCGATTGAAG AAGCGAAGAA AGCGCAGGCA ATTCTGTTCA GCAGCAAGTT GGAGCTGCGT
GAGAAGATCA TCGCTAGCAT TCGCGACACC CTGAAGAATC ACGTTACCGA GCTGGCAGAG
TTGGCAGTTA AAGAAACCGG TATGGGTCGT GTCGCGGACA AAGAGTTGAA AAACAAAATC
GCTATTGAAA AGACCCCGGG TTTGGAAGAT CTGAAGGCAT TCGCATTCAG CGGTGATGAT
GGCCTGACGG TTATGGAACT GTCCCCGTAT GGTGTGATTG GCGCAATTAC GCCGAGCACC
AACCCGAGCG AAACGGTGAT CTGTAACAGC ATCGGCATGA TCGCCGCTGG TAATGCGGTG
ATTTTCGCAC CGCATCCGGG TGCCAAGCGC ACCAGCATCC GCACCGTCGA GCTGATCAAT
GAGGCGATCC GTAAGGTTGG TGGCCCTGAT AATCTGGTTG TTACCATCCG TGAGCCTAGC
ATTGAGAATA CCGAGAAAAT CATTGCCAAT CCAAATATCA AAATGCTGGT TGCTACCGGC
GGTCCGGGCG TTGTCAAAAC CGTTATGAGC AGCGGTAAGA AGGCGATTGG TGCCGGTGCG
GGCAATCCAC CGGTCCTGGT CGATGAAACC GCGGACATCG AGAAAGCCGC GAAAGACATT
ATTGCGGGCT GTAGCTTTGA CAACAATCTG CCGTGCACTG CCGAGAAAGA GGTCGTTGCA
GTTGATTCTA TCGTGAACTA CCTGATCTTT GAGATGCAAA AGAACGGCGC GTATCTGCTG
AAGGACAAAG AACTGATTGA AAAGCTGCTG AGCCTGGTGC TGAAGAACAA CAGCCCGGAT
CGTAAGTACG TCGGTCGTGA CGCCAAGTAT TTGCTGAAAC AGATCGGTAT CGAGGTGGGT
GATGAAATCA AGGTCATTAT CGTCGAAACG GACAAGAACC ACCCGTTCGC TGTGGAAGAG
TTGCTGATGC CGATTCTGCC GATCGTCAAA GTTAAAGACG CCCTGGAAGG TATCAAAGTC
GCGAAAGAGC TGGAGCGTGG CCTGCGTCAT ACTGCGGTGA TCCACTCTAA GAATATTGAT
ATTCTGACCA AATACGCGCG TGAGATGGAA ACGACGATCC TGGTGAAAAA CGGTCCGAGC
TACGCGGGTA TTGGTATCGG CGGTGAGGGC CACGTTACGT TTACCATTGC AGGCCCGACG
GGCGAGGGTC TGACCTCGGC GAAATCCTTC GCGCGCAACC GCCGTTGCGT ATTGGTGGGC
GGTTTTAGCA TTAAATAA

ALDH15

ATGAATTTGG AAGCAAACAA CATGGACGAA ATTGTGGCAC TGATTATGAA AGAACTGAAG
AAAACCGACA TTAAGGCGGG TTGTCAATCT TGTGAGAGCT TGAAAAACGG CGTTTTCAGC
AGCATGGATG AGGCCATTGC TGCAGCGAAG AAGGCGCAGG AGATCCTGTT CAGCTCCCGT
CTGGAGATGC GTGAGAAGAT TGTCGCGAGC ATTCGCGAAG TGATGAAGGA CTATGTTGTG
GAGCTGGCCG AGCTGGGTGT GAAAGAAACC GGTATGGGTC GTGCCGCAGA CAAAGCGCTG
AAACACCAGG TGACGATCGA GAAAACCCCG GGTGTTGAGG ACTTGCGCGC CTTTGCGTTT
AGCGGCGATG ATGGTCTGAC CGTCATGGAG CTGAGCCCGT ATGGCGTGAT TGGCGCGATC
ACCCCAAGCA CCAATCCGTC CGAAACGATC ATCTGCAATA GCATTGGCAT GATCTCCGCT
GGCAATTCTG TTGTTTTCGC GCCACATCCG GGTGCGAAAC GCACGTCGAT TAAGACTGTC
GAAATCATTA ACGAGGCCGT TCGCCGTGCA GGCGGTCCGG AGAACCTGGT GGTCACGATC
GCGGAGCCGA GCATCGAAAA CACCAATCGT ATGATGGAGA ATCCGGATAT CAAGATGCTG
GTCGCCACGG GTGGTCCGGG TGTGGTTAAA AGCGTCATGA GCAGCGGTAA GAAAGCGATT
GGCGCAGGCG CAGGCAATCC GCCGGTGCTG GTTGATGAAA CCGCTGATAT CGAGAAGGCG
GCACGTGACA TCGTCGCCGG CTGTAGCTTT GACAATAATC TGCCGTGCAT TGCTGAGAAA
GAAGTCGTTG CGGTTGATTC TATCACCGAC TACCTGATTT TTGAGATGCA AAAGAACGGC
GCGTATCTGA TTAAAGACAA ATCCGTGATT GACCGCCTGG TGGCGATGGT TCTGAAGAAC
GGTAGCCCGA ACCGCGCGTA CGTTGGCAAA GATGCGAGCT ACATCCTGAA AGACCTGGGT
ATTAACGTTG GCGACGAGAT TCGTGTGATC ATCACCGAAA CCGACAAGGA TCACCCGTTT
GCAGTTGAAG AGCTGCTGAT GCCTATCCTG CCGATCATCC GTGTCAAGAA CGCGCTGGAA
GGTATTGAGG TAAGCAAGAA ATTGGAACAC GGTCTGCGCC ATACCGCGAT GATTCATAGC
AAAAACATTG ATATCTTGAC GAAGTACGCG CGTGATATGG AAACGACCAT CCTGGTCAAG

AATGGCCCGA GCTTCGCAGG CATCGGTGTG GGTGGTGAGG GTCACACGAC TTTCACCATT
GCCGGTCCTA CGGGTGAAGG TCTGACCAGC GCAAAGTCTT TCGCTCGTAA TCGTCGTTGC
GTGTTGGTCG GTGGTCTGAG CATTAAATAA

ALDH46

ATGAATAAAG ACACCCTGAT TCCGACCACG AAAGATCTGA AAGTTAAGAC TAACGGCGAG
AACATTAACC TGAAGAATTA CAAAGACAAT AGCAGCTGTT TTGGCGTCTT TGAAAATGTG
GAGAATGCGA TTTCTTCTGC GGTGCACGCG CAAAAGATTC TGTCCCTGCA CTATACGAAG
GAGCAGCGCG AGAAAATCAT TACTGAAATC CGTAAAGCGG CCCTGCAGAA TAAAGAGGTG
CTGGCAACCA TGATTTTGGA AGAAACGCAC ATGGGTCGCT ACGAAGATAA GATTCTGAAA
CATGAGCTGG TCGCGAAATA CACCCCGGGT ACCGAGGACT TGACCACTAC CGCGTGGAGC
GGCGACAACG GTCTGACCGT CGTCGAGATG AGCCCGTACG GTGTCATTGG TGCAATCACG
CCGAGCACCA ACCCGACGGA AACGGTGATC TGCAACAGCA TTGGTATGAT CGCTGCAGGC
AACGCGGTCG TTTTCAATGG CCACCCGTGT GCGAAGAAGT GTGTTGCCTT TGCTGTTGAG
ATGATCAACA AAGCGATTAT CAGCTGTGGC GGTCCGGAGA ATCTGGTCAC GACCATTAAG
AATCCGACCA TGGAATCCCT GGACGCAATC ATTAAGCACC CGTCGATTAA ACTGCTGTGC
GGCACCGGTG GTCCAGGTAT GGTTAAGACG CTGCTGAACA GCGGTAAGAA AGCAATCGGT
GCTGGCGCTG GTAACCCGCC TGTCATCGTT GACGATACGG CAGACATTGA AAAGGCGGGT
CGTTCCATCA TTGAGGGCTG CAGCTTCGAT AACAACCTGC CGTGCATTGC GGAGAAAGAG
GTTTTCGTGT TTGAGAATGT GGCAGACGAT CTGATCAGCA ACATGCTGAA GAATAACGCG
GTAATCATTA ACGAGGACCA AGTTAGCAAG CTGATCGACC TGGTTTTGCA GAAAAACAAC
GAAACCCAAG AGTACTTCAT CAATAAGAAA TGGGTGGGTA AGGATGCGAA GTTGTTCCTG
GATGAGATCG ATGTGGAAAG CCCAAGCAAT GTGAAATGCA TCATCTGCGA AGTTAATGCC
AATCATCCGT TCGTTATGAC CGAACTGATG ATGCCGATCT TGCCGATCGT GCGTGTCAAA
GATATCGATG AGGCCATTAA GTATGCGAAG ATCGCCGAAC AGAATCGTAA ACATAGCGCT
TATATCTACA GCAAAAACAT TGACAATCTG AATCGCTTCG AACGTGAGAT TGACACCACG
ATTTTTGTGA AAAACGCAAA AAGCTTTGCG GGTGTGGGCT ATGAGGCGGA AGGCTTCACC
ACCTTTACCA TTGCAGGTTC TACCGGTGAA GGTATCACGA GCGCCCGTAA CTTCACGCGC
CAACGTCGTT GTGTTCTGGC CGGCTAA

Appendix 2.6: gBlock sequences for ADHs

ADH1 G1 (Accession No. B6YQP9_AZOPC)

CGAGCGCCCG TAACTTCACG CGCCAACGTC GTTGTGTTCT GGCCGGCTAA TCTAGACTCG
CCATATCCGA CCCACCCAAG GACAACTCAT ATGAACAACT TCCGTTTCTG CAGCCCTACC
GAATTCATTT TTGGTAAAAA CACCATCTGT AAAGTGGCTC AGCTGGTTAA ACAGTATGGT
GGCTCTAAAG TTCTGATCCA TTACGGCAAT AAATCTGCGA AAAAATCTGG TCTGCTGACC
CAGATCGAGA ACTGCTTCCA GAACGAATTT ATCGAATATG TCAAACTGGG TGGTGTTCAG
CCGAACCCGA TCGACGAACT GGTCTACAAG GGTATCGAAC TGGGCCGTAA AGAAAAAGTT
AACTTCATCC TGGCTATCGG TGGCGGTAGC GTTATCGACT CTGCTAAAGC AATCGCTGCG
GGCATTCTGT ACAACGGTGA TTTCTGGAAC TTTTTCGAAG GCATCGTTAC CATTAACCAC
GCCCTGCCAA TTGCAACTGT TCTGACCCTG CCTGCTGCGG GCTCTGAGGG TTCTCCGAAC
ACTGTCATCA CGAAAACCGA CGGTATGCTG AAACGTGGCA TCGGTTCTTC CTTCATCCGC
CCAGTCTTCT CTATCATGGA TCCAGTGCTG ACGTTCACCC TGCCGACCTG TCAGACCGTT
TATGGCATCG CAGATATGAT GGCCCACGTT ATGGA

ADH1 G2 (Accession No. B6YQP9_AZOPC)

ACCTGTCAGA CCGTTTATGG CATCGCAGAT ATGATGGCCC ACGTTATGGA ACGCTACTTC
ACCCAGACCC AGGGTGTGGA TATTACTGAC CGCATGTGCG AGTCTATCCT GCTGTCTATT
ATCCACAGCG CGAAAACTCT GATTCGCGAA CCGGAAAACT ACGACGCTCG TGCCAACATC
ATGTGGGCCT CCACGATCGC GCACAACGGT ATCTGCGGCG TGGGTCGTGA AGAAGACTGG
GCGACCCATG CTCTGGAACA TGAACTGTCC GCGCTGTATA ACATCGCACA CGGCGCCGGC
CTGGCTGTGA TGTTTCCGGC GTGGATGCAA TACGTATACA CCGCGGGTAT CGACCGTTTC
GTGCAATTTG CTACCCGCGT TTGGAACATC GAAAACATCG GCTCTAAAAA AGAGATTGCC
CTGAAAGGTA TCCACGCTCT GAAAGACTTT TTCTCCTCCA TCAAACTGCC AATCAACTTT
GAACAGCTGG GCGCACAGAA AAGCGATATT GACAAACTGA TTGACACCCT GAAAATTAAC
ACCAAAGGTA AACTGGGTAA CTTCCTGCTG CTGGACATGA ACGATGCTCG TGCAATCTAC
GAAATTGCTG CTAAGCGTTA AACTAGTATC GATGATAAGC TGTCAAACAT GAGCAGATCT
GAGCCCGCCT AATGAGC

ADH2 G1 (Accession No. A0RQF7_CAMFF)

CGAGCGCCCG TAACTTCACG CGCCAACGTC GTTGTGTTCT GGCCGGCTAA TCTAGAAGAT
TAAACTCTAA GCGAGGAATA CATGGTCAAC TTTTCCTACT GCAATCCAAC CCGTATCGAA
TTCGGCAAAG GTAAAGAAAA CTCCATCGGT GAATACCTGA ACGAATATGG CGCAAAAAAC
GTGCTGATTC TGTTCGGCTC CGACCGCGTT AAAAAAGACG GTCTGTTTGA CAAAGCGACT
GCGTCCCTGA CCAAATTCGG CATCAAATTC TCCGAACTGG GTGACATTGT GAGCAATCCA
GTACTGTCCA AAGTTTATGA AGCTATCAAC CTGGCCCGCA AAAACGGCGT GGATAGCGTT
CTGGCGATCG GCGGTGGTTC TGTCCTGGAT ACTGCCAAAT CCGTAGCAGC CGGTGCAAAA
TACGACGGTG ACGTTTGGGA TCTGTTCCTG GCCAAAGCTC CGATTAAAGA TGCTCTGATG
GTTTTCGATA TTATGACCCT GGCTGCAACT GGTAGCGAAA TGAACAGCTT CGCCGTTGTC
ACCAACGAAG ACACTAAAGA GAAAATCTCT ATCACCTCTT CCCTGGTGAA CCCAAAAGTA
AGCGTAATCA ATCCGGAACT GATGAAATCC ATTTCTAAAA ACTACCTGGT GTACTCCGCG
GCCGACATCA TCGCGCATTC TATCGAAGGC TACCTGACCG CAACTCATCA CCCGGAAATT
ATCTCCAAAC TGGTTGAAGC GAATATCTCC

ADH2 G2 (Accession No. A0RQF7_CAMFF)

CAACTCATCA CCCGGAAATT ATCTCCAAAC TGGTTGAAGC GAATATCTCC ACTATTATTA
AAACGACCGA AATCCTGCTG GCTGACCCAG ACAACTACGA CGCACGTGCG GAATTTGCGT
GGGCAGCAAC TTGTGCTCTG AACGGCACCA CTTACGTTGG CGTTGGTGGT TACTCCTACC
CGAACCACAT GATCGAACAT TCCATCTCTG CACTGTACGG TGTACCGCAT GGTGCGGGTC
TGTCCGTAGT AATGCCGGCA TGGATGAAAT GGTATAAGGA CAAAAATGAA GCCCAGTTCT
CTCGCTTCGC TAAAGTAATC TTCGGTAAAA ACAGCGCTGA TGAAGGTATT GAAGCCCTGA
AGACGTGGTT CAAAAAAATC GGCACCCCGA CCAAACTGCG CGACTTCGGC CTGGACATGT

CCGTATCTGA CATCACCACT GCTGCGCTGC ATCACGCTAA AGCATTTGGT ATCGCTGATA
TCTATACCAA AGACGTTCTG GAAGAAATTC TGAACCTGGC TTACTAAACT AGTATCGATG
ATAAGCTGTC AAACATGAGC AGATCTGAGC CCGCCTAATG AGC

ADH3 G1 (Accession No. G5F136_9ACTN)

CGAGCGCCCG TAACTTCACG CGCCAACGTC GTTGTGTTCT GGCCGGCTAA TCTAGAGATA
CCTCTCCCTT AAGAGCGAGG TCATTATGAT TAACTTCGAC TATTGCGTGC CGACTAAAGT
TGTTTTCGGT CATGGTGTTG AATCTAACGT TGGCAAATAC GTAAAAGAGT TCGGTGGTAC
CAAAGCGATG ATTCACTGGG GCGGTGACTA TGTTCGCGAT ACGGGTCTGC TGGACCGTGT
CGAAAAATCT CTGTCCGCGG AAGGTATCGG CTACGTTGAG TTTGAAGGCG TCGTACCGAA
CCCGCGCCTG TCCACCGCTA AAGAGGGCCT GGCTCTGGCG AAACGTGAAG GTGTAGATTT
CCTGCTGGCT ATCGGCGGCG GTTCTGCAAT CGATAGCAGC AAAACCATCG CATACGGTCT
GGCGAACGAT TTCGAGCTGG AAGACCTGTT CCTGGGTAAA GTAAGCACTG ACCGTATCGC
GGGCCTGGGT GCGATCTCTA CCCTGGCCGG CACCGGTTCT GAAACCTCTA ACTCTACTGT
TATCAACATC GATACGATGG GTGACGTCGA GCTGAAACGT AGCTACAACC ACGAATGTGC
CCGTCCGAAA TTCGCGATCA TGGATCCGGA ACTGACCTAT ACCGTTCCGG CATGGCAGAC
GGCCGCCGCT GGCTGCGACA TTATGATGCA CACTA

ADH3 G2 (Accession No. G5F136_9ACTN)

TTCCGGCATG GCAGACGGCC GCCGCTGGCT GCGACATTAT GATGCACACT ATGGAACGTT
TCTTCACTAC CGTTTCTCAT ACGGAACTGA TCGATCAAAT GTCCCTGGGT CTGCTGCGTG
CTGTCAAAAC CGCGATTCCA CTGGCTCTGG CTGAGCCGGA TGACTATGAT GCACGCGCCA
CCCTGCTGTG GGCGGGCTCT CTGTCTCACA ACGGTCTGAC CGGCACCGGT CAGCAGGGTG
ACTTCGCATC CCATGCAATT GAACACGAAA TGGGTGCTCT GTACAACTGC ACCCACGGCG
CAGGTCTGTG CGCGATGTGG TCTTCCTGGG CTCGTTATGT CATTGATGTG CGTCCGGAAC
GTTTCGCACA GTTCGGTGTG GAAGTCTTCG GTGTGGTAAA CGACTACTCT GATCCGAAAG
GTACCGGTCT GCGCGGTATC GAGGCTTGGG AAAAATTCTG CAAATCTGTG GGTATGCCGG
TACGTATGAG CGAACTGGCA ATCAACCCGA CTGATGAGGA GATCCGTCAT ATGGCTCAGG
GCGCCATTGA CGCCCGTGGT GGTGATCATT GCGGTTCTTT CATGGAACTG CGTGTTGATG
ACGTCGTAAA AATTCTGGAA ATGGCCCGCT AAACTAGTAT CGATGATAAG CTGTCAAACA
TGAGCAGATC TGAGCCCGCC TAATGAGC

ADH4 G1 (Accession No. B1C7G7_9FIRM)

CGAGCGCCCG TAACTTCACG CGCCAACGTC GTTGTGTTCT GGCCGGCTAA TCTAGACCCA
CCTTCCAAAA CTCCCAGAGG TATTCATGCA GAAATTTGAC TACTATACTC CGACCAAAGT
TATCTTTGGC AAAGGCACCG AAAACAAAGT GGGTAAAGAG ATGAAAAAAG ACGGTGCTAA
GAAGGCTTAT ATCGTTTACG GCGGCAAATC CGCGAAAAAA AGCGGTCTGC TGGACAAAGT
GGAGAAATCT CTGAAAGACG AAAACATTGA ATACAAAATG ATCGGTGGCG TGAAACCGAA
CCCTCGCCTG TCTCTGGCTC GCGAAGGTGT GAAGGAAGCG AAGGAATTCG GTGCCGATTT
TATTCTGGCG GTTGGTGGTG GCTCTGTTAT CGATACCGCA AAAGGCATCG CACATGGCGT
AGCAAACCCT GACACTGACA TCTGGGATTT CTGGGAAGGT AAAGCCAAGG TTGAAAAATC
CCTGCCTGTT GGCGTTATCC TGACCATTTC TGCTGCGGGT TCTGAAATGA GCAACTCCGC
GGTGCTGACG AATGAAGAAA CTGGCATGAA GCGTGGCCTG TCCACCGATT TCAACCGTCC
GAAATTCGCC ATCATGGACC CGGAACTGAC CTACACGCTG CCGGATTACC AGGTTGGTTG
CGGTGTGGTA GACATCATGA TGCACACCAT GGATC

ADH4 G2 (Accession No. B1C7G7_9FIRM)

ATTACCAGGT TGGTTGCGGT GTGGTAGACA TCATGATGCA CACCATGGAT CGTTATTTCA
CTGACCTGAC TGATTGCCAG AACGATCTGA CCGATGAAAT CGCAGAGTCT CTGCTGCGTA
TCGTTATCAA AAACGGTCGT GTAGCTTGCA AGAATAAAGA AGACTACCAC GCTATGAGCG
AAATCATGTG GGCAGGTTCC CTGTCCCATA ACGGCCTGAC CGGTCTGGGC GCCCCGATGG
ACTTTGCAAC GCACCGCCTG GGTCACTCTC TGTCCGCGAA ATTTGATGTT GCACACGGTG
CGTCCCTGTC CGCCATGTGG CCGCACTGGG CTAACTACGT AAAACATAAA GACATCGAGC
GTTTTGCACG CTATGCGCGT AACGTTTGGG GCATTACGGA AGGCACCGAT GAAGAACTGG
CTGATAAAGG TATTGAAGCG ACCGTGGAAT TCTTCAAATC TATCAACATG CCGACCTGCT

TTAGCGAACT GGGTATCGGC ATCCAGGATG AGGATGGCCT GCGTGAGCTG ACCAACCGTT
GCTTCTACGT GAAAGGTACC AAAGTAGGTA AACTGATTCC GCTGACCGAA GAAGATATTT
ACCCGATCTA TGTATCTGCG AACAAATAAA CTAGTATCGA TGATAAGCTG TCAAACATGA
GCAGATCTGA GCCCGCCTAA TGAGC

ADH5 G1 (Accession No. YUGK_BACSU)

CGAGCGCCCG TAACTTCACG CGCCAACGTC GTTGTGTTCT GGCCGGCTAA TCTAGATAAC
ACACCTATCA AGAAATAATT CAGAGGTCCC AATGGAAAAC TTCACCTACT ACAACCCGAC
CAAACTGATC TTCGGCAAAG GCCAGCTGGA ACAGCTGCGC AAAGAATTTA AACGTTATGG
TAAAAACGTT CTGCTGGTTT ATGGTGGCGG CTCCATCAAA CGCAACGGTC TGTACGACCA
GGTCACCGGC ATCCTGAAAG AGGAGGGCGC GGTGGTTCAC GAACTGAGCG GTGTTGAACC
GAACCCGCGC CTGGCTACCG TGGAAAAGGG CATTGGTCTG TGCCGTGAAC ACGATATCGA
TTTTCTGCTG GCCGTCGGTG GTGGCTCTGT CATTGACTGC ACCAAAGCAA TCGCGGCGGG
TGTAAAATAC GATGGTGACG CTTGGGATAT CTTTTCCAAA AAGGTTACCG CCGAAGACGC
TCTGCCGTTT GGCACCGTAC TGACCCTGGC CGCTACCGGT TCCGAGATGA ACCCGGATTC
CGTTATCACC AACTGGGAAA CTAACGAAAA ATTCGTCTGG GGTTCCAACG TTACCCACCC
GCGCTTCTCT ATCCTGGACC CGGAAAACAC CTTTACCGTA CCGGAAAACC AGACAGTGTA
TGGCATGGTT GACAT

ADH5 G2 (Accession No. YUGK_BACSU)

AAACACCTTT ACCGTACCGG AAAACCAGAC AGTGTATGGC ATGGTTGACA TGATGTCTCA
CGTTTTCGAA CAGTATTTCC ATAACGTAGA AAACACTCCG CTGCAGGATC GTATGTGCTT
TGCTGTGCTG CAGACCGTCA TCGAAACGGC TCCGAAGCTG CTGGAAGACC TGGAAAATTA
CGAACTGCGT GAAACCATTC TGTACGCGGG TACCATTGCG CTGAACGGTA CTCTGCAGAT
GGGTTACTTC GGTGATTGGG CGTCTCACAC TATGGAACAC GCAGTGAGCG CAGTGTACGA
CATTCCGCAC GCGGGCGGTC TGGCGATTCT GTTTCCGAAT TGGATGCGTT ACACGCTGGA
TACTAACGTG GGTCGTTTCA AAAACCTGAT GCTGAACATG TTCGATATCG ATACGGAAGG
CAAAACTGAC AAGGAGATCG CCCTGGAAGG TATTGACAAA CTGTCCGCAT TTTGGACGAG
CCTGGGCGCG CCGTCCCGTC TGGCCGATTA CAACATCGGC GAAGAAAAAC TGGAGCTGAT
CGCAGACATT GCTGCGAAAG AGATGGAGCA CGGCGGCTTC GGCAACTTTC AGAAGCTGAA
TAAAGACGAC GTACTGGCGA TCCTGCGTGC ATCTCTGTAA ACTAGTATCG ATGATAAGCT
GTCAAACATG AGCAGATCTG AGCCCGCCTA ATGAGC

ADH6 G1 (Accession No. A8SGI9_9FIRM)

CGAGCGCCCG TAACTTCACG CGCCAACGTC GTTGTGTTCT GGCCGGCTAA TCTAGACCTC
TCCCGGTACG ATAATAAGGA GGCATCAATG AACAACTTCC TGTTCGAAAA CAAAACCAAA
GTATACTTCG GTAAGGGTGG TGTTAAAGAA TATCTGGGTT GTCTGCTGGA ACATTATGGT
GACACCGTTA TGCTGGCCTA TGGCGGCGGC TCCATCAAAC ATAACGGTGT ATATGATGAA
ATTGTGGGCA TCCTGAACGC CGAAGGCAAA CGCATCGTTG AATTCCCGGG TATCATGCCG
AACCCGACGT ATGCTAAGGT GCAAGAAGGT GCTAAACTGG CGCGTGAAAA CCACGTAGAC
CTGATCCTGG CCGTTGGCGG TGGTAGCGTT TCCGACTGCT GCAAAGTTGT GAGCGCGCAG
GCAAAAGTAG ATGAAGATCT GTGGGAGCTG GAAAACACTA AACACACTCG CCCGACTGCA
TTCATTCCGC TGGGTACCAT TGTGACCGTT TTTGGTACTG GCAGCGAAAT GAACAACGGC
GCTGTAATCA CCCACGAGGA GAAAAAAATT AAAGGTGCTC TGTGGGGCGC ACAGGCGGAC
TTTGCATTCC TGGACCCGAC TTATACTCTG TCCGTGCCGA TGAAACAGGT TATTAGCGGT
GCGTTCGACA CTCTG

ADH6 G2 (Accession No. A8SGI9_9FIRM)

ACTCTGTCCG TGCCGATGAA ACAGGTTATT AGCGGTGCGT TCGACACTCT GAGCCACGCT
ATGGAAACTT ATTTCGGCAA ACCGGATGAG AACAATCTGT CCGACGACAT CAACGAAGCG
GTGATGCGTT CCGTTATCCG TAACATTCGT GTGCTGCTGA CCGACAAGGA TAACTACGAA
GCACGCTCCG AACTGACCTG GGCTTCTGCG ATGGCAGAAA ACGGTATTCT GAAAATCGGT
AAAGTAACTG ACTTTCAATG CCACATGATC GAACATCAGC TGGGCGCATA CACTAACTGT
AACCACGGCG CTGGTCTGGC GGTTATCCAC CCGGTTCTGT ATCGTCATCT GCTGCCGGCG
AACACCGCAC GTTTCGCGCG TTTCGCTCAA AACGTTTGGG GCATCGATCC AGCAGGTAAA

TCCGAACTGA AACTGGCGCA GGCGGGTGTG GAAGCTCTGG CGGCGTTTAT CAAGGAAATT
GGCATGCCGA CTACCTTCGC TGAGCTGGGC GTTCCGGCGG ACACCGATCT GAAAGCCGTA
GCTGACTCTA CCGTCCTGAC CGGTGGTTGT TGCAAAAAAC TGTCTCGTGA AGAGCTGCTG
GACATCCTGA ACGAATGTAA ATAAACTAGT ATCGATGATA AGCTGTCAAA CATGAGCAGA
TCTGAGCCCG CCTAATGAGC

ADH7 G1 (Accession No. E2SQ66_9FIRM)

CGAGCGCCCG TAACTTCACG CGCCAACGTC GTTGTGTTCT GGCCGGCTAA TCTAGATATC
CAGCCATTCC CCAGGAGAAA CCACTATGCG TAACTTTACC TACCACAACC CGGTCCGTAT
CCTGTTCGGC GATCATGCTC TGGACCAGCT GCCGGATCTG TTCCGTGAAT TCCACGTGTC
TAACCTGCTG CTGGTGTATT CTGGCGATTT TATTAAAGAA CTGGGCATCT GGGATGCCGT
TTACAACGCT TGCGCGGAAA ATGGTATCGC ATTTTACGAA GAAGGTGGTG TAGTCCCGAA
CCCGAAAATT GAACTGGTTC GTGAACTGGT CGCACTGGGC AAAAAAAAAA AGATCGACTT
CATTCTGGCT GTAGGCGGTG GTTCTTCCAT CGACACTGCT AAGGCTGTTG CCGCAGGCAT
CCCGTACGCC CACGACGTGT GGGACTTCTT CGAATACACT GCGGTTCCGG AAACGGCGGT
GCCGATCGGT GTAATCACCA CGATCCCAGC GTCTGGTTCC GAATGTTCTA ATTGCAGCAT
TATCTCCAAC GGTCTGCACA AATGCGGTAT TGAGTACGAT TGCATCATCC CACAGTTTGC
CATCATGAAC CCGGAGTACA CCCGTACCCT GCCTGCGTAC CAGACCTCCG CAGGCATCGC
GGACATTCTG TCCCA

ADH7 G2 (Accession No. E2SQ66_9FIRM)

GTACCCTGCC TGCGTACCAG ACCTCCGCAG GCATCGCGGA CATTCTGTCC CACATGCTGG
AACGCTACTT CACGAACACT ACTCACGTTG ACACCACCGA CTACATGCTG GAAGGTACCA
TGCAGGCTCT GATGGTCAAC GCGCGCCGCC TGATGAAACA GCCGGATGAC ATCCACGCGC
GCGCAGAAGT TCAGTGTCTG GCTTTCCTGG CACATAACAA CCTGCTGGAC ATCGGTCGCG
AATCTGACTG GGGCCCGCAT CGTATTGAAC ACGAACTGTC CGCACAGTAC GGCATTACCC
ACGGTGAAGG TATGGCAGTT GTAACCATCG CGTGGGCACG CTACATGGCT GCACACCACC
CGGACAAACT GGCACAGCTG GCCTCCCGTA TCTTCGGTGC TGATCCGTTT GTACATTCCA
AAGAGGATAT GGCACTGCTG CTGGCTGACC ACCTGGAAGA GTTTTTCAAA TCCCTGCACC
TGAAAACCAC CCTGCACGAA ATGGGTATCG ACGATACCCA CTTTGAAGAG ATGGCAAACC
GTGCCACCAA TAACGGTAAG GATTGTGTTG GCCACTACGT GGCTCTGAAC AAACAGATCT
TTATCGACAT TCTGCACATG GCCCTGTAAA CTAGTATCGA TGATAAGCTG TCAAACATGA
GCAGATCTGA GCCCGCCTAA TGAGC

ADH8 G1 (Accession No. E1QYZ8_OLSUV)

CGAGCGCCCG TAACTTCACG CGCCAACGTC GTTGTGTTCT GGCCGGCTAA TCTAGAGACT
TAGTAGTCAC ACGCAAGAGG AGGATTCCAG TATGTACGAC TTCATGTTCC ACGTACCGAC
CAAGATCTAC TTCGGCCGCG GCCAGATCTC TCACCTGGCA GAACTGTCTG ATTTTGGCCA
GAAAGCGCTG CTGGTTTACG GTGGCGGCAG CATCAAACGT AACGGCATTT ACGACGAAGC
GATTCGTATT CTGACCCATG CGGGTATCGA AGTTGTAGAA CTGAGCGGCG TTGAACCGAA
CCCGCGTATT GAAACCGTGC GTCGCGGTGT CGGTCTGTGC GCTCGCGAAG GTGTTGACAT
GGTTCTGGCT ATCGGCGGCG GTAGCACCAT CGATTGCGCT AAAGTAGTTG CGGCCGGCGC
GCGTTACGAT GGCGACCCGT GGGACCTGGT ACTGGACGGT TCTAAGGCGG CTTCCGCGCT
GCCAATCTTT TCTGTGCTGA CCCTGTCCGC GACCGGTTCT GAGATGGATG CATTCGCTGT
CATCAGCGAT ATGAGCAAAA ATGAAAAGTG GGGTACCGGC GCAGAGTGTA TGAAACCGAC
CATGTCTGTG CTGGACCCGT CTTACACCTT CAGCGTGAGC CCTAAACAGA CCGCGGCTGG
CACCGCCGAT ATGAT

ADH8 G2 (Accession No. E1QYZ8_OLSUV)

ACACCTTCAG CGTGAGCCCT AAACAGACCG CGGCTGGCAC CGCCGATATG ATGAGCCATA
CCTTCGAATC TTATTTTTCC ATGGACGAAG GTGCGTACGT CCAGAAGCGT CTGGCAGAAG
GTCTGCTGGG CACTATGATC CACTTCGGCC CGATTGCCCT GGCACATCCG GACGACTACG
ATGCGCGTGC GAACCTGATG TGGGCGGCTT CTCACGCAAT TAACGGCCTG GTTTCTGATG
GTTGTAGCCC TGCCTGGTGC GTTCACCCGA TGGAACACGA GCTGTCTGCA TTCTACGATA
TCACTCACGG CGAGGGTCTG GCGATCCTGA CGCCGGCATG GATGGAGCAC GTTCTGGATG

CTCAGACTGC TCCTCTGTTT GCTGCATACG GTTGCAACGT ATGGGGTCTG TCCGGCGTAG
ATGACATGAA AGTTGCTCGT GAAGCAATCA GCCGCACTCG TGCGTTTTTT GTTGAAGCTA
TGCATCTGCC GGCAACCCTG CGCGAGGTCG GCATTACCGA TGAAAAAAAC TTCGAAGTTA
TGGCTCGCAA AGCCGCCGAT GGTTGCAAAG GCAGCTTCGT TGCGCTGTCT CAGGACGACA
TCGTAGAAAT CTACCGTGCT GCTCTGTAAA CTAGTATCGA TGATAAGCTG TCAAACATGA
GCAGATCTGA GCCCGCCTAA TGAGC

ADH9 G1 (Accession No. F5X0G1_STRG1)

CGAGCGCCCG TAACTTCACG CGCCAACGTC GTTGTGTTCT GGCCGGCTAA TCTAGACTCC
TTCAATAAGC CCAGGGAGGA TTAAAGCATG AATGATTTCC AGTTTCAGAA CACTACCAAA
GTTTATTTCG GTAAACATCA GCTGCAACAC CTGCACCAGG AAGTGCTGAA ATACGGTCAG
AAAGTGCTGA TCGCTGATGG CGGTGAATTC ATCCGTCAGT CTCCGCTGTA TGCTCAAGTT
CTGAAAGAAC TGACGGACAA CGGCATCCAG ATCTTCGAAC TGGGTTCTGT GGAGCCGAAT
CCGCGCCACA CCACCGTTAA CCGCGAAGTA AAACTGTGTA AAGGCAACAA CATCCAGACC
GTACTGGCCG TTGGCGGCGG CTCCACGATT GACTGCTGTA AAGCGATCGC GGCGACCTCT
TGCACCGACG AAGACGACGT TTGGACCCTG ATCGAAAAAC GTGAACCGAT CAACCAAGCG
CTGGCGGTTA TCGCTATGCC GACCATCGCG TCCACGGGCT CTGAAATGGA CAAGAGCTGC
GTGATTGCCA ACGAAGAGCT GCACCTGAAA AAGGGTCTGA ACGGCGAAGC TATCCGTCCG
AAAGCGGCTT TTCTGAACCC GGAAAACACC TTCACCGTTC CGGCACGTCA GACCGCGTGT
GGTGGCTTCG ACATCATGAT GCATCTGCTG GATAT

ADH9 G2 (Accession No. F5X0G1_STRG1)

CGTCAGACCG CGTGTGGTGG CTTCGACATC ATGATGCATC TGCTGGATAT GAACTATTTT
GTAGACTCTG ATAAATATCC GCTGCAGTTC AATGTGGTAG AAACCCTGCT GCGCACTATT
CGTGAGCAGC TGCCGATCGC GCTGCGTGAG CCGGAAAACT ACGAGGCTCG TGCGACCCTG
CTGTGGGGTG CTTCCTGGGC GCTGAACTCT TTCTGTACCT CCGGTTTCAA AACCGCACCG
AGCAACCACG GTCTGGAACA ATTCTCTGCG TTCTACGATC AGACGCATGG TCTGGGTCTG
GCTCTGGTGG TTACCAAATG GATGACCTAC CTGCTGGAAA AGGACCCGAC CGTGGCACCA
GATTTCGCTC GTCTGGGCAC CAATGTGCTG GGCTGTCAGC CAGTTGACGA TGTGATCGAG
GGCGCAAAAA ACGCTATCAA AGCCTTTGAC GCATTCATTG TGAATGACCT GGGTCTGCCG
CGTACCATGA CTGAAATCGG TCTGAACGAC TCTAAGCTGA GCGAGATGGC TCATGCTGCG
GTAACCGGTT ATGGCGACGG CACGCTGAAG GGCTACCGTG AACTGACTGA AGCGAACTGC
CTGGCCATTT ATAAAATGTG CCTGTAAACT AGTATCGATG ATAAGCTGTC AAACATGAGC
AGATCTGAGC CCGCCTAATG AGC

ADH10 G1 (Accession No. E6W4G5_DESIS)

CGAGCGCCCG TAACTTCACG CGCCAACGTC GTTGTGTTCT GGCCGGCTAA TCTAGAGCCT
TAATCCCCGT AAGCACAGGA GATCCACAAT GCAGAATTTC GTTTTTCACA ACCCGACCCG
TATCGTTTTC GGCCGTGACA AGACGGCGAG CATCGGCAAG GCGACCCTGC CGTATGGTCG
CCGCGTTCTG CTGCTGACGG GTCAGGGTTC CGTCGTGAAA CACGGTATCC TGGCGAAAGT
GACCTCTTCC CTGTCTACTG CGGGTATCTC CTGGGTTGAG TGTAGCGGTG TGCAGCCGAA
CCCGGTTCTG GGCTTCGTGC GTCAGGCCAT CGACACTTTC CGTCGTGAAA ACCTGGACGC
CATTGTAGCG GTTGGCGGTG GCTCCGTGAT CGACACCGCG AAGGCGGTGG CTGCGGGCGT
TCGTTACGAA GGCGATGTTT GGGACTTCTT TACCGGTAAA GCTAACGTCC TGGACGCGGC
CCCGATCACT GTAGTGCTGA CTCTGCCGGC GGCTGCATCC GAGATGAACA GCGGCGGTGT
TATCACTAAT GAACAAACTC GTCAAAAATT CAACCTGGGC GGCGAACCGC TGTCTCCGAA
AGTTTCTATC CTGGACCCGG TCAACAGCTT TAGCGCCCCG GTGAATCACT CCCTGTACGG
TGTTGTTGAC GCGAT

ADH10 G2 (Accession No. E6W4G5_DESIS)

ACAGCTTTAG CGCCCCGGTG AATCACTCCC TGTACGGTGT TGTTGACGCG ATGGTTCATC
TGCTGGAGGG CTACTTCAAC GGCTCTGACC CGTGGACTCC ACTGCAGGAC CGTTACGCGG
AAGGTATCAT TCGCACTCTG ATGGAATGCG CTGCCATTAT TCGTGAACAG CCAGACCACT
ACGACGCACG TGCTAACATC ATGTGGGGCG CGACTCTGGC TTTCAACGGC CTGGCACCGT
GCGGTATCGG CCCGGCAGGT TTTCCGATGC ACATGATCGA ACACAGCCTG TCTGCACTGT

ATGATGTATC TCATGGTGCG GGTCTGGCGA TGATCCTGCC GGGTTGGCTG AAGTACCACT
CCGATTCCAG CCCGCGCAAA GTTAACCAGT TTGGCCGTCG TATTTTTGAA CTGGATCACC
AGGATGATCG TCAGGGCGCT CAAGCAGCCA TTGCCGAGCT GGAACGTTGG CTGCGTTCCA
TGGATATCCC GGCATCCCTG CACGAAGGTG GCATCCCGAT CGATGAGATC CCAGCAATTG
CGGAGAACGC TGTGATGCTG GCGCAGAAAT GGGGTCTGAA AGCTTACACT CAGGCCGTTA
TCGAAGACGT TCTGCGTCGC GCTTCTCGCT AAACTAGTAT CGATGATAAG CTGTCAAACA
TGAGCAGATC TGAGCCCGCC TAATGAGC

ADH11 G1 (Accession No. E6K7W2_9BACT)

CGAGCGCCCG TAACTTCACG CGCCAACGTC GTTGTGTTCT GGCCGGCTAA TCTAGATGAT
CCCTCCACAA CTAAAGGCGG TATTCAAATG AAAGACTTCA ACTTCTACGC ACCGACCCGT
GTAGTGTTCG GCAAACAGAG CGAAGAGCAG CTGCCGCGCC TGCTGAAAGA AGCGGGTGGT
AAAAAGGTTC TGGTACACTA TGGTGGCGGC TCTGCAAAAC GTTCTGGCCT GCTGGATAAA
GTGTATGGTA TGCTGGACGA CGCGGGCATC GAACATGTAG GTCTGGGCGG TGTAGTACCG
AACCCGCTGC TGTCCAAAGT AAACGAAGGC ATTGACCTGT GCCGTCGTAA AGGTGTAAAC
TTCATTCTGG CTGTAGGCGG CGGCTCCGTA ATCGATAGCG CGAAAGCAAT TGCGTATGGT
GTGCCGTACG AGGGTGACGT TTGGGATTTC TGGAATGGTA AGCCGGCAAC CGCTGCCCTG
CCGGTCGGTG CAATGCTGAC TATCCCGGCT GCTGGCTCTG AAATGAGCAA TTCTTGCGTG
ATTACTAAAG ACGAAGGTGC TGTTAAACGT GGCTTCAACA ACGATCTGTG CCGCTGTAAA
TTCGCGATCA TGAACCCAGA ACGCACTTAC ACGCTGCCGC CGTACCAGAC TGCCGCGGGT
GCGACCGACA TCATG

ADH11 G2 (Accession No. E6K7W2_9BACT)

CACTTACACG CTGCCGCCGT ACCAGACTGC CGCGGGTGCG ACCGACATCA TGATGCACAC
CATGGAACGC TACTTTTCCA AACATGAAGA CATGACCCTG ACCGACGCAA TTGCGGAAGC
CCTGCTGCGC ACGGTTAAAG AAAGCACCTT CGAAGTGCTG AAACACCCGG AGGACTACCG
TAACCGCGCT CAGATTATGT GGGCCGGCTC CCTGTCTCAT AACGATCTGA CCGAATGTGG
TCTGGAAAAG GATTTCGCGA CTCACCGCCT GGAACACGAG CTGTCTGCGC TGTTCGGCGT
TACCCATGGC GCCGGCCTGG CAGCCGTGTG GCCTGCATGG GCGCGTTATG TGATGAAGAA
ACACATTTCC CGCTTCGTTC AGTTCGCGGT CAACGTGATG GGCGTTCCGA ACGACTTTTC
TAACCCGGAA GCTACCGCTG AGAAAGGTAT CTGTCGTATG GAACACTTCT TCCACGCGAT
CGGTATGCCG ACCTCCATCA AAGAACTGCT GGGTCATGAT ATCACCGAAG CGCAGATTGA
CGAAATGGTT GACAAATGCT CTCGTGGTGG TACTATCACT GTTGGTGCCA TGGAGGTGAT
TGCCCCAGAC GACATGCGTG CGATCTACCG TATGGCACGC TAAACTAGTA TCGATGATAA
GCTGTCAAAC ATGAGCAGAT CTGAGCCCGC CTAATGAGC

ADH12 G1 (Accession No. B1C4Z8_9FIRM)

CGAGCGCCCG TAACTTCACG CGCCAACGTC GTTGTGTTCT GGCCGGCTAA TCTAGATGGT
TCTACAATAA TAGGAGGACT CTACACATGC TGGGCGACTT TACCTACTCC AACCCGACGA
AAATTTATTT CGGCGAGAAC TCTCTGGACA ACCTGTCTAC CGAACTGAAA AACTATGGCA
AGAACGTGCT GCTGGTATAC GGTGGTGGTT CTATCAAAAA AAACGGTATC TACGATAAGG
TTATCGACAT TCTGAAAAAG TGTGATAAGA CTATTATTGA GGATGCGGGC GTAATGCCTA
ATCCGACTGT TGAAAAGCTG TATGAAGGTT GCAAACTGGC TCGTGAAGGT AACGTTGACC
TGATTCTGGC GGTTGGCGGT GGCAGCGTGT GTGACTACGC GAAAGCAGTT AGCGTCAGCA
CGTATTGCAA CGAGGATCCG TGGGAAAAGT ACTACCTGCG TATGGAGGAC GTTGATAACA
AAATTATCCC AGTTGGTTGT ATCCTGACCA TGGTTGGTAC TGGTTCCGAA ATGAATGGCG
GCTCTGTTAT CACCAATCAT GAACAGAAAC TGAAAATTGG TCACGTTTTC GGCGACAATG
TGTTCCCGAA GTTCTCCATT CTGAACCCGA CCTTCACCTA CACGCTGCCG AAATATCAGA
TGATCGCTGG TTTCT

ADH12 G2 (Accession No. B1C4Z8_9FIRM)

AACCCGACCT TCACCTACAC GCTGCCGAAA TATCAGATGA TCGCTGGTTT CTACGACATC
ATGTCCCATA TCCTGGAACA GTACTTTAGC GGTGAAGACG ACAACACCTC TGATTATATC
ATGGAAGGTC TGCTGAAATC TCTGATCCAT TCTAGCAAAA TTGCCGTGAA CGATCCTACC

AACTACGAGG CTCGTTCTAA CATCATGTGG ATTGCAACCT GGGCTCTGAA CACCCTGGTG
GCTAAAGGCA AAACCACGGA TTGGATGGTT CACATGATCG GCCAGAGCAT CGGTGCTTAC
ACCGACGCCA CGCATGGTAT GACCCTGGCT GCCGTGTCCA TTCCGTACTA CAAGTACATT
TGTCCATACG GCCTGAACAA ATTCAAACGC TATGCGATTA ACGTTTGGGA TGTTCTGTCT
GAAGGCAAAA CTGACGAGCA GATCGCTAAC GAAGGTCTGG AATGTATGGA AAAATACATG
CGTGACCTGG GTCTGGTAAT GAACATTTCC GATCTGGGCG TCAAAGAAGA GATGCTGGAG
GGTATCGCTG AAGGTACGTT CATCATGAAC GGCGGTTATA AAGTACTGAC CAAAGACGAA
ATTATCACCA TCCTGAAACA ATCCATGAAA TAAACTAGTA TCGATGATAA GCTGTCAAAC
ATGAGCAGAT CTGAGCCCGC CTAATGAGC

ADH13 G1 (Accession No. G4L3E3_TETHN)

CGAGCGCCCG TAACTTCACG CGCCAACGTC GTTGTGTTCT GGCCGGCTAA TCTAGAACGT
AAGGCCACTA CATTAACTAA GGAGCAAAAT ATGGAAAATT TCGATTTCCA CGTTACTACT
GATATCCGCT TTGGCAAAGA CCGTCTGGGT GAACTGCCGC AGGTTCTGAA CAACTTCGGC
AAAAACGTGC TGCTGGTTTA CGGTGGTGGC TCCATCAAGC GTAATGGTCT GTACGACAAA
CTGTACGAAC TGTTCAACCA GAACGACAAT AACGTTGTTG AACTGGCGGG TGTAGACCCG
AACCCGCGCA TTGAAACCGT GCAAAAAGGT GTCCAGCTGT GTAAGGAACA CGCGATCGAC
GTCGTGCTGC CGGTAGGTGG CGGCTCTGTG ATTGACTGCT CCAAAGCTGT GGCGGCTTGC
GTCTTTGTTA GCGGTGACCT GTGGGAAAAC TTCGTGCTGC AGAAAAACTA TAAAGGCCCG
GCACTGCCGA TTGTCACCAT TCTGACGCTG GCCGCTACGG GCTCTGAGAT GAACGGTACG
TGCGTAATCT CTAACATGGA TGCGCAGATT AAACTGGGCG TCCACGGTAC CACCAACCTG
CTGCCAAAGG TATCCTTCCT GGATCCGACT AACACCTTCT CTGTTGGTGC ATACCAGACT
GCAGCTGGCT CCGCTGACAT CCTGAGCCAC CTGAT

ADH13 G2 (Accession No. G4L3E3_TETHN)

TGGTGCATAC CAGACTGCAG CTGGCTCCGC TGACATCCTG AGCCACCTGA TGGAGAACTA
TTTCAACGCG ACCGAAGGCA CCGAAGTTCA GGATGAAATC GCTGAAGGCC TGATGAAAAC
GGTGATCAAA TATCTGCCGG TGGCGCTGGA CGAACCGGAC AACTATATTG CCCGTGCTAA
CCTGATGTGG GCCTCTACTC TGGCGCTGAA CGGCCTGGTT GGCAAAGGTA AAAAAGGCAG
CTGGTCTTGT CATGCTATGG AACACGAACT GTCCGCTTTC TATGACATCA CTCACGGCGT
CGGCCTGGCT ATGCTGACCC CGCGTTGGAT GGCACACATC CTGGACGAAG ACACCCTGCC
GAAATTTCAA CGTTTTGCTG AAGAGGTCTG GAATGTTAAA GAAAAGGAAC CGAAACGTAC
GGCGGAGATC GGCATTCAGA AACTGTACGA TTTTTTCGTC TCCTGCAACA TCCCTATGAC
CCTGTCCGGT GTGGGCATCC AGACCGAAGA AAATTTTGAA GAAATGGGTC AGCGTGCCGT
TGCTCACTCC TCCATCTCTA ATCAGGGCTT CGTACCGCTG CACGAGGACG ACGTGGTCTC
CATCTATCGC GACTGCATGT CCGAGTCTTC TTTCGTCTAA ACTAGTATCG ATGATAAGCT
GTCAAACATG AGCAGATCTG AGCCCGCCTA ATGAGC

ADH14 G1 (Accession No. E8LLW8_9GAMM)

CGAGCGCCCG TAACTTCACG CGCCAACGTC GTTGTGTTCT GGCCGGCTAA TCTAGAAGTA
TATTTCCCGC TCAATATAAG GAGGAGTACA TATGGAATCT TTCGATTTTT TCCGTCGCAC
TCGTATCATC TTTGGCCAGT CTGCGGACAA CGAAGTAGGT CAGATTATCA AATATCAAGG
TGGCACTCGT GTGCTGCTGC TGCACGGTGA AAAAGCAGCG ATCAAGTACG GTATTGTGGA
GCGTATTGGT CGTACTCTGG ACCGTTCCGG TCTGAAATAC TTCTCCAAAG GCGGCATCAA
GAGCAACCCG CATATTGATA AAGTTTACGA ATGCATTGAA TTCTGCCTGT CCAACTCCAT
TAATTATATC CTGGCTGTGG GTGGTGGTTC CGTGATCGAC ACCGCCAAAA TCGTCGCGGC
GGGCGTATTC TTCGACGGCG ACATCTGGGA CATGTTTGAA AAACATCGCG AACCGTACCG
TTCCCTGCCG CTGGGCTGCG TAGTTACCGT TCCTGCAAGC GGTACTGAAT GCAGCAACTC
TTCTTCCCTG ATGCGTGAAA AAGACGGCCG CCGTGAAAAA CTGATCGCGT ATTCTAACAG
CTTCGTACCG GAGTTCGCCA TTCTGAACCC GGACCTGACG CTGTCTCTGT CTCCGCGTGT
GACCGCTAGC GGTTGCGTTG ATATGATTAA CCATG

ADH14 G2 (Accession No. E8LLW8_9GAMM)

CTCTGTCTCC GCGTGTGACC GCTAGCGGTT GCGTTGATAT GATTAACCAT GTCCTGGAAG
GTTATTTCTC CAACTCTACC GGTGTACTGC TGAGCGATAA GCTGTGTGAA GCGGTTCTGA
GCTCTATTAT CGAACTGCTG CCGCAGATCT ATGAAGATCC GAATAACATT GATGCGCGCG

CAAACCTGAT GCTGGCAGCA ACCCTGTCTC ACAATGATAT CTGCTGCATG GGCCGCAAGT
CCGACAACGT TATCACGAAA CTGGCCAACC AGCTGGTGGT TGAAAACGAT TGTCCGTTCG
GTGATGCACT GGCTGTTCTG ATCCCGGCTT GGATGGAATA TGTTGTTCAG TTTAACCCGC
TGCGCATCGC ACAATTCTCC AACCGCGTTT TTGGTATCGC AATCAACTTC GAAGATCCGA
AAATTACCGC GTATGACGGT ATCAAAGCCC TGCGCGCTTT TTTCAAAAAT GTAAAACTGC
CGTGCAACTT CGTTGAACTG GGTATCAAGA CCGAAGCAAT CGCGGACATC GTAAACGCTC
TGGACCTGAA AGAAGGTAAA ACTCTGGGTT CTTTTGTGCC GCTGGACGCT GTGGCCTGCG
AAGCAATCCT GTCCCTGGCC GCCAATTACT GCGAAGGTCG CGATATTTTC TAAACTAGTA
TCGATGATAA GCTGTCAAAC ATGAGCAGAT CTGAGCCCGC CTAATGAGC

Appendix 2.7: Open reading frames

pcnB

```
   1  GTGCTAAGCC GCGAGGAAAG CGAGGCTGAA CAGGCAGTCG CCCGTCCACA GGTGACGGTG
  61  ATCCCGCGTG AGCAGCATGC TATTTCCCGC AAAGATATCA GTGAAAATGC CCTGAAGGTA
 121  ATGTACAGGC TCAATAAAGC GGGATACGAA GCCTGGCTGG TTGGCGGCGG CGTGCGCGAC
 181  CTGTTACTTG GCAAAAAGCC GAAAGATTTT GACGTAACCA CTAACGCCAC GCCTGAGCAG
 241  GTGCGCAAAC TGTTCCGTAA CTGCCGCCTG GTGGGTCGCC GTTTCCGTCT GGCTCATGTA
 301  ATGTTTGGCC CGGAGATTAT CGAAGTTGCG ACCTTCCGTG GACACCACGA AGGTAACGTC
 361  AGCGACCGCA CGACCTCCCA ACGCGGGCAA AACGGCATGT TGCTGCGCGA CAACATTTTC
 421  GGCTCCATCG AAGAAGACGC CCAGCGCCGC GATTTCACTA TCAACAGCCT GTATTACAGC
 481  GTAGCGGATT TTACCGTCCG TGATTACGTT GGCGGCATGA AGGATCTGAA GGACGGCGTT
 541  ATCCGTCTGA TTGGTAACCC GGAAACGCGC TACCGTGAAG ATCCGGTACG TATGCTGCGC
 601  GCGGTACGTT TTGCCGCCAA ATTGGGTATG CGCATCAGCC CGGAAACCGC AGAACCGATC
 661  CCTCGCCTCG CTACCCTGCT GAACGATATC CCACCGGCAC GCCTGTTTGA AGAATCGCTT
 721  AAACTGCTAC AAGCGGGCTA CGGTTACGAA ACCTATAAGC TGTTGTGTGA ATATCATCTG
 781  TTCCAGCCGC TGTTCCCGAC CATTACCCGC TACTTCACGG AAAATGGCGA CAGCCCGATG
 841  GAGCGGATCA TTGAACAGGT GCTGAAGAAT ACCGATACGC GTATCCATAA CGATATGCGC
 901  GTGAACCCGG CGTTCCTGTT TGCCGCCATG TTCTGGTACC CACTGCTGGA GACGGCACAG
 961  AAGATCGCCC AGGAAAGCGG CCTGACCTAT CACGACGCTT TCGCGCTGGC GATGAACGAC
1021  GTGCTGGACG AAGCCTGCCG TTCACTGGCA ATCCCGAAAC GTCTGACGAC ATTAACCCGC
1081  GATATCTGGC AGTTGCAGTT GCGTATGTCC CGTCGTCAGG GTAAACGCGC ATGGAAACTG
1141  CTGGAGCATC CTAAGTTCCG TGCGGCTTAT GACCTGTTGG CCTTGCGAGC TGAAGTTGAG
1201  CGTAACGCTG AACTGCAGCG TCTGGTGAAA TGGTGGGGTG AGTTCCAGGT TTCCGCGCCA
1261  CCAGACCAAA AAGGGATGCT CAACGAGCTG GATGAAGAAC CGTCACCGCG TCGTCGTACT
1321  CGTCGTCCAC GCAAACGCGC ACCACGTCGT GAGGGTACCG CATGA
```

rpoC

```
   1  gtgAAAGATT TATTAAAGTT TCTGAAAGCG CAGACTAAAA CCGAAGAGTT TGATGCGATC
  61  AAAATTGCTC TGGCTTCGCC AGACATGATC CGTTCATGGT CTTTCGGTGA AGTTAAAAAG
 121  CCGGAAACCA TCAACTACCG TACGTTCAAA CCAGAACGTG ACGGCCTTTT CTGCGCCCGT
 181  ATCTTTGGGC CGGTAAAAGA TTACGAGTGC CTGTGCGGTA AGTACAAGCG CCTGAAACAC
 241  CGTGGCGTCA TCTGTGAGAA GTGCGGCGTT GAAGTGACCC AGACTAAAGT ACGCCGTGAG
 301  CGTATGGGCC ACATCGAACT GGCTTCCCCG ACTGCGCACA TCTGGTTCCT GAAATCGCTG
 361  CCGTCCCGTA TCGGTCTGCT GCTCGATATG CCGCTGCGCG ATATCGAACG CGTACTGTAC
 421  TTTGAATTCT ATGTGGTTAT CGAAGGCGGT ATGACCAACC TGGAACGTCA GCAGATCCTG
 481  ACTGAAGAGC AGTATCTGGA CGCGCTGGAA GAGTTCGGTG ACGAATTCGA CGCGAAGATG
 541  GGGGCGGAAG CAATCCAGGC TCTGCTGAAG AGCATGGATC TGGAGCAAGA GTGCGAACAG
 601  CTGCGTGAAG AGCTGAACGA AACCAACTCC GAAACCAAGC GTAAAAAGCT GACCAAGCGT
 661  ATCAAACTGC TGGAAGCGTT CGTTCAGTCT GGTAACAAAC CAGAGTGGAT GATCCTGACC
 721  GTTCTGCCGG TACTGCCGCC AGATCTGCGT CCGCTGGTTC CGCTGGATGG TGGTCGTTTC
 781  GCGACTTCTG ACCTGAACGA TCTGTATCGT CGCGTCATTA ACCGTAACAA CCGTCTGAAA
 841  CGTCTGCTGG ATCTGGCTGC GCCGGACATC ATCGTACGTA ACGAAAAACG TATGCTGCAG
 901  GAAGCGGTAG ACGCCCTGCT GGATAACGGT CGTCGCGGTC GTGCGATCAC CGGTTCTAAC
 961  AAGCGTCCTC TGAAATCTTT GGCCGACATG ATCAAAGGTA AACAGGGTCG TTTCCGTCAG
1021  AACCTGCTCG GTAAGCGTGT TGACTACTCC GGTCGTTCTG TAATCACCGT AGGTCCATAC
1081  CTGCGTCTGC ATCAGTGCGG TCTGCCGAAG AAAATGGCAC TGGAGCTGTT CAAACCGTTC
1141  ATCTACGGCA AGCTGGAACT GCGTGGTCTT GCTACCACCA TTAAAGCTGC GAAGAAAATG
1201  GTTGAGCGCG AAGAAGCTGT CGTTTGGGAT ATCCTGGACG AAGTTATCCG CGAACACCCG
1261  GTACTGCTGA ACCGTGCACC GACTCTGCAC CGTCTGGGTA TCCAGGCATT TGAACCGGTA
1321  CTGATCGAAG GTAAAGCTAT CCAGCTGCAC CCGCTGGTTT GTGCGGCATA TAACGCCGAC
1381  TTCGATGGTG ACCAGATGGC TGTTCACGTA CCGCTGACGC TGGAAGCCCA GCTGGAAGCG
1441  CGTGCGCTGA TGATGTCTAC CAACAACATC CTGTCCCCGG CGAACGGCGA ACCAATCATC
1501  GTTCCGTCTC AGGACGTTGT ACTGGGTCTG TACTACATGA CCCGTGACTG TGTTAACGCC
1561  AAAGGCGAAG GCATGGTGCT GACTGGCCCG AAAGAAGCAG AACGTCTGTA TCGCTCTGGT
```

```
1621  CTGGCTTCTC TGCATGCGCG CGTTAAAGTG CGTATCACCG AGTATGAAAA AGATGCTAAC
1681  GGTGAATTAG TAGCGAAAAC CAGCCTGAAA GACACGACTG TTGGCCGTGC CATTCTGTGG
1741  ATGATTGTAC CGAAAGGTCT GCCTTACTCC ATCGTCAACC AGGCGCTGGG TAAAAAAGCA
1801  ATCTCCAAAA TGCTGAACAC CTGCTACCGC ATTCTCGGTC TGAAACCGAC CGTTATTTTT
1861  GCGGACCAGA TCATGTACAC CGGCTTCGCC TATGCAGCGC GTTCTGGTGC ATCTGTTGGT
1921  ATCGATGACA TGGTCATCCC GGAGAAGAAA CACGAAATCA TCTCCGAGGC AGAAGCAGAA
1981  GTTGCTGAAA TTCAGGAGCA GTTCCAGTCT GGTCTGGTAA CTGCGGGCGA ACGCTACAAC
2041  AAAGTTATCG ATATCTGGGC TGCGGCGAAC GATCGTGTAT CCAAAGCGAT GATGGATAAC
2101  CTGCAAACTG AAACCGTGAT TAACCGTGAC GGTCAGGAAG AGAAGCAGGT TTCCTTCAAC
2161  AGCATCTACA TGATGGCCGA CTCCGGTGCG CGTGGTTCTG CGGCACAGAT TCGTCAGCTT
2221  GCTGGTATGC GTGGTCTGAT GGCGAAGCCG GATGGCTCCA TCATCGAAAC GCCAATCACC
2281  GCGAACTTCC GTGAAGGTCT GAACGTACTC CAGTACTTCA TCTCCACCCA CGGTGCTCGT
2341  AAAGGTCTGG CGGATACCGC ACTGAAAACT GCGAACTCCG GTTACCTGAC TCGTCGTCTG
2401  GTTGACGTGG CGCAGGACCT GGTGGTTACC GAAGACGATT GTGGTACCCA TGAAGGTATC
2461  ATGATGACTC CGGTTATCGA GGGTGGTGAC GTTAAAGAGC CGCTGCGCGA TCGCGTACTG
2521  GGTCGTGTAA CTGCTGAAGA CGTTCTGAAG CCGGGTACTG CTGATATCCT CGTTCCGCGC
2581  AACACGCTGC TGCACGAACA GTGGTGTGAC CTGCTGGAAG AGAACTCTGT CGACGCGGTT
2641  AAAGTACGTT CTGTTGTATC TTGTGACACC GACTTTGGTG TATGTGCGCA CTGCTACGGT
2701  CGTGACCTGG CGCGTGGCCA CATCATCAAC AAGGGTGAAG CAATCGGTGT TATCGCGGCA
2761  CAGTCCATCG GTGAACCGGG TACACAGCTG ACCATGCGTA CGTTCCACAT CGGTGGTGCG
2821  GCATCTCGTG CGGCTGCTGA ATCCAGCATC CAAGTGAAAA ACAAAGGTAG CATCAAGCTC
2881  AGCAACGTGA AGTCGGTTGT GAACTCCAGC GGTAAACTGG TTATCACTTC CCGTAATACT
2941  GAACTGAAAC TGATCGACGA ATTCGGTCGT ACTAAAGAAA GCTACAAAGT ACCTTACGGT
3001  GCGGTACTGG CGAAAGGCGA TGGCGAACAG GTTGCTGGCG GCGAAACCGT TGCAAACTGG
3061  GACCCGCACA CCATGCCGGT TATCACCGAA GTAAGCGGTT TTGTACGCTT TACTGACATG
3121  ATCGACGGCC AGACCATTAC GCGTCAGACC GACGAACTGA CCGGTCTGTC TTCGCTGGTG
3181  GTTCTGGATT CCGCAGAACG TACCGCAGGT GGTAAAGATC TGCGTCCGGC ACTGAAAATC
3241  GTTGATGCTC AGGGTAACGA CGTTCTGATC CCAGGTACCG ATATGCCAGC GCAGTACTTC
3301  CTGCCGGGTA AAGCGATTGT TCAGCTGGAA GATGGCGTAC AGATCAGCTC TGGTGACACC
3361  CTGGCGCGTA TTCCGCAGGA ATCCGGCGGT ACCAAGGACA TCACCGGTGG TCTGCCGCGC
3421  GTTGCGGACC TGTTCGAAGC ACGTCGTCCG AAAGAGCCGG CAATCCTGGC TGAAATCAGC
3481  GGTATCGTTT CCTTCGGTAA AGAAACCAAA GGTAAACGTC GTCTGGTTAT CACCCCGGTA
3541  GACGGTAGCG ATCCGTACGA AGAGATGATT CCGAAATGGC GTCAGCTCAA CGTGTTCGAA
3601  GGTGAACGTG TAGAACGTGG TGACGTAATT TCCGACGGTC CGGAAGCGCC GCACGACATT
3661  CTGCGTCTGC GTGGTGTTCA TGCTGTTACT CGTTACATCG TTAACGAAGT ACAGGACGTA
3721  TACCGTCTGC AGGGCGTTAA GATTAACGAT AAACACATCG AAGTTATCGT TCGTCAGATG
3781  CTGCGTAAAG CTACCATCGT TAACGCGGGT AGCTCCGACT TCCTGGAAGG CGAACAGGTT
3841  GAATACTCTC GCGTCAAGAT CGCAAACCGC GAACTGGAAG CGAACGGCAA AGTGGGTGCA
3901  ACTTACTCCC GCGATCTGCT GGGTATCACC AAAGCGTCTC TGGCAACCGA GTCCTTCATC
3961  TCCGCGGCAT CGTTCCAGGA GACCACTCGC GTGCTGACCG AAGCAGCCGT TGCGGGCAAA
4021  CGCGACGAAC TGCGCGGCCT GAAAGAGAAC GTTATCGTGG GTCGTCTGAT CCCGGCAGGT
4081  ACCGGTTACG CGTACCACCA GGATCGTATG CGTCGCCGTG CTGCGGGTGA AGCTCCGGCT
4141  GCACCGCAGG TGACTGCAGA AGACGCATCT GCCAGCCTGG CAGAACTGCT GAACGCAGGT
4201  CTGGGCGGTT CTGATAACGA GTAA
```

Appendix 2.8: Genome sequencing results from evolved strains

A. Unassigned new junctions. Each new junction consists of two row, one describing one side of the junction in the reference sequence.

Product	Number	Position	Annotation	Gene	Product	Strain
HB	1	2787052	intergenic (-8/-579)	*ECDH1_RS13890/pgaA*	GGDEF domain-containing protein/poly-beta-1,6 N-acetyl-D-glucosamine export porin PgaA	2404
		3970989	intergenic (-39/+14)	*ECDH1_RS19625/ECDH1_ RS19630*	tyrosine recombinase/transposase	
	2	2787061	intergenic (-17/-570)	*ECDH1_RS13890/pgaA*	GGDEF domain-containing protein/poly-beta-1,6 N-acetyl-D-glucosamine export porin PgaA	2404
		3850814	intergenic (+252/-249)	*ECDH1_RS19025/ECDH1_ RS19030*	30S ribosomal protein S20/transposase	
	3	2991264	intergenic (-234/-36)	*ECDH1_RS14840/ECDH1_ RS14845*	hypothetical protein/transporter	2404
		3851581	intergenic (+15/+176)	*ECDH1_RS19030/ECDH1_ RS19035*	transposase/transcriptional activator NhaR	
	4	2991272	intergenic (-242/-28)	*ECDH1_RS14840/ECDH1_ RS14845*	hypothetical protein/transporter	
		3850814	intergenic (+252/-249)	*ECDH1_RS19025/ECDH1_ RS19030*	30S ribosomal protein S20/transposase	2404

Product	Number	Position	Annotation	Gene	Product	Strain
BDO	1	1967355	coding (176/213 nt)	*ECDH1_RS09640*	HTH domain-containing protein	2405, 2407
		2200475	intergenic (-424/-100)	*ECDH1_RS10830/ECDH1_RS10835*	hypothetical protein/NAD(P) transhydrogenase subunit alpha	
	2	1968549	pseudogene (3/624 nt)	*ECDH1_RS09650*	DNA-binding transcriptional regulator KdgR	2405, 2407
		2200472	intergenic (-421/-103)	*ECDH1_RS10830/ECDH1_RS10835*	hypothetical protein/NAD(P) transhydrogenase subunit alpha	
	3	1586747	intergenic (+99/-67)	ECDH1_RS07770/ECDH1_RS07775	hypothetical protein/IS5 family transposase	2405
		2568298	coding (291/720 nt)	*ECDH1_RS12695*	protein TonB	
	4	1586746	intergenic (+98/-68)	*ECDH1_RS07770/ECDH1_RS07775*	hypothetical protein/IS5 family transposase	2406
		2200475	intergenic (-424/-100)	*ECDH1_RS10830/ECDH1_RS10835*	hypothetical protein/NAD(P) transhydrogenase subunit alpha	
	5	1776306	intergenic (+146/-481)	*ECDH1_RS08610/ECDH1_RS08615*	IS5 family transposase/phosphogluconate dehydrogenase (NADP(+)-dependent, decarboxylating)	2406
		2200472	intergenic (-421/-103)	*ECDH1_RS10830/ECDH1_RS10835*	hypothetical protein/NAD(P) transhydrogenase subunit alpha	
	6	3970034	intergenic (-120/-94)	*fimA/ECDH1_RS19620*	type-1 fimbrial protein, A chain/hypothetical protein	2406, 2412
		3970348	intergenic (+32/+48)	*ECDH1_RS19620/ECDH1_RS19625*	hypothetical protein/tyrosine recombinase	

7	3970042	intergenic (-128/-86)	*fimA/ECDH1_RS19620*	type-1 fimbrial protein, A chain/hypothetical protein	2406, 2412
	3970338	intergenic (+22/+58)	*ECDH1_RS19620/ECDH1_RS19625*	hypothetical protein/tyrosine recombinase	
8	2200472	intergenic (-421/-103)	*ECDH1_RS10830/ECDH1_RS10835*	hypothetical protein/NAD(P) transhydrogenase subunit alpha	2408
	2449684	intergenic (-2/-68)	*ECDH1_RS12040/ECDH1_RS12045*	enterobacterial Ail/Lom family protein/IS5 family transposase	
9	2200475	intergenic (-424/-100)	*ECDH1_RS10830/ECDH1_RS10835*	hypothetical protein/NAD(P) transhydrogenase subunit alpha	2408
	2450879	pseudogene (211/216 nt)	*ECDH1_RS12050*	enterobacterial Ail/Lom family protein	
10	3872608	coding (20/141 nt)	*ECDH1_RS19135*	hypothetical protein	2409
	4006344	coding (265/267 nt)	*ECDH1_RS19830*	transposase	
11	3872612	coding (16/141 nt)	*ECDH1_RS19135*	hypothetical protein	2409
	4005127	coding (265/369 nt)	*ECDH1_RS19820*	transposase	
12	2668860	pseudogene (5/345 nt)	*ECDH1_RS13215*	hypothetical protein	2410, 2411
	2670689	coding (289/789 nt)	*ECDH1_RS13230*	integrase	
13	2668875	pseudogene (20/345 nt)	*ECDH1_RS13215*	hypothetical protein	2410, 2411
	2670672	coding (306/789 nt)	*ECDH1_RS13230*	integrase	
14	1776305	intergenic (+145/-482)	*ECDH1_RS08610/ECDH1_RS08615*	IS5 family transposase/phosphogluconate dehydrogenase	2410

					(NADP(+)-dependent, decarboxylating)	
		2568303	coding (286/720 nt)	*ECDH1_RS12695*	protein TonB	
	15	1771774	coding (946/1167 nt)	*ECDH1_RS08585*	O-antigen polymerase	2411
		1775112	pseudogene (447/450 nt)	*ECDH1_RS08605*	rhamnosyltransferase	

Product	**Number**	**Position**	**Annotation**	**Gene**	**Product**	**Strain**
n-butanol	1	1361359	intergenic (+30/+170)	*ECDH1_RS06715/ECDH1_RS06720*	sensor domain-containing phosphodiesterase/IS4 family transposase	2616
		1998641	coding (112/360 nt)	*ECDH1_RS09820*	hypothetical protein	
	2	1362696	intergenic (-55/+32)	*ECDH1_RS06720/ECDH1_RS06725*	IS4 family transposase/nucleoside permease NupC	2616
		1998651	coding (122/360 nt)	*ECDH1_RS09820*	hypothetical protein	
	3	3970034	intergenic (-120/-94)	*fimA/ECDH1_RS19620*	type-1 fimbrial protein, A chain/hypothetical protein	2616, 2619, 2620, 2621, 2622, 2628, 2630, 2686, 2687
		3970348	intergenic (+32/+48)	*ECDH1_RS19620/ECDH1_RS19625*	hypothetical protein/tyrosine recombinase	
	4	3970042	intergenic (-128/-86)	*fimA/ECDH1_RS19620*	type-1 fimbrial protein, A chain/hypothetical protein	2616, 2619, 2620, 2621, 2622, 2628, 2630, 2686, 2687
		3970338	intergenic (+22/+58)	*ECDH1_RS19620/ECDH1_RS19625*	hypothetical protein/tyrosine recombinase	

5	2668860	pseudogene (5/345 nt)	*ECDH1_RS13215*	hypothetical protein	2620, 2626, 2687, 2750
	2670689	coding (289/789 nt)	*ECDH1_RS13230*	integrase	
6	2668875	pseudogene (20/345 nt)	*ECDH1_RS13215*	hypothetical protein	2620, 2626, 2687, 2650
	2670672	coding (306/789 nt)	*ECDH1_RS13230*	integrase	
7	3971755	intergenic (-249/+491)	*ECDH1_RS19630/ECDH1_RS19635*	transposase/tyrosine recombinase	2620
	3978093	coding (698/1017 nt)	*ECDH1_RS19665*	hypothetical protein	
8	300335	coding (391/417 nt)	*ECDH1_RS01430*	hypothetical protein	2626
	1606068	coding (66/1557 nt)	*ECDH1_RS07850*	protein Rtn	
9	1606076	coding (58/1557 nt)	*ECDH1_RS07850*	protein Rtn	2626
	3851581	intergenic (+15/+176)	*ECDH1_RS19030/ECDH1_RS19035*	transposase/transcriptional activator NhaR	
10	3203305	coding (400/1428 nt)	*ECDH1_RS15885*	HscC co-chaperone, uncharacterized J domain-containing protein	2626
	3577068	intergenic (+146/-287)	*ECDH1_RS17700/ECDH1_RS17705*	IS5 family transposase/hypothetical protein	
11	4075591	coding (508/939 nt)	ECDH1_RS20165	hypothetical protein	2628, 2630, 2685, 2686, 2687
	4079177	pseudogene (1938/1959 nt)	*ECDH1_RS20175*	2',3'-cyclic-nucleotide 2'-phosphodiesterase	

12	4079177	pseudogene (1938/1959 nt)	*ECDH1_RS20175*	2',3'-cyclic-nucleotide 2'-phosphodiesterase	2628
	4154154	coding (1508/1539 nt)	*ECDH1_RS20595*	transcriptional regulator	
13	4079944	intergenic (+15/+144)	*ECDH1_RS20180/ECDH1_RS20185*	transposase/HxlR family transcriptional regulator	2628
	4154146	coding (1500/1539 nt)	*ECDH1_RS20595*	transcriptional regulator	
14	3505052	coding (403/3075 nt)	*lacZ*	beta-galactosidase	2630
	3505061	coding (412/3075 nt)	*lacZ*	beta-galactosidase	
15	3503661	coding (94/960 nt)	lacI	lac repressor	2686
	3503728	coding (161/960 nt)	*lacI*	lac repressor	
16	2678611	intergenic (+1/+29)	*ECDH1_RS13290/ECDH1_RS13295*	integrase/transposase	2750
	3044783	intergenic (-15/+126)	*ECDH1_RS15080/ECDH1_RS15085*	dehydrogenase/DNA-binding protein YbiB	
17	360753	coding (868/960 nt)	*lacI*	lac repressor	2726, 2729, 2730
	360815	coding (806/960 nt)	*lacI*	lac repressor	
18	1203246	coding (290/630 nt)	*BW25113_RS05990*	hypothetical protein	2728, 2729
	1205075	intergenic (-6/-66)	*BW25113_RS06005/BW25113_RS06010*	phage tail protein/DNA-invertase from lambdoid prophage e14	
19	361460	coding (161/960 nt)	*lacI*	lac repressor	2729
	361540	coding (81/960 nt)	*lacI*	lac repressor	

20	1203261	coding (305/630 nt)	*BW25113_RS05990*	hypothetical protein	2729
	1205058	coding (12/495 nt)	*BW25113_RS06005*	phage tail protein	
21	376716	coding (28/444 nt)	*BW25113_RS01855*	transferase	2730
	563704	intergenic (+1/-67)	BW25113_RS02785/BW25113_RS02790	protein ren/multidrug SMR transporter	
22	563698	pseudogene (208/213 nt)	*BW25113_RS02785*	protein ren	2730
	1462169	pseudogene (2521/2526 nt)	*BW25113_RS07350*	hypothetical protein	
23	3313550	coding (1315/1518 nt)	*BW25113_RS16450*	phosphate starvation-inducible protein PsiE	2730
	3576788	coding (391/417 nt)	*BW25113_RS17825*	hypothetical protein	
24	3313557	coding (1308/1518 nt)	*BW25113_RS16450*	phosphate starvation-inducible protein PsiE	2730
	3577555	intergenic (+15/-564)	*BW25113_RS17830/BW25113_RS17835*	transposase/heat-shock protein	
25	3179456	intergenic (+132/-90)	*BW25113_RS15795/BW25113_RS15800*	fimbrial-like adhesin protein/transposase	2748
	3995194	coding (409/951 nt)	*BW25113_RS19815*	magnesium transporter CorA	
26	3995189	coding (404/951 nt)	*BW25113_RS19815*	magnesium transporter CorA	2748
	4489328	intergenic (+11/+166)	*BW25113_RS22180/BW25113_RS23025*	integrase/phosphoethanolamine transferase YjgX	

B. Missing coverage

Product	Start	End	Size		Gene	Strain
BDO	1702686–1703700	1771773	68074–69088	*[ECDH1_RS08295]–[ECDH1_RS08585]*	*[ECDH1_RS08295], ECDH1_RS08300, ECDH1_RS08305, ECDH1_RS08310, ECDH1_RS08315, ECDH1_RS08320, ECDH1_RS08325, ECDH1_RS08330, ECDH1_RS08335, ECDH1_RS08340, ECDH1_RS08350, ECDH1_RS08355, ECDH1_RS08360, ECDH1_RS08365, ECDH1_RS08370, ECDH1_RS08375, ECDH1_RS08380, ECDH1_RS08385, ECDH1_RS08390, ECDH1_RS08395, ECDH1_RS08400, ECDH1_RS08405, ECDH1_RS08410, ECDH1_RS08415, ECDH1_RS08420, ECDH1_RS08425, ECDH1_RS08430, ECDH1_RS08435, ECDH1_RS08440, ECDH1_RS08445, ECDH1_RS08450, ECDH1_RS08455, ECDH1_RS08460, ECDH1_RS08465, ECDH1_RS08470, ECDH1_RS08475, ECDH1_RS08480, ECDH1_RS08485, ECDH1_RS08490, ECDH1_RS08495, ECDH1_RS08500, ECDH1_RS08505, ECDH1_RS08510, ECDH1_RS08515, ECDH1_RS08520, ECDH1_RS08525, ECDH1_RS08530, ECDH1_RS08535, ECDH1_RS08540, ECDH1_RS08545, ECDH1_RS08550, ECDH1_RS08555, ECDH1_RS08560, ECDH1_RS08565, ECDH1_RS08570, ECDH1_RS08575, ECDH1_RS08580, [ECDH1_RS08585]*	2411
n-butanol	3971196–3971755	3978092	6338–6897	*[ECDH1_RS19630]–[ECDH1_RS19665]*	[ECDH1_RS19630], ECDH1_RS19635, ECDH1_RS19640, ECDH1_RS19645, ECDH1_RS19650, ECDH1_RS19655, ECDH1_RS19660, [ECDH1_RS19665]	2620
	3192268–3193273	3203304	10032–11037	*[ECDH1_RS15835]–[ECDH1_RS15885]*	[ECDH1_RS15835], ECDH1_RS15840, ECDH1_RS15845, ECDH1_RS15850, artP, ECDH1_RS15860, ECDH1_RS15865, ECDH1_RS15870, ECDH1_RS15875, ECDH1_RS15880, [ECDH1_RS15885]	2626
	4075592	4079775–4079178	3587–4184	*[ECDH1_RS20165]–[ECDH1_RS20180]*	[ECDH1_RS20165], ECDH1_RS20170, ECDH1_RS20175, [ECDH1_RS20180]	2228, 2630, 2685, 2686, 2687
	2435616	2436637	1022	*ECDH1_RS11980*	ECDH1_RS11980	2630

Appendix 3: *Strains, plasmids, oligonucleotides, and sequences, RNA-sequencing results, and metabolomics data for Chapter 3*

Appendix 3.1: Strains

E. coli DH10B was used for DNA construction. *E. coli* DH1 (ATCC 39936) and all other strains were used for production and evolution experiments.

Organism	Name	Description	Number	Source
E. coli	DH10B	F- endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ-	55	*Invitrogen*
E. coli	DH1Δ*5*	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC*	799	*Dr. Miao Wen*
E. coli	BW25113Δ*5*-T1R	BW25113 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC* Δ*fhuA*, P1 transduced *fhuA*:Km^R from 1637 parent to 1435 then recycled Km marker	1691	*Dr. Matthew Davis*
E. coli	DH1Δ5_2406_pcnB(R149L)	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC pcnB*(R194L)	2806	*This study*
E. coli	DH1Δ5_2406_rpoC(M466L)	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC rpoC*(M466L)	2807	*This study*
E. coli	DH1Δ5_2406_pcnB(R149L)_rpoC(M466L)	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC pcnB*(R194L) *rpoC*(M466L)	2809	*This study*
E. coli	DH1Δ5_2403_pcnB(G141A)	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC pcnB*(G141A)	2880	*This study*
E. coli	DH1Δ5_2403_+TGG_pntA/B	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC* *+TGG_pntA/B*	2876	*This study*
E. coli	HB evolved strain*	DH1 Δ*ackA-pta* Δ*adhE* Δ*ldhA* Δ*poxB* Δ*frdBC pcnB*(G141A) *+TGG_pntA/B (HB evolved strain 2403 without plasmids)*	2883	*This study*
E. coli	DH1Δ5.cadB(stop41R(pseudogene))TGA→AGA	cadB(stop41R(pseudogene))TGA→AGA	3104	*This study*
E. coli	DH1Δ5.pspE(S14P)TCA→CCA	pspE(S14P)TCA→CCA	3103	*This study*
E. coli	DH1Δ5.pyrG(D42E)GAT→GAA	pyrG(D42E)GAT→GAA	3102	*This study*
E. coli	DH1Δ5.pnp(ΔGDISEFAPR)	pnp(ΔGDISEFAPR(546-554))ΔGGCGATATCTCTGAGTTCGCACCGCGT(1636-1662 nt)	3101	*This study*
E. coli	DH1Δ5.pnp(I154N, A153T+)	pnp(I154N, A153T+),ATC→AAC, GCG→ACG+	3100	*This study*
E. coli	BW25113Δ5.rne(R488H, V489L+)	rne(R488H, V489L+)CGC→CAC, CGC→CAC+	3099	*This study*
E. coli	DH1Δ5.rne(K255N)	rne(K255N)AAA→AAC	3098	*This study*
E. coli	DH1Δ5.rne(R374S)	rne(R374S)CGT→AGT	3097	*This study*
E. coli	BW25113Δ5.pcnB*	pcnB(frame shift after D391, total 454aa), ΔG (1176 nt)	3096	*This study*
E. coli	BW25113Δ5.pcnB(N138H)	pcnB(N138H)AAC→CAC	3095	*This study*

E. coli	BW25113Δ5.pcnB(E108A)	pcnB(E108A)GAA→GCA	3094	*This study*
E. coli	DH1Δ5.pcnB(R149P)	pcnB(R149P)CGC→CCA	3093	*This study*
E. coli	DH1Δ5.pcnB(D194E)	pcnB(D194E)GAT→GAG	3092	*This study*
E. coli	DH1Δ5.pcnB(L208W)	pcnB(L208W) TTG→TGG	3091	*This study*
E. coli	DH1Δ5.pcnB(P68T)	pcnB(P68T), CCT→ACT	3090	*This study*
E. coli	DH1Δ5.rpoC(K1192E)	rpoC(K1192E),AAA→GAA	3089	*This study*
E. coli	DH1Δ5.rpoC(G1161R)	rpoC(G1161R), GGT→CGT	3088	*This study*
E. coli	DH1Δ5.rpoC(ΔKKLTKR(215-220)	rpoC(ΔKKLTKR(215-220)), ΔAAAAGCTGACCAAGCGTA(644-661 nt)	3087	*This study*
E. coli	DH1Δ5.rpoB(G467V)	rpoB(G467V) GGC→GTC	3086	*This study*

Appendix 3.2: Plasmids used for production and strain construction

Plasmid	Selection / Origin	Description	Number	Sources
pBT33-phaA.phaB.phaC	Cm; p15a	The phaA.phaB.phaC operon was driven by the arabinose promoter	2692	*Dr. Joseph Gallagher*
pT533-phaA.phaB	Cm; p15a	The phaA.phaB.operon was driven by the T5 promoter	1319	*Dr. Matt Davis*
pT533-phaA.HBD	Cm; p15a	The phaA.hbd.operon was driven by the T5 promoter	1318	*Dr. Matt Davis*
pX_Ter.tesB	Cb; ColE1, 2u, Leu2D	TesB was cloned into a yeast shuttle vector.	2717	*Dr. Zhen Wang*
pAM45	Cb; ColE1,	Trc promoter	139	*J. Keasling Lab*
pTrc-sADS	Cm; p15a	lacUV5	122	*J. Keasling Lab*
pTargetF	Km; pMB1	For the expression of specific guide	2637	*Jiang et. al.*
pCas	Sp; RepA101ts	Cas9 from S. pyogenes MGAS5005; Lambda Red recombinase	2636	*Jiang et. al.*
pCRISPR-Gibson1	Km; ColE1	Derived from pCRISPR, XmaI and SacI cutsites were introduced between the promoter and sgRNA for guide insertion	2786	*This study*
pCRISPR-PcnB2409	Km; ColE1	Express guide target for the pcnB locus	2784	*This study*
pCRISPR-RpoC2406	Km; ColE1	Express guide target for the rpoC locus	2794	*This study*
pCRISPR_gibson_1 guide_2403g2NADP	Km; ColE1	Express guide target for the upstream sequence of pntA/B	2938	*This study*
pCRISPR_Tet_g1Km	Tc; ColE1	Express guide target Km resistant gene	2935	*This study*
pCRISPR_Tet_g3Cb	Tc; ColE1	Express guide target Cb resistant gene	2936	*This study*
pCRISPR_Tet_g1Cm	Tc; ColE1	Express guide target Cm resistant gene	2937	*This study*
pCRISPR_Tet	Tc; ColE1	For the expression of specific guide	2792	*This study*
pKD46-Cas9-RecA-Cure_Sp	Sp; RepA101ts	Cas9 from S. pyogenes	2811	*This study*

Appendix 3.3: Oligos used for plasmid and strain construction

All guide sequences are highlighted in grey. The "*" in the repair fragments indicates the phosphorothioate bond modification.

Name	Sequence
P1151_pCRISPR_gib_guideF	Ataccgctcgccgcagccgaacgccctaggtctagggcggcggatttgtc
P1141*_pCRISPR_gibson_2R	gctgttttgaatggtcccaaaacCCGCGGAAGCTTgttttagagctatgctgttttgaatggtc
P1141_pCRISPR_gibson_3F	gctgttttgaatggtcccaaaacCCGCGGAAGCTTgttttagagctatgctgttttgaatggtc
P1142pCRISPR_gibson_3R	attcaaaacagcatagctctaaaacTCTAGAgttttgggaccattcaaaacagc
P1138_pCRISPR_gibson_1F	atgctgttttgaatggtcccaaaacTCTAGAgttttagagctatgctgttttgaatggtc
P1152_pCRISPR_gib_guideR	Gaggccctttcgtcttcacctcgagtccctatcagtgatagagattgacatcc
P1156_pCRISPR_2409_pcnB_R	aaacagcatagctctaaaacCTACGCTGTAATACAGGCTGgttttgggaccattcaaaac
P1155_pCRISPR_2409_pcnB_F	gttttgaatggtcccaaaacCAGCCTGTATTACAGCGTAGgttttagagctatgctgttt
P1233_g2rpoC_R	aaacagcatagctctaaaacCGGCGAACGGCGAACCAATCgttttgggaccattcaaaac
P1232_g2rpoC_F	gttttgaatggtcccaaaacGATTGGTTCGCCGTTCGCCGgttttagagctatgctgttt
P1257_g1Km_R3	gcatagctctaaaacCCGCATTGCATCAGCCATGAgttttgggaccattc
P1256_g1Km_F3	gaatggtcccaaaacTCATGGCTGATGCAATGCGGgttttagagctatgc
P1255_g3Cb_Tc_R	ggaccattcaaaacagcatagctctaaaacTCGTGTAGATAACTACGATAgttttgggaccattcaaaacagcatagctc
P1254_g3Cb_Tc_F	gagctatgctgttttgaatggtcccaaaacTATCGTAGTTATCTACACGAgttttagagctatgctgttttgaatggtcc
P1274_g1Cm_Tc_R	aaacagcatagctctaaaacTTGGGATATATCAACGGTGGgttttgggaccattcaaaac
P1273_g1Cm_Tc_F	gttttgaatggtcccaaaacCCACCGTTGATATATCCCAAgttttagagctatgctgttt
P1269_g2NADPH_R	aaacagcatagctctaaaacTCGCCTTGCGCAAACCAGGTgttttgggaccattcaaaac
P1268_g2NADPH_F	gttttgaatggtcccaaaacACCTGGTTTGCGCAAGGCGAgttttagagctatgctgttt
P1227_2406_pcnB RF_R	A*CGTAATCACGGACGGTAAAATCtGCTACGCTGTAATACAGGCTGTTGATAGTGAAATCGCGGaGCTGGGCGTCTTCTTCGATGGAGCCGAAAATGT*T
P1226_2406_pcnB RF_F	A*ACATTTTCGGCTCCATCGAAGAAGACGCCCAGCtCCGCGATTTCACTATCAACAGCCTGTATTACAGCGTAGCaGATTTTACCGTCCGTGATTACG*T
P1230_2406_rpoC_RF_F	C*TTCGATGGTGACCAGcTGGCTGTTCACGTACCGCTGACGCTGGAAGCCCAGCTGGAAGCGCGTGCGCTGATGATGTCTACCAACAACATCCTGTCaCCGGCGAACGGCGAACCAATCATCGTTCCGTCTCAGG*A
P1231_2406_rpoC_RF_R	T*CCTGAGACGGAACGATGATTGGTTCGCCGTTCGCCGGtGACAGGATGTTGTTGGTAGACATCATCAGCGCACGCGCTTCCAGCTGGGCTTCCAGCGTCAGCGGTACGTGAACAGCCAgCTGGTCACCATCGAA*G
P1258_2403_pcnB_RF	G*CTGCGCGACAACATTTTCGcCTCCATCGAAGAAGACGCCCAGCGCCGCGATTTCACTATCAACAGCCTGTATTACAGCGTAGCaGATTTTACCGTCCGTGATTACGTT*G
P1275_2403_pcnB mutant RF_R	C*AACGTAATCACGGACGGTAAAATCtGCTACGCTGTAATACAGGCTGTTGATAGTGAAATCGCGGCGCTGGGCGTCTTCTTCGATGGAGgCGAAAATGTTGTCGCGCAG*C
P1267_2403_NADPH transhydrogenase RF	G*TTTCTCGTTAATAACAATACCAccaGTACCTGGTTTGCGCAAGGCGAAaGATTATTTTTATGAAGCTTAAGAACACCCTCCTGGCGTC*G
P1276_2403_NADPH transhydrogenase RF_R	G*TTTCTCGTTAATAACAATACCAccaGTACCTGGTTTGCGCAAGGCGAAaGATTATTTTTATGAAGCTTAAGAACACCCTCCTGGCGTC*G

The following primers were used to construct point mutations and indels arose from evolution using the Cas9 system described in Jiang *et al*. [6] The targeting vectors were constructed from pTargetF vector by reverse PCR using 459 and different -target primers, and subsequent self-ligation. The repair fragments were generated by primer pairs -1 & -2 and -3 & -4 using *E. coli* 799 genomic DNA as template, and subsequent SOE-PCR for fusion of above two fragments. The following primers were used for these studies.

Primer	Sequence
459-pTargetF-F2	ACTAGTATTATACCTAGGACTGAGCTAGCTGTCAAG
335-V2-target	TCCTAGGTATAATACTAGTGCGGAAAACCAGTTCCGCGTGTTTTAGAGCTAGAAATAGC
337-V4-target	TCCTAGGTATAATACTAGTGCAATCCTGGCTGAAATCAGGTTTTAGAGCTAGAAATAGC
338-V5-target	TCCTAGGTATAATACTAGTCGGATCGCTACCGTCTACCGGTTTTAGAGCTAGAAATAGC
339-V6-target	TCCTAGGTATAATACTAGTAACAGTTTGCGCACCTGCTCGTTTTAGAGCTAGAAATAGC
340-V7-target	TCCTAGGTATAATACTAGTGCTGATGCGCATACCCAATTGTTTTAGAGCTAGAAATAGC
341-V8-target	TCCTAGGTATAATACTAGTAACGCGCTACCGTGAAGATCGTTTTAGAGCTAGAAATAGC
343-V10-target	TCCTAGGTATAATACTAGTGGCTGTTGATAGTGAAATCGGTTTTAGAGCTAGAAATAGC
344-V11-target	TCCTAGGTATAATACTAGTGGCTCATGTAATGTTTGGCCGTTTTAGAGCTAGAAATAGC
345-V12-target	TCCTAGGTATAATACTAGTGCGCTGGGCGTCTTCTTCGAGTTTTAGAGCTAGAAATAGC
346-V13-target	TCCTAGGTATAATACTAGTGGAGCATCCTAAGTTCCGTGGTTTTAGAGCTAGAAATAGC
347-V14-target	TCCTAGGTATAATACTAGTTCTGCGTGAAGCGGTGCGTCGTTTTAGAGCTAGAAATAGC
348-V15-target	TCCTAGGTATAATACTAGTTATCGCTGCATTAGGTCGCCGTTTTAGAGCTAGAAATAGC
350-V17-target	TCCTAGGTATAATACTAGTCGTGCTGCGCGTGCGTAAAGGTTTTAGAGCTAGAAATAGC
351-V18-target	TCCTAGGTATAATACTAGTTATCCTGGGCGTAATGGAACGTTTTAGAGCTAGAAATAGC
352-V19-target	TCCTAGGTATAATACTAGTCAGGCGATCAACGCGCCGCGGTTTTAGAGCTAGAAATAGC
354-V21-target	TCCTAGGTATAATACTAGTCAATGTGACCATCATGAAACGTTTTAGAGCTAGAAATAGC
355-V22-target	TCCTAGGTATAATACTAGTAGGCTTACTTGCTCTGGCACGTTTTAGAGCTAGAAATAGC
358-V25-target	TCCTAGGTATAATACTAGTCAATACCACCGATACTTGCTGTTTTAGAGCTAGAAATAGC
363-V2-1	AGTTACCAGGTCTTCTACGAAGTGGCCTTC
364-V2-2	GAAATGGCGGAAAACCAGTTCCGCGTTGTCCTGGTACGTGTAGAGCGTGCGGTGAAAG
365-V2-3	CTTTCACCGCACGCTCTACACGTACCAGGACAACGCGGAACTGGTTTTCCGCCATTTC
366-V2-4	GACTATATTGATGAGTCTACCGGCGAGCTG
371-V4-1	GATGAAGGACTCGGTTGCCAGAGACGCTTT
372-V4-2	CCGGCAATCCTGGCTGAAATCAGCCGTATCGTTTCCTTCGGTAAAGAAACCAAAG
373-V4-3	CTTTGGTTTCTTTACCGAAGGAAACGATACGGCTGATTTCAGCCAGGATTGCCGG
374-V4-4	GGCGAAAGGCGATGGCGAACAGGTTGCTGG
375-V5-1	GGATCAGACGACCCACGATAACGTTCTCTT
376-V5-2	GGTTATCACGCCGGTAGACGGTAGCGATCCGTACGAAGAGATGATTCCGGAATGGCGTC
377-V5-3	GACGCCATTCCGGAATCATCTCTTCGTACGGATCGCTACCGTCTACCGGCGTGATAACC
378-V5-4	CGCACACCATGCCGGTTATCACCGAAGTAA
379-V6-1	CGATCCACGCCCGGTACTAATCGCGGCAC
380-V6-2	GTTACGGAACAGTTTGCGCACCTGCTCAGTCGTGGCGTTAGTGGTTACGTCAAAATCT
381-V6-3	AGATTTTGACGTAACCACTAACGCCACGACTGAGCAGGTGCGCAAACTGTTCCGTAAC
382-V6-4	GTAGCCCGCTTGTAGCAGTTTAAGCGATTC
383-V7-1	CAGGCTCAATAAAGCGGGATACGAAGCCTG
384-V7-2	CGGTTTCCGGGCTGATGCGCATACCCCATTTCGCGGCAAAACGTACCGCGCGCAGC
385-V7-3	GCTGCGCGCGGTACGTTTTGCCGCGAAATGGGGTATGCGCATCAGCCCGGAAACCG
386-V7-4	CATGCGCGTTTACCCTGACGACGGGACATAC
387-V8-1	GAGCAGCATGCTATTTCCCGCAAAGATATC
388-V8-2	CGTACCGCGCGCAGCATACGTACTGGCTCTTCACGGTAGCGCGTTTCCGGGTTACCA
389-V8-3	TGGTAACCCGGAAACGCGCTACCGTGAAGAGCCAGTACGTATGCTGCGCGCGGTACG
390-V8-4	GGTTAATGTCGTCAGACGTTTCGGGATTGCC
395-V10-1	GAGGTGTACTATTTTTACCCGAGTCGCTAA
396-V10-2	AATACAGGCTGTTGATAGTGAAATCGCGTGGCTGGGCGTCTTCTTCGATGGAGCCGAAA
397-V10-3	TTTCGGCTCCATCGAAGAAGACGCCCAGCCACGCGATTTCACTATCAACAGCCTGTATT
398-V10-4	ATCTTCTGTGCCGTCTCCAGCAGTGGGTAC
399-V11-1	CACACTGGCAGGATTTCAGCGTCGAGCAAA

400-V11-2	CGTGGTGTCCACGGAAGGTCGCAACTGCGATAATCTCTGGGCCAAACATTACATGA
401-V11-3	TCATGTAATGTTTGGCCCAGAGATTATCGCAGTTGCGACCTTCCGTGGACACCACG
402-V11-4	CAATGATCCGCTCCATCGGGCTGTCGCCAT
403-V12-1	GTCCTGAATGATGTTTGACACTACCGAGGTG
404-V12-2	CGTCTTCTTCGATAGAGCCGAAAATGTGGTCGCGCAGCAACATGCCGTTTTGCCCG
405-V12-3	CGGGCAAAACGGCATGTTGCTGCGCGACCACATTTTCGGCTCTATCGAAGAAGACG
406-V12-4	CGGGTTCACGCGCATATCGTTATGGATACG
407-V13-1	ATATCCCACCGGCACGCCTGTTTGAAGAATC
408-V13-2	TTACGCTCAACTTCAGCTCGCAAGGCCAAAGGTCATAAGCGGCACGGAACTTAGGATG
409-V13-3	CATCCTAAGTTCCGTGCCGCTTATGACCTTTGGCCTTGCGAGCTGAAGTTGAGCGTAA
410-V13-4	CGTTCAGTATTTATCACTTCATTACCAAACAG
477-V14-1new	CACCCGCATTGAACCGAGTCTGGAAGCTGCTTTTG
478-V14-2new	GAGAAATATGGCTGATTTGAATACTCGCACGGTCTTGGCGAACGGCTTCACGCAGACG
479-V14-3new	CGTCTGCGTGAAGCCGTTCGCCAAGACCGTGCGAGTATTCAAATCAGCCATATTTCTC
480-V14-4new	CTTGTTGCGCCTGACGTTTATCATCATTACGGCGGC
415-V15-1	ATTTCCCTGCTAACTACAGTGCTCATGGTCGTCC
416-V15-2	GATCTCGCCGGTGTACAGTTTGATGTTGCTGCTGAAATCTGGGCGACCTAATGCAG
417-V15-3	CTGCATTAGGTCGCCCAGATTTCAGCAGCAACATCAAACTGTACACCGGCGAGATC
418-V15-4	CGCTTCTTCTTCGATCAGACGCAGAATAGAGAGC
422-V16-4	TCACGCTGGAAGGCGGACTCGATCTG
485-V17-1new	GGCATTTCTCGCCGTATCGAAGGCGACGACC
486-V17-2new	CTTAAGGTTGGGGTTTCTTCACCTTTGCGAAGGTGCAGCACGTGGTAGTGCGGGGTTTC
487-V17-3new	GAAACCCCGCACTACCACGTGCTGCACCTTCGCAAAGGTGAAGAAACCCCAACCTTAAG
488-V17-4new	ACTTTGCCAGAGGCCAGTTCCGGAGACGC
427-V18-1	TTCCTGTTCCACTACAACTTCCCTCCGTACTC
428-V18-2	GTGCGAACTCAGAGATATCGCCACGCGGCGCGTTGTTCGCTTGTTCCATTACGCCCAGG
429-V18-3	CCTGGGCGTAATGGAACAAGCGAACAACGCGCCGCGTGGCGATATCTCTGAGTTCGCAC
430-V18-4	CTCGCCCTGTTCAGCAGCCGGAG
431-V19-1	TTCCTGTTCCACTACAACTTCCCTCCGTACTCC
432-V19-2	GATCTTGTCCGGGTTGATCTTGATGGTATGGATACGCGGCGCGTTGATCGCCTGTTCC
433-V19-3	GGAACAGGCGATCAACGCGCCGCGTATCCATACCATCAAGATCAACCCGGACAAGATC
434-V19-4	CTCGCCCTGTTCAGCAGCCGGAGCTTC
439-V21-1	AAAGCGAACGAAAAATTCGAGCGTCGTTTTC
440-V21-2	GTACCTGGATCGACGTTGATGTACGGTTCAAGTTTCATGATGGTCACATTGAGGCCAC
441-V21-3	GTGGCCTCAATGTGACCATCATGAAACTTGAACCGTACATCAACGTCGATCCAGGTAC
442-V21-4	GATCTGAACGACAAATCAGGATGTCAGGCTGG
443-V22-1	TCCACCGTAGATTTCGTCCAGGTAATCCG
444-V22-2	GGCTTACTTGCTCTGGCACTTGTGTTTCCACTGCCCGTTTTCGCCGCTGAACACTG
445-V22-3	CAGTGTTCAGCGGCGAAAACGGGCAGTGGAAACACAAGTGCCAGAGCAAGTAAGCC
446-V22-4	TCATTTTGTCATTTGCGCTTGATCCAATGCC
455-V25-1	TGTTCCTTATGTTGTACCTTATCTCGACAAATTTC
456-V25-2	GATAATCCAACCTCTGATAGCAATACCACCGATACTTGCTAGATTCGCAGGTAATAATG
457-V25-3	CATTATTACCTGCGAATCTAGCAAGTATCGGTGGTATTGCTATCAGAGGTTGGATTATC
458-V25-4	TTTCGGGTTTTTAACCATACCAGTACTTACAGCTGC

Appendix 3.4: RNA-Seq data for parent strains and evolved *E. coli* strains

After the Sleuth analysis, data was then filtered by the p value < 0.05. Data is then further filtered with the ß ≥ 2 and ≤2 to obtained the up-regulated and down-regulated data set respectively (n=3).

A1. BDO parent strain and BDO evolved strain 2406 (Up-regulated).

target_id	pval	qval	b	se_b	mean_obs	var_obs	tech_var	sigma_sq	smooth_sigma_sq	final_sigma_sq	K12 ID (uniprot)	gene name	annotation
ECDH1_RS10835	0	0	7.72	0.12	13.00	17.89	0.00	0.00	0.02	0.02	P07001	pntA	NAD(P) transhydrogenase subunit alpha
ECDH1_RS10840	0	0	7.59	0.13	13.01	17.31	0.00	0.02	0.02	0.02	P0AB67	pntB	NAD(P) transhydrogenase subunit beta
ECDH1_RS10670	6.86E-284	7.71E-281	4.14	0.12	12.25	5.16	0.00	0.01	0.02	0.02	P77304	dtpA	Dipeptide and tripeptide permease A
ECDH1_RS11215	2.59E-112	1.06E-109	2.82	0.13	9.46	2.40	0.01	0.02	0.02	0.02	P31122	sotB	Sugar efflux transporter
ECDH1_RS12685	6.76E-89	2.53E-86	2.42	0.12	9.81	1.76	0.00	0.00	0.02	0.02	P0A710	yciB	Probable intracellular septation protein A
ECDH1_RS16615	1.06E-87	3.39E-85	3.10	0.16	8.49	2.90	0.01	0.00	0.02	0.02	P77307	fetB	Probable iron export permease protein FetB
ECDH1_RS12595	1.02E-85	2.88E-83	2.31	0.12	9.91	1.61	0.00	0.01	0.02	0.02	P0AG14	sohB	Probable protease SohB
ECDH1_RS15275	1.02E-85	2.88E-83	4.66	0.24	9.55	6.59	0.01	0.07	0.02	0.07	P37329	modA	Molybdate-binding periplasmic protein
ECDH1_RS04460	5.17E-83	1.37E-80	2.61	0.14	9.05	2.05	0.01	0.01	0.02	0.02	P63340	yqeG	Inner membrane

													transport protein YqeG
ECDH1_RS04065	6.75E-80	1.55E-77	2.16	0.11	10.26	1.41	0.00	0.01	0.02	0.02	P0C0S1	mscS	Small-conductance mechanosensitive channel
ECDH1_RS16680	1.57E-76	3.36E-74	2.12	0.11	10.13	1.35	0.00	0.01	0.02	0.02	P39830	ybaL	Inner membrane protein YbaL
ECDH1_RS15265	3.59E-76	7.34E-74	3.31	0.18	9.40	3.33	0.01	0.04	0.02	0.04	P09833	modC	Molybdenum import ATP-binding protein ModC
ECDH1_RS08445	1.02E-75	1.99E-73	2.15	0.12	9.84	1.40	0.00	0.01	0.02	0.02	P76389	yegH	UPF0053 protein YegH
ECDH1_RS16625	9.52E-73	1.78E-70	2.09	0.12	10.16	1.32	0.00	0.00	0.02	0.02	P0AA53	qmcA	Protein QmcA
modB	1.21E-58	1.81E-56	4.08	0.25	9.02	5.07	0.01	0.09	0.02	0.09	P0AF01	modB	Molybdenum transport system permease protein ModB
ECDH1_RS13735	1.02E-55	1.39E-53	2.01	0.13	9.14	1.21	0.01	0.00	0.02	0.02	P0ACV0	lpxL	Lipid A biosynthesis lauroyltransferase
ECDH1_RS09120	1.65E-52	2.12E-50	2.28	0.15	9.71	1.59	0.00	0.03	0.02	0.03	P31064	yedE	UPF0394 inner membrane protein YedE
ECDH1_RS10070	2.91E-47	2.72E-45	2.50	0.17	7.89	1.89	0.02	0.00	0.03	0.03	P76219	ydjX	TVP38/TMEM64 family membrane protein YdjX
ECDH1_RS10850	4.99E-46	4.49E-44	2.42	0.17	7.95	1.79	0.02	0.02	0.03	0.03	P69212	mdtJ	Spermidine export protein MdtJ
ECDH1_RS09815	4.92E-45	4.02E-43	2.81	0.20	7.34	2.38	0.03	-0.01	0.03	0.03	P76249	leuE	Leucine efflux protein

ECDH1_RS17655	1.57E-44	1.26E-42	2.18	0.16	8.54	1.45	0.01	0.03	0.02	0.03	P0AAA1	yagU	Inner membrane protein YagU
ECDH1_RS10065	1.03E-43	8.02E-42	2.17	0.16	8.40	1.44	0.01	0.03	0.02	0.03	P76220	ydjY	Uncharacterized protein YdjY
ECDH1_RS07650	4.99E-43	3.74E-41	2.03	0.15	13.59	1.26	0.00	0.03	0.02	0.03	P06996	ompC	Outer membrane protein C
ECDH1_RS19475	1.65E-37	9.75E-36	2.28	0.18	9.24	1.60	0.01	0.04	0.02	0.04	P39386	mdtM	Multidrug resistance protein MdtM
ECDH1_RS11210	8.53E-35	4.51E-33	2.28	0.19	7.55	1.58	0.02	0.00	0.03	0.03	P0AEY1	marC	UPF0056 inner membrane protein MarC
ECDH1_RS05320	1.48E-31	6.87E-30	2.28	0.20	9.59	1.61	0.00	0.05	0.02	0.05	P14175	proV	Glycine betaine/proline betaine transport system ATP-binding protein ProV
ECDH1_RS12250	4.11E-30	1.76E-28	2.19	0.19	7.47	1.48	0.02	0.04	0.03	0.04	P0AEB5	ynaI	Low conductance mechanosensitive channel YnaI
ECDH1_RS16620	1.19E-26	4.27E-25	2.59	0.24	8.34	2.09	0.02	0.07	0.02	0.07	P77279	fetA	Probable iron export ATP-binding protein FetA
ECDH1_RS05315	2.14E-26	7.51E-25	2.48	0.23	9.40	1.91	0.01	0.08	0.02	0.08	P14176	proW	Glycine betaine/proline betaine transport system permease protein ProW
ECDH1_RS07315	4.88E-26	1.65E-24	2.09	0.20	11.90	1.35	0.00	0.06	0.02	0.06	P0AFE8	nuoM	NADH-quinone oxidoreductase subunit M
ECDH1_RS07320	2.89E-25	9.20E-24	2.02	0.19	11.76	1.26	0.00	0.06	0.02	0.06	P0AFF0	nuoN	NADH-quinone oxidoreductase subunit N

ECDH1_RS10060	1.44E-23	4.01E-22	2.03	0.20	7.98	1.29	0.01	0.05	0.03	0.05	P76221	ydjZ	TVP38/TMEM64 family inner membrane protein YdjZ
ECDH1_RS16640	2.17E-23	5.83E-22	2.19	0.22	10.83	1.50	0.00	0.07	0.02	0.07	P77400	ybaT	Inner membrane transport protein YbaT
ECDH1_RS05310	4.39E-23	1.15E-21	2.63	0.27	9.28	2.16	0.01	0.10	0.02	0.10	P0AFM2	proX	Glycine betaine/proline betaine-binding periplasmic protein
tqsA	3.59E-22	8.75E-21	2.56	0.26	7.88	2.06	0.02	0.09	0.03	0.09	P0AFS5	tqsA	AI-2 transport protein TqsA
ECDH1_RS07425	2.59E-18	4.58E-17	2.16	0.25	6.98	1.47	0.04	0.06	0.04	0.06	P76472	arnD	Probable 4-deoxy-4-formamido-L-arabinose-phosphoundecaprenol deformylase ArnD
ECDH1_RS12075	6.86E-14	8.57E-13	2.15	0.29	7.63	1.48	0.02	0.10	0.03	0.10	P23849	trkG	Trk system potassium uptake protein TrkG
ECDH1_RS16515	1.63E-13	1.94E-12	2.28	0.31	5.71	1.62	0.08	0.00	0.06	0.06	P77328	ybbY	Putative purine permease YbbY
ECDH1_RS05480	1.26E-12	1.41E-11	2.97	0.42	5.45	2.85	0.15	0.11	0.07	0.11	P52138	yfjW	Uncharacterized protein YfjW
ECDH1_RS14010	1.68E-12	1.84E-11	2.30	0.33	5.46	1.65	0.09	0.00	0.07	0.07	P56614	ymdF	Uncharacterized protein YmdF
ECDH1_RS15995	7.56E-11	6.88E-10	2.37	0.36	8.72	1.84	0.01	0.19	0.02	0.19	P37002	crcB	Putative fluoride ion transporter CrcB
ECDH1_RS13310	1.75E-10	1.55E-09	2.43	0.38	5.02	1.92	0.13	0.06	0.09	0.09	P75968	ymfE	Uncharacterized protein YmfE

ECDH1_RS03685	3.52E-10	2.96E-09	3.87	0.62	6.91	4.96	0.07	0.50	0.04	0.50	P64574	yghW	Uncharacterized protein YghW
ECDH1_RS11535	3.15E-06	1.60E-05	2.07	0.44	5.06	1.52	0.12	0.18	0.09	0.18	P19317	narW	Probable nitrate reductase molybdenum cofactor assembly chaperone NarW
ECDH1_RS19850	3.40E-06	1.72E-05	2.60	0.56	4.28	2.26	0.33	-0.04	0.14	0.14	P39352	yjhB	Putative metabolite transport protein YjhB
ECDH1_RS11405	7.98E-05	0.000329591	2.17	0.55	13.06	1.78	0.00	0.46	0.02	0.46	P63235	gadC	Probable glutamate/gamma-aminobutyrate antiporter
ECDH1_RS11145	0.001689779	0.005136544	2.10	0.67	9.72	1.87	0.00	0.67	0.02	0.67	P64463	ydfZ	Putative selenoprotein YdfZ

A2. BDO parent strain and BDO evolved strain 2406 (Down-regulated)

target_id	pval	qval	b	se_b	mean_obs	var_obs	tech_var	sigma_sq	smooth_sigma_sq	final_sigma_sq	K12 ID (uniprot)	gene name	annotation
ECDH1_RS19600	0	0	-7.05	0.18	8.38	14.92	0.03	-0.02	0.02	0.02	P30130	fimD	Outer membrane usher protein FimD
ECDH1_RS00090	1.69E-169	1.52E-166	-5.95	0.21	9.73	10.66	0.01	0.06	0.02	0.06	P0C058	ibpB	Small heat shock protein IbpB
ECDH1_RS19605	3.12E-163	2.34E-160	-7.42	0.27	7.60	16.60	0.06	0.06	0.03	0.06	P31697	fimC	Chaperone protein FimC
fimA	1.10E-123	7.08E-121	-7.45	0.32	11.86	16.77	0.00	0.15	0.02	0.15	P04128	fimA	Type-1 fimbrial protein, A chain
ECDH1_RS19610	3.08E-121	1.73E-118	-8.05	0.34	8.76	19.57	0.03	0.15	0.02	0.15	P39264	fimI	Fimbrin-like protein FimI
ECDH1_RS03625	2.10E-116	1.05E-113	-4.30	0.19	12.49	5.58	0.00	0.05	0.02	0.05	Q46856	yqhD	Alcohol dehydrogenase YqhD
ECDH1_RS03635	1.72E-113	7.75E-111	-3.42	0.15	9.34	3.53	0.00	0.03	0.02	0.03	Q46855	yqhC	Uncharacterized HTH-type transcriptional regulator YqhC
ECDH1_RS19585	1.43E-88	4.93E-86	-4.11	0.21	8.10	5.13	0.01	0.05	0.03	0.05	P08191	fimH	Protein FimH
ECDH1_RS03330	6.02E-81	1.50E-78	-2.10	0.11	10.95	1.33	0.00	0.01	0.02	0.02	P0ABS5	dnaG	DNA primase
ECDH1_RS01810	6.89E-80	1.55E-77	-2.08	0.11	10.64	1.32	0.00	0.01	0.02	0.02	P08201	nirB	Nitrite reductase

													(NADH) large subunit
ECDH1_RS24725	2.52E-62	3.91E-60	-2.34	0.14	8.11	1.67	0.00	0.03	0.03	0.03	P0CF11	insA5	Insertion element IS15 protein InsA
ECDH1_RS08770	4.43E-56	6.43E-54	-2.51	0.16	9.44	1.92	0.00	0.03	0.02	0.03	P39180	flu	Antigen 43
ECDH1_RS00375	2.54E-49	2.72E-47	-2.58	0.17	7.39	2.00	0.01	-0.01	0.03	0.03	P05523	mutM	Formamidopyrimidine-DNA glycosylase
ECDH1_RS10595	1.35E-42	9.75E-41	-2.89	0.21	10.26	2.56	0.00	0.06	0.02	0.06	P67430	nemR	HTH-type transcriptional repressor NemR
lacI	6.96E-41	4.81E-39	-2.80	0.21	18.76	2.39	0.00	0.05	0.07	0.07	P03023	lacI	Lactose operon repressor
ECDH1_RS19590	4.52E-39	2.86E-37	-5.76	0.44	6.68	10.20	0.05	0.25	0.04	0.25	P08190	fimG	Protein FimG
ECDH1_RS13440	5.67E-36	3.23E-34	-2.58	0.21	6.75	2.02	0.02	0.01	0.04	0.04	P75954	ycfS	Probable L,D-transpeptidase YcfS
ECDH1_RS17755	1.32E-35	7.43E-34	-2.53	0.20	7.63	1.97	0.01	0.05	0.03	0.05	P75682	yagE	Probable 2-keto-3-deoxy-galactonate aldolase YagE
ECDH1_RS00085	3.13E-33	1.58E-31	-4.10	0.34	10.17	5.18	0.00	0.17	0.02	0.17	P0C054	ibpA	Small heat shock protein IbpA
ECDH1_RS14420	1.26E-30	5.50E-29	-2.84	0.25	9.85	2.49	0.00	0.09	0.02	0.09	P02931	ompF	Outer membrane protein F

ECDH1_RS13445	9.57E-29	3.88E-27	-3.48	0.31	10.35	3.75	0.00	0.14	0.02	0.14	P0AB40	bhsA	Multiple stress resistance protein BhsA
ECDH1_RS21785	2.81E-28	1.10E-26	-2.01	0.18	7.19	1.23	0.02	0.00	0.03	0.03	P0A712	kdgT	2-keto-3-deoxygluconate permease
ECDH1_RS10585	8.80E-24	2.54E-22	-2.29	0.23	10.85	1.64	0.00	0.08	0.02	0.08	P0AC81	gloA	Lactoylglutathione lyase
ECDH1_RS10590	5.13E-23	1.30E-21	-2.52	0.25	12.61	1.98	0.00	0.10	0.02	0.10	P77258	nemA	N-ethylmaleimide reductase
ECDH1_RS03620	8.28E-21	1.79E-19	-2.12	0.23	11.68	1.42	0.00	0.08	0.02	0.08	Q46857	dkgA	2,5-diketo-D-gluconic acid reductase A
ECDH1_RS22625	1.33E-20	2.77E-19	-2.62	0.28	10.83	2.16	0.00	0.12	0.02	0.12	P00963	asnA	Aspartate--ammonia ligase
ECDH1_RS19580	7.74E-20	1.54E-18	-2.31	0.25	5.92	1.62	0.04	-0.02	0.06	0.06	P0AC94	gntP	High-affinity gluconate transporter
ECDH1_RS19245	1.13E-19	2.22E-18	-3.63	0.40	11.25	4.15	0.00	0.24	0.02	0.24	P0A6L0	deoC	Deoxyribose-phosphate aldolase
ECDH1_RS19240	5.32E-19	9.93E-18	-3.32	0.37	11.76	3.48	0.00	0.21	0.02	0.21	P07650	deoA	Thymidine phosphorylase
ECDH1_RS19595	1.14E-18	2.07E-17	-6.86	0.78	6.24	14.83	0.15	0.75	0.05	0.75	P08189	fimF	Protein FimF
ECDH1_RS20225	9.26E-17	1.48E-15	-3.30	0.40	4.59	3.29	0.12	-0.09	0.12	0.12	P39308	yjfZ	Uncharacterized protein YjfZ

ECDH1_RS08305	6.18E-16	9.52E-15	-2.37	0.29	5.43	1.70	0.05	-0.02	0.07	0.07	P0A9S3	gatD	Galactitol-1-phosphate 5-dehydrogenase
ECDH1_RS04370	1.29E-15	1.91E-14	-2.11	0.26	5.76	1.41	0.04	0.06	0.06	0.06	Q46797	ygeQ	Putative uncharacterized protein YgeQ
ECDH1_RS16705	3.44E-15	4.96E-14	-2.08	0.26	11.87	1.38	0.00	0.10	0.02	0.10	P0A6Z3	htpG	Chaperone protein HtpG
ECDH1_RS22805	2.87E-14	3.66E-13	-3.25	0.43	4.27	3.19	0.13	-0.11	0.14	0.14	P0A853	tnaA	Tryptophanase
ECDH1_RS21045	2.90E-13	3.43E-12	-3.14	0.43	6.31	3.17	0.03	0.25	0.05	0.25	P32696	pspG	Phage shock protein G
ECDH1_RS19255	3.49E-12	3.71E-11	-2.21	0.32	9.55	1.59	0.00	0.15	0.02	0.15	P37342	yjjI	Uncharacterized protein YjjI
ECDH1_RS17750	5.93E-12	6.13E-11	-2.23	0.32	8.13	1.61	0.01	0.15	0.03	0.15	P77596	yagF	Uncharacterized protein YagF
ECDH1_RS20035	8.75E-12	8.84E-11	-2.44	0.36	9.88	1.95	0.00	0.19	0.02	0.19	P0ABB8	mgtA	Magnesium-transporting ATPase, P-type 1
ECDH1_RS12390	5.34E-10	4.33E-09	-2.20	0.35	10.49	1.60	0.00	0.19	0.02	0.19	P0AFM6	pspA	Phage shock protein A
ECDH1_RS06760	9.70E-10	7.68E-09	-2.31	0.38	4.54	1.66	0.09	-0.02	0.12	0.12	P77579	fryC	Fructose-like permease IIC component 1

ECDH1_RS04895	2.71E-09	2.04E-08	-2.30	0.39	6.63	1.76	0.02	0.20	0.04	0.20	P38036	ygcB	CRISPR-associated endonuclease/helicase Cas3
ECDH1_RS13925	3.30E-09	2.46E-08	-3.18	0.54	6.08	3.38	0.05	0.39	0.05	0.39	P0A9K1	phoH	Protein PhoH
ECDH1_RS00790	3.19E-08	2.13E-07	-2.73	0.49	5.14	2.53	0.07	0.30	0.09	0.30	P19768	insJ	Insertion element IS150 uncharacterized 19.7 kDa protein
ECDH1_RS19730	4.23E-08	2.76E-07	-3.65	0.67	3.17	4.44	0.35	0.20	0.32	0.32	P39360	yjhI	Uncharacterized HTH-type transcriptional regulator YjhI
ECDH1_RS14835	5.40E-08	3.48E-07	-3.54	0.65	7.60	4.27	0.02	0.62	0.03	0.62	P68688	grxA	Glutaredoxin 1
ECDH1_RS20320	4.37E-07	2.52E-06	-2.24	0.44	4.19	1.59	0.14	-0.03	0.15	0.15	P33222	yjfC	Putative acid--amine ligase YjfC
ECDH1_RS06180	6.60E-07	3.73E-06	-2.52	0.51	3.75	2.13	0.18	0.11	0.21	0.21	P65290	yfgH	Uncharacterized lipoprotein YfgH
ECDH1_RS12365	1.17E-06	6.28E-06	-2.72	0.56	3.51	2.36	0.22	-0.05	0.25	0.25	P76041	ycjM	Putative sucrose phosphorylase
ECDH1_RS02600	6.03E-05	0.0002544	-2.48	0.62	4.83	2.31	0.09	0.48	0.10	0.48	P28721	gltF	Protein GltF

ECDH1_RS03830	7.80E-05	0.00032381	-4.26	1.08	1.13	5.47	0.18	-0.15	1.56	1.56	Q46835	yghG	Uncharacterized lipoprotein YghG
ECDH1_RS00250	0.000153837	0.00059831	-2.30	0.61	3.16	1.79	0.24	0.01	0.32	0.32	P31436	setC	Sugar efflux transporter C
ECDH1_RS16100	0.000197038	0.00075047	-2.38	0.64	3.24	1.77	0.31	-0.22	0.30	0.30	P77746	ybdO	Uncharacterized HTH-type transcriptional regulator YbdO
ECDH1_RS03825	0.000271646	0.00099696	-2.14	0.59	3.41	1.59	0.25	0.03	0.26	0.26	Q46836	pppA	Leader peptidase PppA
ECDH1_RS17905	0.000353466	0.00126803	-2.49	0.70	3.45	2.45	0.24	0.49	0.26	0.49	P02932	phoE	Outer membrane pore protein E
ECDH1_RS12360	0.000403123	0.00142474	-3.55	1.00	2.75	4.99	0.20	1.31	0.44	1.31	P76042	ycjN	Putative ABC transporter periplasmic-binding protein YcjN
ECDH1_RS00785	0.000405341	0.00142959	-3.03	0.86	4.60	3.64	0.14	0.96	0.12	0.96	P19769	insK	Putative transposase InsK for insertion sequence element IS150
ECDH1_RS09395	0.000476256	0.00164649	-2.25	0.64	4.21	2.02	0.13	0.49	0.15	0.49	P52005	torY	Cytochrome c-type protein TorY

ECDH1_RS20980	0.000478174	0.00165347	-2.85	0.81	7.11	3.23	0.02	0.98	0.04	0.98	P0A9E2	soxS	Regulatory protein SoxS
ECDH1_RS19325	0.000688787	0.00229615	-2.16	0.64	4.04	1.89	0.16	0.45	0.17	0.45	P55914	yjjZ	Uncharacterized protein YjjZ
ECDH1_RS03080	0.000773195	0.00254596	-2.79	0.83	2.35	2.63	0.43	-0.06	0.60	0.60	P0AGF6	tdcB	L-threonine dehydratase catabolic TdcB
ECDH1_RS06755	0.000907616	0.00293713	-2.20	0.66	3.09	1.88	0.32	0.20	0.34	0.34	P69808	fryB	PTS system fructose-like EIIB component 1
ECDH1_RS20340	0.001093065	0.00346297	-2.05	0.63	3.19	1.58	0.28	0.11	0.31	0.31	P0AF78	yjfJ	Uncharacterized protein YjfJ
ECDH1_RS18945	0.001245093	0.00390324	-2.15	0.67	3.26	1.93	0.32	0.35	0.30	0.35	P31551	caiD	Carnitinyl-CoA dehydratase
chiP	0.001348759	0.00418497	-2.18	0.68	3.59	1.98	0.22	0.47	0.23	0.47	P75733	chiP	Chitoporin
ECDH1_RS07005	0.0040732	0.01127654 3	-2.06	0.72	2.87	1.50	0.37	-0.09	0.40	0.40	P77288	yfcV	Uncharacterized fimbrial-like protein YfcV
ECDH1_RS06420	0.005332161	0.01427837 7	-2.15	0.77	2.64	2.09	0.41	0.47	0.48	0.48	P76556	eutP	Ethanolamine utilization protein EutP
ECDH1_RS13175	0.006114148	0.01606616 4	-2.33	0.85	2.74	2.51	0.36	0.73	0.44	0.73	P75991	ycgZ	Probable two-component-system

													connector protein YcgZ
ECDH1_RS02960	0.007168837	0.01843934	-2.28	0.85	2.36	1.58	0.48	-0.45	0.60	0.60	P42914	yraI	Probable fimbrial chaperone YraI
sgbH	0.007614765	0.01948328	-2.02	0.76	2.61	1.42	0.37	-0.13	0.49	0.49	P37678	sgbH	3-keto-L-gulonate-6-phosphate decarboxylase SgbH
ECDH1_RS11835	0.009516079	0.02361164	-2.08	0.80	2.54	1.51	0.45	-0.18	0.52	0.52	P76091	ynbB	Uncharacterized protein YnbB
ECDH1_RS20325	0.009855995	0.02438775 6	-2.03	0.79	2.68	1.99	0.40	0.53	0.47	0.53	P39295	yjfM	Uncharacterized protein YjfM
ECDH1_RS17445	0.011434791	0.02768488	-2.06	0.81	2.34	1.63	0.39	0.05	0.61	0.61	P75692	yahM	Uncharacterized protein YahM
ECDH1_RS06815	0.015916688	0.03715546 6	-2.42	1.00	1.91	2.97	0.46	1.05	0.85	1.05	O32528	ypdI	Uncharacterized lipoprotein YpdI
ECDH1_RS20335	0.019152111	0.04379852	-2.15	0.92	2.96	2.40	0.31	0.95	0.37	0.95	P39293	yjfK	Uncharacterized protein YjfK
ECDH1_RS15520	0.023538524	0.05223553 9	-2.29	1.01	2.11	2.79	0.33	1.20	0.73	1.20	P37909	ybgD	Uncharacterized fimbrial-like protein YbgD
ECDH1_RS06835	0.026035 86	0.05704543 3	-2.00	0.90	2.85	2.17	0.36	0.86	0.41	0.86	P0AA49	yfdV	Uncharacterized

													transporter YfdV
ECDH1_RS01210	0.04142 6875	0.085 1647 13	-2.77	1.36	2.69	4.51	1.29	1.48	0.46	1.48	P2891 2	yhhI	H repeat-associated protein YhhI
ECDH1_RS10775	0.04520 7032	0.092 1772 42	-2.70	1.35	1.66	4.38	0.34	2.39	1.03	2.39	Q4770 6	uidC	Membrane-associated protein UidC

B1. HB parent strain and HB evolved strain 2403 (Up-regulated)

target_id	pval	qval	b	se_b	mean_obs	var_obs	tech_var	sigma_sq	smooth_sigma_sq	final_sigma_sq	K12 ID (uniprot)	gene name	annotation
ECDH1_RS10835	4.97E-296	2.28E-292	5.05	0.14	12.08	7.65	0.00	0.00	0.03	0.03	P07001	pntA	NAD(P) transhydrogenase subunit alpha
ECDH1_RS10840	3.23E-277	7.40E-274	4.81	0.14	11.93	6.95	0.00	0.01	0.03	0.03	P0AB67	pntB	NAD(P) transhydrogenase subunit beta
ECDH1_RS16645	3.23E-209	3.71E-206	4.47	0.14	10.09	6.01	0.01	0.02	0.02	0.02	P77454	glsA1	Glutaminase 1
ECDH1_RS16640	3.26E-152	2.49E-149	4.16	0.16	9.46	5.22	0.01	0.03	0.02	0.03	P77400	ybaT	Inner membrane transport protein YbaT
ECDH1_RS11405	8.74E-58	3.64E-55	3.88	0.24	11.68	4.60	0.00	0.09	0.02	0.09	P63235	gadC	Probable glutamate/gamma-aminobutyrate antiporter
ECDH1_RS01080	1.87E-49	6.13E-47	3.65	0.25	9.19	4.06	0.01	0.08	0.02	0.08	P37194	slp	Outer membrane protein slp
ECDH1_RS09805	3.16E-24	2.59E-22	3.56	0.35	9.56	3.94	0.01	0.18	0.02	0.18	P76251	dmlA	D-malate dehydrogenase [decarboxylating]
ECDH1_RS11400	8.28E-74	5.43E-71	3.41	0.19	11.76	3.53	0.00	0.05	0.02	0.05	P69910	gadB	Glutamate decarboxylase beta
ECDH1_RS03685	7.49E-08	8.59E-07	2.72	0.51	6.37	2.53	0.05	0.34	0.05	0.34	P64574	yghW	Uncharacterized protein YghW
ECDH1_RS01170	3.04E-29	3.25E-27	2.66	0.24	8.16	2.19	0.01	0.07	0.03	0.07	P37630	yhiM	Inner membrane protein YhiM
ECDH1_RS01025	2.98E-70	1.71E-67	2.53	0.14	9.05	1.93	0.01	0.01	0.02	0.02	P37639	gadX	HTH-type transcriptional regulator GadX

ECDH1_RS01380	0.00034462	0.00189129	2.52	0.70	5.83	2.50	0.08	0.66	0.06	0.66	P0AG80	ugpB	sn-glycerol-3-phosphate-binding periplasmic protein UgpB
ECDH1_RS01030	5.54E-61	2.83E-58	2.52	0.15	8.46	1.92	0.01	0.02	0.02	0.02	P63201	gadW	HTH-type transcriptional regulator GadW
ECDH1_RS20315	6.87E-23	4.57E-21	2.47	0.25	9.16	1.91	0.01	0.09	0.02	0.09	P33224	aidB	Putative acyl-CoA dehydrogenase AidB
ECDH1_RS08910	1.46E-10	2.43E-09	2.47	0.39	5.28	2.01	0.10	0.13	0.08	0.13	P76344	zinT	Metal-binding protein ZinT
ECDH1_RS09595	1.09E-18	4.52E-17	2.43	0.28	9.23	1.86	0.01	0.11	0.02	0.11	P64503	yebV	Uncharacterized protein YebV
ECDH1_RS01155	1.93E-24	1.64E-22	2.38	0.23	8.27	1.76	0.01	0.07	0.03	0.07	P0A8S5	uspB	Universal stress protein B
ECDH1_RS21180	2.82E-05	0.00019936 5	2.36	0.56	3.95	1.75	0.30	-0.20	0.17	0.17	P32688	yjbG	Uncharacterized protein YjbG
ECDH1_RS01020	2.51E-33	3.29E-31	2.35	0.20	11.62	1.70	0.00	0.05	0.02	0.05	P69908	gadA	Glutamate decarboxylase alpha
ECDH1_RS06935	2.38E-07	2.57E-06	2.34	0.45	5.60	1.89	0.10	0.21	0.07	0.21	P77326	tfaS	Putative protein TfaS
ECDH1_RS00995	1.70E-26	1.53E-24	2.32	0.22	9.27	1.66	0.01	0.07	0.02	0.07	P37642	yhjD	Inner membrane protein YhjD
ECDH1_RS09890	1.51E-23	1.14E-21	2.28	0.23	9.28	1.63	0.01	0.07	0.02	0.07	P76235	yeaH	UPF0229 protein YeaH
ECDH1_RS01050	1.13E-05	8.86E-05	2.27	0.52	7.99	1.87	0.01	0.39	0.03	0.39	P63204	gadE	Transcriptional regulator GadE
ECDH1_RS11460	1.21E-16	4.11E-15	2.20	0.27	6.20	1.46	0.06	-0.04	0.05	0.05	P76127	bdm	Protein bdm
ECDH1_RS01195	6.99E-18	2.61E-16	2.15	0.25	8.27	1.46	0.01	0.08	0.03	0.08	P37626	yhiI	Uncharacterized protein YhiI

ECDH1_RS06855	0.000102892	0.000651128	2.14	0.55	4.06	1.55	0.29	-0.06	0.16	0.16	P52599	emrK	Probable multidrug resistance protein EmrK
ECDH1_RS07920	0.043237791	0.109237326	2.14	1.06	2.10	2.71	0.57	1.10	0.67	1.10	P0A9E9	yeiL	Regulatory protein YeiL
ECDH1_RS21190	8.66E-06	6.96E-05	2.12	0.48	5.33	1.63	0.09	0.25	0.08	0.25	P0AF45	yjbE	Uncharacterized protein YjbE
ECDH1_RS01075	2.13E-15	6.44E-14	2.12	0.27	6.10	1.39	0.05	0.00	0.05	0.05	P37195	dctR	HTH-type transcriptional regulator DctR
ECDH1_RS14170	2.13E-35	3.36E-33	2.12	0.17	9.42	1.38	0.00	0.04	0.02	0.04	P19932	hyaF	Hydrogenase-1 operon protein HyaF
ECDH1_RS21955	0.000243498	0.001403477	2.08	0.57	3.98	1.43	0.31	-0.15	0.17	0.17	P32139	yihR	Uncharacterized protein YihR
ECDH1_RS01385	0.000749223	0.003716146	2.06	0.61	3.76	1.46	0.36	-0.12	0.20	0.20	P10905	ugpA	sn-glycerol-3-phosphate transport system permease protein UgpA
ECDH1_RS00280	3.00E-46	8.59E-44	2.03	0.14	9.35	1.26	0.00	0.03	0.02	0.03	P0AGM9	xanP	Xanthine permease XanP

B2. HB parent strain and HB evolved strain 2403 (Down-regulated)

target_id	pval	qval	b	se_b	mean_obs	var_obs	tech_var	sigma_sq	smooth_sigma_sq	final_sigma_sq	K12 ID (uniprot)	gene name	annotation
ECDH1_RS19245	1.22E-224	1.86E-221	-4.19	0.13	11.78	5.28	0.00	0.02	0.02	0.02	P0A6L0	deoC	Deoxyribose-phosphate aldolase
ECDH1_RS19240	3.33E-158	3.05E-155	-3.66	0.14	12.37	4.04	0.00	0.02	0.03	0.03	P07650	deoA	Thymidine phosphorylase
ECDH1_RS17755	9.31E-59	4.27E-56	-5.46	0.34	7.43	9.09	0.03	0.14	0.03	0.14	P75682	yagE	Probable 2-keto-3-deoxy-galactonate aldolase YagE
ECDH1_RS19230	1.35E-50	4.75E-48	-2.06	0.14	12.46	1.28	0.00	0.01	0.03	0.03	P0ABP8	deoD	Purine nucleoside phosphorylase DeoD-type
ECDH1_RS17750	1.19E-48	3.63E-46	-4.73	0.32	7.85	6.83	0.02	0.14	0.03	0.14	P77596	yagF	Uncharacterized protein YagF
ECDH1_RS19235	3.41E-45	9.19E-43	-2.19	0.16	13.73	1.46	0.00	0.02	0.04	0.04	P0A6K6	deoB	Phosphopentomutase
ECDH1_RS10830	1.56E-32	1.93E-30	-2.27	0.19	8.29	1.59	0.01	0.05	0.03	0.05	P76177	ydgH	Protein YdgH
ECDH1_RS08305	5.64E-27	5.28E-25	-2.61	0.24	6.18	2.09	0.04	0.03	0.05	0.05	P0A9S3	gatD	Galactitol-1-phosphate 5-dehydrogenase
ECDH1_RS03865	1.60E-23	1.18E-21	-2.23	0.22	8.44	1.55	0.01	0.07	0.02	0.07	P0AFF4	nupG	Nucleoside permease NupG
ECDH1_RS01810	1.63E-23	1.19E-21	-2.21	0.22	12.20	1.52	0.00	0.07	0.03	0.07	P08201	nirB	Nitrite reductase (NADH) large subunit
ECDH1_RS04835	7.33E-08	8.45E-07	-4.44	0.83	2.16	6.32	0.38	0.12	0.64	0.64	P76633	ygcW	Uncharacterized oxidoreductase YgcW
ECDH1_RS14835	2.46E-07	2.65E-06	-3.41	0.66	8.24	4.02	0.01	0.65	0.03	0.65	P68688	grxA	Glutaredoxin 1
ECDH1_RS19730	4.16E-07	4.32E-06	-2.15	0.42	4.18	1.51	0.12	0.04	0.15	0.15	P39360	yjhI	Uncharacterized HTH-type transcriptional regulator YjhI
ECDH1_RS11070	0.0001 33512	0.0008 26659	-2.22	0.58	3.51	1.77	0.27	0.09	0.24	0.24	P76160	ydfR	Uncharacterized protein YdfR
ECDH1_RS11350	0.0004 3224	0.0022 99153	-2.17	0.62	3.35	1.50	0.30	-0.19	0.27	0.27	P77588	ydeQ	Uncharacterized fimbrial-like protein YdeQ
ECDH1_RS11840	0.0043 33833	0.0164 32749	-2.65	0.93	1.71	2.73	0.40	0.38	0.89	0.89	P76090	ynbA	Inner membrane protein YnbA

Appendix 3.5: Metabolomic data

All significant samples are highlighted in pink (n=5).

A. HB parent and HB evolved strain 2403

Data from the top and the bottom section of the table were collected from two independent experiments.

Name	HB parent Average	HB parent SEM	HB evolved strain 2403 Average	HB evolved strain 2403 SEM	ttest
glyoxylic acid	1	0.125065212	0.921871345	0.174796311	0.725644767
pyruvate	1	0.342587707	1.070467322	0.354833505	0.889925832
lactic acid	1	0.660038469	0.313763004	0.126600444	0.337098892
glycerol	1	0.18769823	0.755770624	0.242528847	0.448800371
cytosine	1	0.269381631	0.276548373	0.088046379	0.034030872
uracil	1	0.270442822	0.586558609	0.2299247	0.27768698
fumarate	1	0.258341987	0.718754295	0.146244263	0.371174287
succinate	1	0.262138231	0.795282005	0.250403863	0.587749482
thymine	1	0.385944022	0.934843073	0.438896131	0.913978872
oxaloacetate	1	0.308483691	0.525368078	0.144719758	0.201127881
malate	1	0.168660352	0.848261927	0.150017327	0.520377415
adenine	1	0.246917778	1.04613418	0.208506329	0.89001587
hypoxanthine	1	0.327283957	0.809016498	0.261720931	0.660692105
phosphorylethanolamine	1	0.095118576	0.540785346	0.039744973	0.002126334
alpha ketoglutarate	1	0.469675616	0.786990583	0.243826753	0.697845229
xanthine	1	0.269923033	0.504033937	0.111422444	0.127858189
phenyl pyruvate	1	0.758048015	0.457787877	0.167384202	0.504689177
PEP	1	0.317740865	0.912180511	0.361164914	0.859683395
glyceraldehyde 3-phosphate (G3P) or DHAP	1	0.144488272	0.412732134	0.066894169	0.006144737
transaconitate	1	0.698426589	0.229722474	0.07460484	0.304710305
inositol	1	0.176473034	0.680866112	0.208912	0.276826863
glucose old	1	0.280722896	0.734322045	0.218880241	0.476812554
glucose new	1	0.283881413	0.738212725	0.219965745	0.48682107
D-glycerate 3-phosphate	1	0.719482604	1.533111288	0.366011645	0.527538121
citrate	1	0.302939001	0.417170269	0.141772153	0.119582199
erythrose-4-phosphate	1	0.132140915	0.746352841	0.215595354	0.345194273
pantothenate	1	0.127275863	0.746785884	0.155510478	0.243161139
ribose 5-phosphate	1	0.184853244	0.52962798	0.201286253	0.123529474
ribulose-5-phosphate or xylulose5P	1	0.312661775	0.440838714	0.17481402	0.157153661

uridine	1	0.249642492	0.941311381	0.363143184	0.897340949
inositol 4-phosphate	1	0.114688858	0.438539853	0.106896107	0.00717602
fructose-6-phosphate	1	0.115992485	0.474387711	0.121310854	0.013981248
glucose 1-phosphate or glucose-6-phosphate	1	0.118030051	0.439787169	0.132079543	0.013341216
glucose 6-phosphate	1	0.163000972	0.450894958	0.067418877	0.014381019
glucose 1-phosphate	1	0.156939413	0.458400397	0.09639269	0.018691575
phosphonogluconic acid	1	0.174106261	0.806128278	0.227505803	0.517660457
sedoheptulose-7-phosphate	1	0.320224802	0.445071677	0.222035411	0.19223363
glutathione, reduced GSH	1	0.237890387	0.629843605	0.266805465	0.330710163
dUMP	1	0.120456491	0.756993269	0.139705506	0.224195664
R15BP	1	0.152139121	0.69683846	0.175836723	0.228560431
CMP	1	0.050773751	1.525723619	0.372531483	0.199572742
UMP	1	0.065554786	1.602606737	0.425668106	0.199316766
cAMP	1	0.15691561	0.625039801	0.170721068	0.144530132
inositol 1,4-bisphosphate	1	0.303266204	12.2817522	7.245467253	0.158386352
fructose 1,6bp	1	0.287063771	13.29842643	7.571854665	0.143232149
AMP	1	0.076189057	1.874091135	0.286568111	0.01848779
guanosine 5' monophosphate	1	0.075304103	1.309669186	0.339783165	0.399547866
IP3 (1,4,5) or IP3 (1,3,4)	1	0.219854372	0.516731237	0.14513999	0.103927122
ADP	1	0.308364875	0.607158121	0.110892488	0.264910991
C18:1 Phe	1	0.385935812	0.40023239	0.164029835	0.190520572
folic acid	1	0.048547748	0.518261038	0.185673768	0.036361905
dUTP	1	0.17892997	0.727617184	0.176887212	0.310542228
CTP	1	0.202333615	0.386463014	0.093319877	0.024922064
UTP	1	0.137987214	0.6117321	0.104045157	0.054849535
ATP	1	0.274926375	0.521961548	0.113178476	0.146529791
GTP	1	0.212418187	0.76698366	0.177592212	0.424458459
uridine 5-disphosphoglucuronic acid	1	0.237147995	2.041505783	0.532443562	0.111773548
glutathione, oxidized GSSG	1	0.27021294	3.234257374	0.586644067	0.008577583
NAD	1	0.181445667	1.503042778	0.208997187	0.10665305
NADH	1	0.263499086	0.491975756	0.22114149	0.177971069
NADP	1	0.128354395	0.833043794	0.095111238	0.326528868
coenzyme A	1	0.348971794	0.255101699	0.0762312	0.070523293
acetyl CoA	1	0.389447412	1.226004486	0.499624824	0.730492888

lactic acid	1	0.292007935	0.174877059	0.045566541	0.023489399
cytosine	1	0.400381495	0.02930935	0.007632593	0.041591418
uracil	1	0.41779158	0.079941663	0.021602058	0.059062326
fumerate	1	0.247407841	0.236433888	0.048978961	0.016371347
succinate T1	1	0.399670249	0.209824473	0.049655064	0.085395656
succinate T2	1	0.369474361	0.19191314	0.045254142	0.061731644
thymine	1	0.466365604	0.075915262	0.019460444	0.083082518
oxaloacetate	1	0.38326104	0.155742323	0.039896822	0.059828803
malate	1	0.334995335	0.225659106	0.050320098	0.051597772
adenine	1	0.407624964	0.154389514	0.041508721	0.072927224
alpha ketoglutarate	1	0.321146536	0.29287004	0.071978802	0.063915866
PEP	1	0.430768919	0.030349306	0.008315799	0.054519885
glyceraldehyde 3-phosphate (G3P) or DHAP	1	0.282124884	0.18534355	0.056672589	0.022115452
glycerol-3-phosphate	1	0.32720101	0.123137188	0.033433267	0.028538299
glucose	1	0.297904363	0.284978926	0.067172594	0.047313186
D-glycerate 3-phosphate	1	0.407019704	0.011118804	0.003070378	0.041234602
Citrate	1	0.267342864	0.163063696	0.030081615	0.014424944
ribulose-5-phosphate	1	0.293086471	0.297250111	0.087168965	0.050608015
palmitate C12	1	0.406055762	0.091977244	0.022709613	0.056059198
fructose-6-phosphate	1	0.233516344	0.061107327	0.017867554	0.003900743
glucose 1-phosphate or glucose-6-phosphate	1	0.27387672	0.060793177	0.016905787	0.009050596
glucose-6-phosphate	1	0.431204309	0.262120822	0.135728844	0.141269907
glucose 1-phosphate	1	0.377416006	0.086359712	0.020463255	0.042031523
1,3-bisphosphoglycerate	1	0.511801102	0.051292248	0.013541318	0.101007913
phosphonogluconic acid	1	0.335335852	0.053421602	0.015448991	0.022501283
glutathione, reduced	1	0.3624029	0.408642319	0.102535232	0.155024585
fructose 1,6bp	1	0.318625798	0.010954735	0.003090204	0.014578326
UDP-glucose	1	0.218119087	0.304612933	0.07417265	0.016600825
Glutathione, oxidized	1	0.28564606	0.317611737	0.073684908	0.049441078
NAD	1	0.230629802	0.665878695	0.173649105	0.280512398
NADP	1	0.623128203	0.554096906	0.374035388	0.556555879
acetyl-coa	1	0.287632091	0.789049183	0.226135547	0.58009628
acetoacetyl CoA fragment	1	0.294733023	0.389527298	0.140854114	0.098584692
acetoacetyl CoA parent	1	0.316888752	0.036206862	0.011640215	0.016078572
l-alanine	1	0.084067868	1.780492332	0.175593433	0.003899973
serine	1	0.103664969	0.487895109	0.039515097	0.001719108

proline	1	0.183834988	0.722150831	0.017896453	0.170917688
threonine	1	0.157815371	0.90230718	0.208473491	0.718388099
leucine	1	0.086163177	0.350550327	0.038723267	0.000127707
isoleucine	1	0.183963095	0.331115446	0.087226758	0.011098498
asparagine	1	0.114101299	1.055605422	0.03284843	0.652055285
glutamine	1	0.099732866	0.644813292	0.044983974	0.011763801
lysine	1	0.029315515	0.896630259	0.151448775	0.521662512
glutamic acid	1	0.082331781	0.525255099	0.036028337	0.000743872
methionine	1	0.037978344	1.938990035	0.14313599	0.000222815
arginine	1	0.073336091	0.350504878	0.022123128	2.8665E-05
citrulline	1	0.092501508	0.385822249	0.02490838	0.000206635
AMP	1	0.116616402	1.456718097	0.037519936	0.005802696
ADP	1	0.131788288	0.368425975	0.035191025	0.00168781
ATP	1	0.592084926	0.503908263	0.335111583	0.486687916

B. BDO parent and BDO evolved strain 2406

Data from the top and the bottom section of the table were collected from two independent experiments.

Name	BDO parent Average	BDO parent SEM	BDO evolved strain 2403 Average	BDO evolved strain 2403 SEM	ttest
glyoxylic acid	1	0.067778583	0.971666652	0.104330918	0.825565438
pyruvate	1	0.057794851	1.709213498	0.322403462	0.062277352
lactic acid	1	0.089789083	0.515409869	0.074366568	0.003180555
glycerol	1	0.079082776	0.820861908	0.137223725	0.290794424
cytosine	1	0.15097688	0.287677025	0.047552629	0.002001684
uracil	1	0.105992005	0.603209925	0.077236884	0.016421207
fumarate	1	0.110022851	0.825682949	0.071995592	0.221525872
succinate	1	0.130913633	0.678331113	0.083500158	0.072050443
thymine	1	0.172499184	0.71235168	0.255407259	0.377968748
oxaloacetate	1	0.171151114	0.648960046	0.090370291	0.107274646
malate	1	0.060535941	1.068008965	0.098044666	0.571337361
adenine	1	0.053280985	0.959051368	0.098829414	0.724779944
hypoxanthine	1	0.089905489	1.174589365	0.279119689	0.568047598
phosphorylethanolamine	1	0.058400091	1.077349841	0.157857866	0.658065302
alpha ketoglutarate	1	0.089206663	0.611568577	0.098516653	0.019212966
xanthine	1	0.173558784	0.685840693	0.034739075	0.113832815
phenyl pyruvate	1	0.147532257	1.524600937	0.450567232	0.300667701
PEP	1	0.356897211	1.880266143	0.374850392	0.127410771
glyceraldehyde 3-phosphate (G3P) or DHAP	1	0.119237871	1.13561905	0.079997694	0.372558612
transaconitate	1	0.213867367	0.532754907	0.077371213	0.073996942
inositol	1	0.132738906	0.61823287	0.059794501	0.030541334
glucose old	1	0.09536059	0.797615625	0.098646309	0.178428778
glucose new	1	0.093728342	0.804207728	0.102797596	0.196944126
D-glycerate 3-phosphate	1	0.305351027	0.986538506	0.287098715	0.975164677
citrate	1	0.317336459	0.473722199	0.06187453	0.142224424
erythrose-4-phosphate	1	0.198352396	0.834519898	0.186904873	0.560564736
pantothenate	1	0.056669565	0.738849882	0.073657579	0.022841808
ribose 5-phosphate	1	0.099093057	0.74650643	0.140594299	0.178776025
ribulose-5-phosphate or xylulose5P	1	0.08983587	0.692799861	0.144224897	0.108228229
uridine	1	0.101695469	0.789903331	0.111131989	0.200614261
inositol 4-phosphate	1	0.050286133	1.129939899	0.134604815	0.392253851
fructose-6-phosphate	1	0.045900802	1.138794457	0.150180508	0.402569725
glucose 1-phosphate or glucose-6-phosphate	1	0.051183699	1.099141665	0.134647654	0.510752487
glucose 6-phosphate	1	0.14733267	1.456352881	0.219715074	0.122795882

glucose 1-phosphate	1	0.148989403	1.2475291	0.150000366	0.275370469
phosphonogluconic acid	1	0.173270257	0.646621161	0.053840042	0.087314085
sedoheptulose-7-phosphate	1	0.059262044	0.465066979	0.086203021	0.000914303
glutathione, reduced GSH	1	0.093069601	0.507019631	0.04801776	0.001526972
dUMP	1	0.075596919	0.500215672	0.089167159	0.002704659
R15BP	1	0.057759511	0.951959863	0.124641505	0.735589301
CMP	1	0.066034647	1.340269877	0.144975244	0.06518607
UMP	1	0.084425569	1.351440407	0.151129549	0.076841576
cAMP	1	0.09637917	0.698848667	0.072190679	0.036890965
inositol 1,4-bisphosphate	1	0.417390313	7.573226308	0.436956737	4.51353E-06
fructose 1,6bp	1	0.175101311	15.57264728	1.081019334	9.71534E-07
AMP	1	0.126648927	1.574750189	0.177835145	0.030058555
guanosine 5' monophosphate	1	0.161229748	0.892511768	0.157736088	0.646429019
IP3 (1,4,5) or IP3 (1,3,4)	1	0.118504191	0.687849574	0.125863774	0.108607902
ADP	1	0.149787987	2.324150849	0.441184327	0.021744209
C18:1 Phe	1	0.115472668	0.557842754	0.053239889	0.008352607
folic acid	1	0.079340159	0.148254715	0.024824295	7.07993E-06
dUTP	1	0.110986079	0.842348747	0.082718974	0.287687017
CTP	1	0.168492495	0.584050519	0.082361478	0.057372884
UTP	1	0.147848546	0.89206181	0.224327187	0.698383557
ATP	1	0.213501078	0.79717318	0.05015841	0.38210686
GTP	1	0.288386582	7.370494392	1.018275111	0.000316479
uridine 5-disphosphoglucuronic acid	1	0.225711361	4.299951405	1.363999981	0.044071248
glutathione, oxidized GSSG	1	0.067020852	1.517340361	0.141945588	0.010928052
NAD	1	0.043556172	0.770546268	0.063181558	0.017333703
NADH	1	0.078516064	0.353997684	0.036477628	7.1842E-05
NADP	1	0.138427413	0.518978264	0.073416808	0.015350983
coenzyme A	1	0.101356841	0.677496005	0.047158944	0.020358181
acetyl CoA	1	0.340032474	25.2558086	2.344565366	7.11696E-06
lactic acid	1	0.135160943	0.582213746	0.061008008	0.022586728
cytosine	1	0.144862687	0.9733709	0.093787737	0.881189796
uracil	1	0.102741614	1.04269776	0.079012586	0.750288819
fumerate	1	0.024892834	1.608024011	0.280415436	0.062808581
succinate T1	1	0.082915678	1.309235038	0.131656451	0.082091164
succinate T2	1	0.116476641	1.555093264	0.291300835	0.114796997
thymine	1	0.15851798	1.978133388	0.185421489	0.003896959
oxaloacetate	1	0.160007957	1.084820298	0.120735244	0.683329671
malate	1	0.031541982	1.874958048	0.166282336	0.000853455
adenine	1	0.169572649	1.316427542	0.229548123	0.299751582

alpha ketoglutarate	1	0.067730902	1.2725738	0.11186624	0.070638016
PEP	1	0.175364417	0.744901945	0.059615892	0.205733309
glyceraldehyde 3-phosphate (G3P) or DHAP	1	0.103185953	3.153109348	0.389120498	0.000687153
glycerol-3-phosphate	1	0.147758431	3.883288742	0.366323865	8.39325E-05
glucose	1	0.047256505	1.286095209	0.110668462	0.044721556
D-glycerate 3-phosphate	1	0.226354412	0.384915035	0.039520724	0.028062276
Citrate	1	0.0887014	0.471488418	0.110858478	0.005850292
ribulose-5-phosphate	1	0.130064016	2.691749541	0.597857322	0.024484496
palmitate C12	1	0.257576795	2.0510305	0.160593415	0.008535269
fructose-6-phosphate	1	0.137362343	0.54617083	0.034770239	0.012558138
glucose 1-phosphate or glucose-6-phosphate	1	0.16004725	0.615660983	0.044852119	0.049509053
glucose-6-phosphate	1	0.337948697	1.708540701	0.591180976	0.328527138
glucose 1-phosphate	1	0.080880169	0.424792497	0.018566441	0.000120632
1,3-bisphosphoglycerate	1	0.186047983	0.546405329	0.080319138	0.055566558
phosphonogluconic acid	1	0.186100585	0.192947394	0.00793883	0.00250284
glutathione, reduced	1	0.099743159	0.486894069	0.062356604	0.002406186
fructose 1,6bp	1	0.151792697	0.138607791	0.008878798	0.000473069
UDP-glucose	1	0.081903712	1.355672794	0.068819001	0.010466119
glutathione, oxidized	1	0.069912505	0.873277383	0.033868325	0.141482036
NAD	1	0.176280407	1.142698514	0.207607361	0.614515626
NADP	1	0.521911913	0.74302875	0.315517481	0.684590193
acetyl-coa	1	0.192751597	1.074640696	0.309740926	0.842996591
acetoacetyl CoA fragment	1	0.178921216	1.186884811	0.315641515	0.620422259
acetoacetyl CoA parent	1	0.185540783	0.055086901	0.013585302	0.000954128
l-alanine	1	0.225837571	2.874093899	0.42218349	0.004454523
serine	1	0.176250994	0.955922597	0.058407076	0.818322163
proline	1	0.135921788	0.894733234	0.05948714	0.498173165
threonine	1	0.171961774	1.360469835	0.150812078	0.153676176
leucine	1	0.133355336	1.180994835	0.073618561	0.26882827
isoleucine	1	0.119541443	1.304555865	0.121511172	0.111798626
asparagine	1	0.116978183	0.922652585	0.047232665	0.556822299
glutamine	1	0.25687949	1.081883836	0.082890134	0.769353964
lysine	1	0.067679637	1.034379751	0.117721987	0.806511708
glutamic acid	1	0.156326205	1.22593631	0.084397672	0.239175068
methionine	1	0.124219404	1.119725673	0.072395823	0.429158596
arginine	1	0.152480395	0.996902911	0.07431959	0.985880089
citrulline	1	0.149684126	1.000807073	0.073086196	0.996252802
AMP	1	0.029775093	4.627129864	0.268591741	9.09319E-07
ADP	1	0.058986119	1.361957532	0.050888473	0.001652798
ATP	1	0.232872395	1.017162401	0.399139409	0.971283775

C. DH1Δ5 n-butanol parent strain and evolved strain 2622

Name	DH1Δ5 n-butanol parent Control Average	DH1Δ5 n-butanol parent Control SEM	DH1Δ5 n-butanol evolved strain 2622 Average	DH1Δ5 n-butanol evolved strain 2622 SEM	ttest
glyoxylic acid	1	0.079961353	1.340346848	0.058544371	0.008898261
pyruvate	1	0.195385747	0.423911669	0.031765345	0.019581414
lactic acid	1	0.120854	0.842872112	0.070637943	0.294217089
glycerol	1	0.311919624	1.135881988	0.218474541	0.730460399
cytosine	1	0.271362435	0.50336266	0.111537917	0.128959852
uracil	1	0.144191779	0.519965993	0.053157939	0.014150872
fumarate	1	0.137099913	0.818991218	0.128959762	0.364370523
succinate	1	0.120972526	0.633422271	0.05230338	0.023872281
thymine	1	0.169416768	0.685301597	0.153050066	0.205405437
oxaloacetate	1	0.176118984	0.492247458	0.046226759	0.023610963
malate	1	0.08188351	1.506380938	0.06712129	0.001385911
adenine	1	0.145223022	0.792375428	0.080441441	0.246401705
hypoxanthine	1	0.133640519	0.017517454	0.004439101	8.01215E-05
phosphorylethanolamine	1	0.122878145	0.723746505	0.072618164	0.088966453
alpha ketoglutarate	1	0.126026572	0.570723788	0.060956364	0.015432303
xanthine	1	0.099881688	0.023033011	0.002676117	1.00366E-05
phenyl pyruvate	1	0.154471482	0.992604267	0.058342749	0.965372932
PEP	1	0.157436774	1.003366231	0.122558981	0.986952003
glyceraldehyde 3-phosphate (G3P) or DHAP	1	0.165695824	0.865170497	0.046198511	0.455723115
transaconitate	1	0.19982276	0.431610525	0.038134761	0.02341191
inositol	1	0.120631183	0.790204989	0.042458075	0.139531127
glucose old	1	0.122182131	0.814541837	0.068349717	0.221860528
glucose new	1	0.123006855	0.816180498	0.069896208	0.230040374
D-glycerate 3-phosphate	1	0.225794864	0.704151094	0.067217614	0.244625836
citrate	1	0.251316893	0.36836591	0.049417448	0.038948117
erythrose-4-phosphate	1	0.126480498	0.679768311	0.06526133	0.054566816
pantothenate	1	0.075720364	0.753353411	0.034407028	0.017993409
ribose 5-phosphate	1	0.286155186	0.826809553	0.096320214	0.581994589
ribulose-5-phosphate or xylulose5P	1	0.299352338	0.899905245	0.106482908	0.760796317
uridine	1	0.154772616	1.15531484	0.098187175	0.421413746
inositol 4-phosphate	1	0.167605324	0.63063779	0.070960208	0.076931912
fructose-6-phosphate	1	0.16051992	0.630467388	0.080081442	0.073362876

glucose 1-phosphate or glucose-6-phosphate	1	0.1584536	0.641313109	0.080152407	0.078064962
glucose 6-phosphate	1	0.17685018	0.615872928	0.081514826	0.084006296
glucose 1-phosphate	1	0.17649996	0.628444272	0.087580794	0.096053789
phosphonogluconic acid	1	0.098927321	0.193177804	0.015386305	4.14237E-05
sedoheptulose-7-phosphate	1	0.296386707	0.644194829	0.091120018	0.284341785
glutathione, reduced GSH	1	0.108069685	0.478198364	0.105700296	0.008671647
dUMP	1	0.172117431	1.135972885	0.143442989	0.560758179
R15BP	1	0.124979506	0.915684581	0.081935183	0.588083944
CMP	1	0.184443859	1.401653474	0.130013589	0.112968665
UMP	1	0.145144091	1.429409242	0.094902288	0.038339628
cAMP	1	0.122849231	0.853531775	0.088090475	0.36096751
inositol 1,4-bisphosphate	1	0.0886803	0.899384903	0.036148236	0.324119049
fructose 1,6bp	1	0.128585259	0.886554739	0.041094613	0.42509618
AMP	1	0.18622492	1.422819883	0.095234574	0.07787956
guanosine 5' monophosphate	1	0.051557782	1.169108542	0.105852732	0.188852539
IP3 (1,4,5) or IP3 (1,3,4)	1	0.231703829	0.681036666	0.071972081	0.225066817
ADP	1	0.193681573	1.177516719	0.045478327	0.398290858
C18:1 Phe	1	0.119374338	0.774663656	0.092831089	0.174526937
folic acid	1	0.107049242	0.899922178	0.073900825	0.463790292
dUTP	1	0.172752556	0.502013633	0.064134052	0.026972157
CTP	1	0.155742444	0.246927589	0.035863699	0.00151765
UTP	1	0.228331883	0.435749571	0.067393391	0.045239132
ATP	1	0.088256797	0.721528033	0.073722216	0.041747956
GTP	1	0.156536279	1.253478034	0.281947427	0.454510727
uridine 5-disphosphoglucuronic acid	1	0.182026665	0.855593677	0.058291461	0.47158777
glutathione, oxidized GSSG	1	0.096992109	0.853307015	0.099872824	0.322809731
NAD	1	0.089921811	1.039808995	0.069076773	0.734596893
NADH	1	0.167974762	0.700988328	0.086062915	0.151793408
NADP	1	0.123580624	0.489884567	0.097057586	0.011766259
coenzyme A	1	0.117664959	0.685289631	0.046925391	0.037853818
acetyl CoA	1	0.049360203	1.531960273	0.127738963	0.004645108

D. BW25113Δ5 n-butanol parent strain and evolved strain 2731

Name	BW25113 Δ5 n-butanol parent Control Average	BW25113Δ5 n-butanol parent SEM	BW25113Δ5 n-butanol evolved strain 2731 Average	BW25113Δ5 n-butanol evolved strain 2731 SEM	ttest
glyoxylic acid	1	0.196504455	3.799533574	0.141326351	2.83718E-06
pyruvate	1	0.098608888	0.269684184	0.0428442	0.00013885
lactic acid	1	0.234497971	8.985038991	1.023357202	6.27159E-05
glycerol	1	0.163512346	0.947717561	0.03982891	0.763996972
cytosine	1	0.054459194	0.667324508	0.119692291	0.035262772
uracil	1	0.181359966	0.673465271	0.135951324	0.187651592
fumarate	1	0.093972885	2.921007893	0.077901422	2.65437E-07
succinate	1	0.093693399	2.282821713	0.230218847	0.000862457
thymine	1	0.213089812	0.978758628	0.120908612	0.933041799
oxaloacetate	1	0.107044702	0.976803099	0.140017923	0.898538937
malate	1	0.192308769	3.747189486	0.100725058	1.42884E-06
adenine	1	0.097593972	0.796710793	0.062611696	0.11765448
hypoxanthine	1	0.139899536	0.907756616	0.125942292	0.637262955
phosphorylethanolamine	1	0.08822893	0.648548268	0.057259187	0.010209711
alpha ketoglutarate	1	0.10268839	1.033961683	0.088413706	0.808418105
xanthine	1	0.166544472	8.503849326	0.527106127	8.33611E-07
phenyl pyruvate	1	0.193978066	0.399617069	0.0635906	0.018678138
PEP	1	0.574193622	0.926485527	0.233965267	0.9085424
glyceraldehyde 3-phosphate (G3P) or DHAP	1	0.093231353	1.095904578	0.135272391	0.575468704
transaconitate	1	0.093044091	0.943062269	0.108164241	0.70028432
inositol	1	0.203973007	2.265202208	0.453830984	0.034559533
glucose old	1	0.071906489	0.677627986	0.088204988	0.022055974
glucose new	1	0.062768845	0.68367991	0.075828556	0.012360708
D-glycerate 3-phosphate	1	0.80303137	1.362765485	0.464072567	0.705912287
citrate	1	0.129447215	0.929082333	0.129702248	0.708842602
erythrose-4-phosphate	1	0.086891667	0.999627223	0.041624339	0.997007652
pantothenate	1	0.178761136	0.805875952	0.048733845	0.325393373
ribose 5-phosphate	1	0.24756624	0.553189762	0.099632869	0.132603448
ribulose-5-phosphate or xylulose5P	1	0.178899173	0.622670446	0.105331622	0.106653038
uridine	1	0.106708036	1.504018812	0.196913063	0.054532208
inositol 4-phosphate	1	0.134860626	0.651910417	0.030225011	0.035885499
fructose-6-phosphate	1	0.14344239	0.624192439	0.041423612	0.03597281

glucose 1-phosphate or glucose-6-phosphate	1	0.1519501	0.638376907	0.04840555	0.0530889
glucose 6-phosphate	1	0.093687143	0.767754707	0.042619794	0.054022219
glucose 1-phosphate	1	0.131333736	0.698541635	0.042832252	0.060650415
phosphonogluconic acid	1	0.113117489	1.183935699	0.234512299	0.49996582
sedoheptulose-7-phosphate	1	0.140240624	0.944916386	0.077682614	0.740011153
glutathione, reduced GSH	1	0.160544799	0.680838246	0.146255892	0.179866888
dUMP	1	0.205177563	0.534080405	0.027594508	0.054520569
R15BP	1	0.160214465	1.016363213	0.087031635	0.930694917
CMP	1	0.144979334	0.706495676	0.039537113	0.086572273
UMP	1	0.082203847	0.711735877	0.05563073	0.019764764
cAMP	1	0.106515913	0.457773141	0.058262612	0.002094001
inositol 1,4-bisphosphate	1	0.074625863	1.979887988	0.81230257	0.264008226
fructose 1,6bp	1	0.105937068	3.358037059	1.500159861	0.155528919
AMP	1	0.06000111	1.087870838	0.091895806	0.446458723
guanosine 5' monophosphate	1	0.103317219	0.969971936	0.086058016	0.828885152
IP3 (1,4,5) or IP3 (1,3,4)	1	0.11159669	0.88173382	0.160219559	0.561497835
ADP	1	0.084359878	1.381660377	0.246913203	0.181692422
C18:1 Phe	1	0.123310892	0.810754849	0.097870428	0.263700162
folic acid	1	0.306010673	1.705662704	0.200780695	0.089991831
dUTP	1	0.066920066	2.454570562	0.23611978	0.000350971
CTP	1	0.196733499	0.499282997	0.054514225	0.039766967
UTP	1	0.127601743	0.904261062	0.105671715	0.579251741
ATP	1	0.118873683	0.60258088	0.145024287	0.066891244
GTP	1	0.188339596	0.850473529	0.193918578	0.595286847
uridine 5-disphosphoglucuronic acid	1	0.123799038	2.055955119	0.346174491	0.020758267
glutathione, oxidized GSSG	1	0.204658687	1.713259178	0.243731003	0.055329113
NAD	1	0.129152713	1.291866107	0.058222478	0.073338528
NADH	1	0.161973432	0.524825514	0.110176856	0.041481157
NADP	1	0.053767233	0.804963204	0.052168807	0.031451929
coenzyme A	1	0.052628591	0.771699424	0.126651306	0.134560037
acetyl CoA	1	0.099678204	12.75947408	3.304134262	0.007428828

Appendix 4: *Strains, plasmids, oligonucleotides, sequences, and RNA-sequencing results for Chapter 4*

Appendix 4.1: Strains

A. *E. coli* strains

E. coli DH10B was used for DNA construction and BL21(de3) Star-T1^R was used for heterologous production of proteins for purification.

Organism	Name	Description	Source
E. coli	DH10B	F- endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ-	Invitrogen
E. coli	BL21 (DE3) Star T1R	RNaseE mutation to increase mRNA stability, Δ*fhuA*	A. Martin

B. *Saccharomyces cerevisiae* strains

BY4741 (MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) and BY4742 (MATα *his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) were used as the parent for all yeast strains generated in this study. BY4741 was obtained from J. Rine Lab. BY4742 and all heat shock protein knockouts were provided by the J. Thorner Lab. Protease knockout strains (BJ1991 and BJ5457) were gifts from J. Cate Lab. Additional modifications to these strains were generated using the CRISPR-Cas9 system as described in the method section using the corresponding plasmids listed in the Constructs for genome engineering studies under the constructs section (Appendix 4.2). See table below for corresponding integration fragments.

B1. Production strains

Organism	Strain	Genotype	Number	Source
S. cerevisiae	BY4741	ΔAdh1	844	*J. Rine Lab*
S. cerevisiae	BY4741	BY4741 Delta YGR252W (GCN5)	1942	*ATCC*
S. cerevisiae	BY4741	ΔAdh1 ΔPBR1	2067	*This study*
S. cerevisiae	BY4741	ΔAdh1 ΔPEP4	2068	*This study*
S. cerevisiae	BY4741	ΔAdh1 ΔPBR1 ΔPEP4	2163	*This study*
S. cerevisiae	BY4741	ΔGPD1::AdhE2	2320	*This study*
S. cerevisiae	BY4741	ΔGCN5 ΔADH1	2325	*This study*
S. cerevisiae	BY4741	ΔGCN5 ΔADH1 ΔGPD1	2388	*This study*
S. cerevisiae	BY4741	ΔADH1 ΔADH5	2572	*This study*
S. cerevisiae	BY4741	ΔADH1 ΔADH6	2573	*This study*
S. cerevisiae	BY4741	ΔADH1 ΔGCY1	2574	*This study*
S. cerevisiae	BY4741	ΔADH1 ΔADH5 ΔADH6 ΔGPD2	2638	*This study*
S. cerevisiae	BY4741	ΔADH1 ΔADH5 ΔADH6 ΔGPD1	2639	*This study*
S. cerevisiae	BY4741	ΔADH1 ΔADH5ΔADH6 ΔDHH1	2640	*This study*

S. cerevisiae	BY4741	ΔADH1 ΔADH5 ΔADH6 ΔCOS12	2641	*This study*
S. cerevisiae	BY4741	ΔADH1 ΔADH5 ΔADH6	2597	*This study*
S. cerevisiae	BY4741	ΔADH1 ΔADH5 ΔADH6 ΔGPD1 ΔGPD2	2666	*This study*
S. cerevisiae	BY4741	ΔGPD1 ΔGPD2 ΔADH1 ΔADH5 ΔADH6 ΔADH4::PGK1p_eutE	2785	*This study*
S. cerevisiae	BY4741	ΔADH1 ΔADH5 ΔADH6 ΔADH4::eutE ΔADH3::pdc ΔGPD1 ΔGPD2	2812	*This study*
S. cerevisiae	BY4741	ΔADH1 ΔADH5 ΔADH6 ΔADH4::eutE ΔADH3::pdc ΔGPD1 ΔGPD2 YPRCΔ15::Pha_hbd_Crt	2942	*This study*
S. cerevisiae	BY4741	ΔADH1 ΔADH5 ΔADH6 ΔADH4::eutE ΔADH3::pdc ΔGPD1 ΔGPD2 YPRCΔ15::Pha_hbd_Crt YPRCτ3::Ter_ADLH21_ADH6	2963	*This study*
S. cerevisiae	BY4741	ΔADH1 ΔADH5 ΔADH6 ΔADH4::eutE ΔADH3::pdc ΔGPD1 ΔGPD2 YPRCΔ15::Pha_hbd_Crt YPRCτ3::Ter_ADLH21_ADH7	2964	*This study*
S. cerevisiae	BY4741	ΔADH1 ΔADH5 ΔADH6 ΔADH4::eutE ΔADH3::pdc ΔGPD1 ΔGPD2 YPRCΔ15::Pha_hbd_Crt YPRCτ3::Ter_ADLH21_ADH9	2965	*This study*
S. cerevisiae	BY4742	ΔHSP30 ΔADH1	2305	*This study*
S. cerevisiae	BY4742	ΔSLX8 ΔADH1	2306	*This study*
S. cerevisiae	BY4742	ΔCPR7 ΔADH1	2312	*This study*
S. cerevisiae	BY4742	ΔRKR1ΔADH1	2313	*This study*
S. cerevisiae	BY4742	ΔHDR1 ΔADH1	2314	*This study*
S. cerevisiae	BY4742	ΔHSP42 ΔADH1	2315	*This study*
S. cerevisiae	BY4742	ΔUTR2 ΔADH1	2316	*This study*
S. cerevisiae	BY4742	ΔSAN1 ΔADH1	2317	*This study*
S. cerevisiae	BY4742	ΔSSM4 ΔADH1	2318	*This study*
S. cerevisiae	BY4742	ΔLHS1 ΔADH1	2319	*This study*

B2. Protein quality control strains.

Organism	Strain	Genotype	Source	Number
S. cerevisiae	BJ5457	Mat alpha, ura3, trp1, lys2, leu2, his3, pep4::his3, prb1, can 1, GAL	*J. Cate Lab*	1877
S. cerevisiae	BJ1991	Mat alpha, leu2, trp1, ura3, prb1, pep4, gal2	*J. Cate Lab*	1876
S. cerevisiae	BY4742	ΔAMS1	*J. Thorner Lab*	2126
S. cerevisiae	BY4742	ΔYDJ1	*J. Thorner Lab*	2125
S. cerevisiae	BY4742	ΔSSB1	*J. Thorner Lab*	2124
S. cerevisiae	BY4742	Δ SSA3	*J. Thorner Lab*	2123
S. cerevisiae	BY4742	ΔSTE3	*J. Thorner Lab*	2122
S. cerevisiae	BY4742	ΔAPE4	*J. Thorner Lab*	2121
S. cerevisiae	BY4742	Δ TDH3	*J. Thorner Lab*	2120
S. cerevisiae	BY4742	ΔATG19	*J. Thorner Lab*	2119
S. cerevisiae	BY4742	ΔSSA2	*J. Thorner Lab*	2118
S. cerevisiae	BY4742	ΔMOT2/NOT4_YER068W	*J. Thorner Lab*	2065
S. cerevisiae	BY4742	ΔSLX8 (YER116C)	*J. Thorner Lab*	2064
S. cerevisiae	BY4742	ΔRPN4 (YDL020C)	*J. Thorner Lab*	2063
S. cerevisiae	BY4742	ΔSSM4/DOA10 (YIL030C)	*J. Thorner Lab*	2062
S. cerevisiae	BY4742	ΔHSP42 (YDR171W)	*J. Thorner Lab*	2061
S. cerevisiae	BY4742	ΔLHS1 (YKL073W)	*J. Thorner Lab*	2060
S. cerevisiae	BY4742	ΔHSP30 (YCR021C)	*J. Thorner Lab*	2059
S. cerevisiae	BY4742	ΔPBR1 (YEL060C)	*J. Thorner Lab*	2058
S. cerevisiae	BY4742	ΔUTR2/CRH2 (YEL040W)	*J. Thorner Lab*	2057
S. cerevisiae	BY4742	ΔSAN1 (YDR143C)	*J. Thorner Lab*	2056
S. cerevisiae	BY4742	ΔUMP1 (YBR173C)	*J. Thorner Lab*	2055
S. cerevisiae	BY4742	ΔPEP4 (YPL154C)	*J. Thorner Lab*	2054
S. cerevisiae	BY4742	ΔHDR1 (YOL013C)	*J. Thorner Lab*	2053
S. cerevisiae	BY4742	ΔRKR1/LTN1 (YMR247C)	*J. Thorner Lab*	2052
S. cerevisiae	BY4742	ΔSSA1 (YAL005C)	*J. Thorner Lab*	2051
S. cerevisiae	BY4742	ΔSSA4	*J. Thorner Lab*	2127

Appendix 2.2: Plasmids used for production and strain construction

A. Constructs for promoter screening

Plasmid	Selection/ Origin	Promoter	Number	Source
pESCLeu2d-ter-adhE2	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	795	Brooks Bond-Watts
pESCLeu2d-AdhE2.TDH3p(5'UTR-PYK2)TdTer	Cb, Leu2d; pUC, 2 micron	pTDH3, pGAL10	1534	This study
pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)TdTer	Cb, Leu2d; pUC, 2 micron	pCCW12, pGAL10	1525	This study
pESCLeu2d-(CCW12p)TdTer-(TDH3p)ALD5-(FBA1p)ADH2	Cb, Leu2d; pUC, 2 micron	pCCW12, pTDH3, PFBA1	2391	This study

B. Constructs for codon usage screening

Plasmid	Selection/ Origin	Promoter	Number	Source
pESCLeu2d-AdhE2.(5'UTR-PYK2)sTdTer(gly)	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	1551	This study
pESCLeu2d-AdhE2.(5'UTR-PYK2)sTdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	1552	This study
pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)sTdTer(gly)	Cb, Leu2d; pUC, 2 micron	pGAL1, pCCW12	1556	This study
pESCLeu2d-AdhE2.TDH3p(5'UTR-PYK2)sTdTer(gly)	Cb, Leu2d; pUC, 2 micron	pGAL1, pTDH3	1557	This study
pESC_Leu_AdhE2_CCW12_5'UTR_PYK2_TdTer(S.c codon optimized))	Cb, Leu2d; pUC, 2 micron	pGAL1, pCCW12	1558	This study
pESCLeu2d-AdhE2.TDH3p(5'UTR-PYK2)sTdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pTDH3	1559	This study
pESCLeu2d-AdhE2.CCW12p(5'UTR-PYK2)sTdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	1568	This study

C. Constructs for screening thiolase homologs

Plasmid	Selection / Origin	Promoter	Description	Number	Source
pESC.His-Bu2	Cb, HIS3; ColE1, 2 micron	pTEF1, pPGK1, pPDC1	phaA from *R. eutropha*	800	Brooks Bond-Watt
pESC_His_Erg10_hbd_crt (C terminal Hisx10)	Cb, HIS3; ColE1, 2 micron	pTEF1, pPGK1, pPDC1	Erg10 from *S. Pombe*	1384	This study
pESC_His_Erg10_hbd_crt	Cb, HIS3; ColE1, 2 micron	pTEF1, pPGK1, pPDC1	Erg10 from *S. Pombe*	1383	This study

D. Constructs for Ter homolog screening

Plasmid	Selection/ Origin	Promoter	Number	Source
pESC_leu2d-adhe2-(eg)ter	Cb, Leu2D; pBR322, 2 micron	pGAL1, pGAL10	1124	*Michiei Sho*
pESC_leu2D_adhe2_(Eg)ter(E.coli_codon)	Cb, Leu2D; pBR322, 2 micron	pGAL1, pGAL10	1067	*Michael Blaisse*
pESC_Leu_adhE2_EgTer (YCO)	Cb, Leu2D; ColE1, 2 micron	pGAL1, pGAL10	1328	*This study*
pESC_Leu_adhE2_Hisx10MECR1_	Cb, Leu2D; ColE1, 2 micron	pGAL1, pGAL10	1429	*This study*
pESC_Leu_adhE2_MECR1	Cb, Leu2D; ColE1, 2 micron	pGAL1, pGAL10	1428	*This study*

E. Constructs for 5′- and 3′-untranslated region (UTR) screening.

Plasmid	Selection / Origin	Promoter	Description	Number	Source
pESCLeu2d-AdhE2.(5'UTR-TPI1)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1413	*This study*
pESCLeu2d-AdhE2.(5'UTR-TDH2-YJR009C)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1414	*This study*

pESCLeu2d-AdhE2.(5'UTR-FBA1-YKL060C)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1415	*This study*
pESCLeu2d-AdhE2.(5'UTR-GPM1-YKL152C)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1416	*This study*
pESCLeu2d-AdhE2.(5'UTR-YLR075W)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1417	*This study*
pESCLeu2d-AdhE2.(5'UTR-YHL001W)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1418	*This study*
pESCLeu2d-AdhE2.(5'UTR-YJL177W)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1419	*This study*
pESCLeu2d-AdhE2.TdTer(3'UTR-FBA1)	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1424	*This study*
pESCLeu2d-AdhE2.TdTer(3'UTR-YJL177W)	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1425	*This study*
pESCLeu2d-AdhE2.(5'UTR-FBA)TdTer(3'UTR-FBA1)	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1426	*This study*
pESCLeu2d-AdhE2.(5'UTR-FBA)TdTer(3'UTR-YJL177W)	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1427	*This study*
pESCLeu2d-AdhE2.(5'UTR-TDH1)TdTer (#1453)	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1453	*This study*
pESCLeu2d-AdhE2.(5'UTR-PYK2)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1454	*This study*
pESCLeu2d-AdhE2.(5'UTR-PGI1)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1455	*This study*

pESCLeu2d-AdhE2.(5'UTR-PFK1)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1456	*This study*
pESCLeu2d-AdhE2.(5'UTR-PFK2)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1457	*This study*
pESCLeu2d-AdhE2.(5'UTR-ENO1)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1458	*This study*
pESCLeu2d-AdhE2.(5'UTR-ENO2)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1459	*This study*
pESCLeu2d-AdhE2.(5'UTR-CDC19)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1460	*This study*
pESCLeu2d-AdhE2.5'UTR-TDH3_TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	1464	*This study*
pESCLeu2d-(5'UTR-PYK2)AdhE2.(5'UTR-PYK2)TdTer	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL10	Ter was driven by the pGAL1 promoter and the ADH1t with UTRs;	2401	*This study*

F. Constructs for Aldh and Adh homolog screening

Plasmid	Selection / Origin	Promoter	Description	Number	Source
pESC_Leu. (5'UTR)Tdter. Aldh21.Adh14	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2805	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh21.Adh13	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2804	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh21.Adh12	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2803	*This study*

pESC_Leu. (5'UTR)Tdter. Aldh21.Adh10	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2802	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh21.Adh9	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2801	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh21.Adh7	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2800	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh21.Adh6	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2799	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh21.Adh5	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2798	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh21.Adh4	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2797	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh21.Adh3	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2796	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh12.Adh22	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2570	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh12.Adh8	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2569	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh12.Adh2	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2568	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh10.Adh22	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2567	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh10.Adh8	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2566	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh10.Adh2	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2565	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh7.Adh22	Cb, Leu2d; pUC, 2 micron		pGal10_5'PYK2_UTR_Td Ter_ADH1t;	2564	*This study*

		pGAL1, pGAL7, pGAL10	pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t		
pESC_Leu. (5'UTR)Tdter. Aldh7.Adh8	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2563	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh7.Adh2	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2562	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh6.Adh22.	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2561	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh6.Adh8	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2560	*This study*
pESC_Leu. (5'UTR)Tdter. Aldh6.Adh2.	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2559	*This study*
pESC_Leu. (5'UTR)Tdter_Aldh5_AD H22	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2558	*This study*
pESC_Leu. (5'UTR)Tdter_Aldh5_AD H8	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2557	*This study*
pESC_Leu. (5'UTR)Tdter_Aldh5_AD H2	Cb, Leu2d; pUC, 2 micron	pGAL1, pGAL7, pGAL10	pGal10_5'PYK2_UTR_Td Ter_ADH1t; pGAL1_Adhs_TPS3t; pGAL7_Aldhs_CYC1t	2556	*This study*

G. Constructs for multi-component optimization

Plasmid	Selection / Origin	Description	Number	Source
pVYY1.0.0_2	Cb, URA3; pUC, 2micron	pCCW12_cutsite_PRM9t; pTDH3_cutsite_SPG5t	1799	*This study*
pVYY1.0.0.5	Cb, URA3; pUC, 2micron	pCCW12_cutsite_PRM9t; pTDH3_gALD5_HIS5t_cutsite_SPG5t	1879	*This study*
pVYY1.C.0	Cb, URA3; pUC, 2micron	pCCW12_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_cutsite_SPG5t	1828	*This study*
pVYY1.1.0	Cb, URA3; pUC, 2micron	pCCW12_5'PYK2_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_cutsite_SPG5t	1821	*This study*
pVYY1.2.0	Cb, URA3; pUC, 2micron	pCCW12_5'PFK1_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_cutsite_SPG5t	1822	*This study*

pVYY1.3.0	Cb, URA3; pUC, 2micron	pCCW12_5'PFK2_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_cutsite_SPG5t	1823	*This study*
pVYY1.4.0	Cb, URA3; pUC, 2micron	pCCW12_5'YHL001W_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_cutsite_SPG5t	1824	*This study*
pVYY1.5.0	Cb, URA3; pUC, 2micron	pCCW12_5'TDH2_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_cutsite_SPG5t	1825	*This study*
pVYY1.6.0	Cb, URA3; pUC, 2micron	pCCW12_5'TDH3_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_cutsite_SPG5t	1826	*This study*
pVYY1.7.0	Cb, URA3; pUC, 2micron	pCCW12_5'VSV_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_cutsite_SPG5t	1848	*This study*
pVYY1.8.0	Cb, URA3; pUC, 2micron	pCCW12_5'VSV_gTdTer_3'VSV_PRM9t; pTDH3_gALD5_HIS5t_cutsite_SPG5t	1827	*This study*

pVYY1.0.1_1	Cb, URA3; pUC, 2micron	pCCW12_cutsite_PRM9t; pTDH3_gALD5_HIS5t_pFBA1_gADH_SPG5t	2001	*This study*
pVYY1.1.1_PYK2	Cb, URA3; pUC, 2micron	pCCW12_5'PYK2_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_pFBA1_gADH_SPG5t	1997	*This study*
pVYY1.2.1_PFK1	Cb, URA3; pUC, 2micron	pCCW12_5'PFK1_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_pFBA1_gADH_SPG5t	1972	*This study*
pVYY1.4.1_YHL001W	Cb, URA3; pUC, 2micron	pCCW12_5'PFK2_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_pFBA1_gADH_SPG5t	1973	*This study*
pVYY1.3.1_PFK2	Cb, URA3; pUC, 2micron	pCCW12_5'YHL001W_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_pFBA1_gADH_SPG5t	2002	*This study*
pVYY1.5.1_TDH2	Cb, URA3; pUC, 2micron	pCCW12_5'TDH2_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_pFBA1_gADH_SPG5t	1998	*This study*
pVYY1.6.1_TDH3	Cb, URA3; pUC, 2micron	pCCW12_5'TDH3_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_pFBA1_gADH_SPG5t	1974	*This study*
pVYY1.7.1_VSV	Cb, URA3; pUC, 2micron	pCCW12_5'VSV_gTdTer_PRM9t; pTDH3_gALD5_HIS5t_pFBA1_gADH_SPG5t	1975	*This study*
pVYY1.8.1_5'VSV_3'VSV	Cb, URA3; pUC, 2micron	pCCW12_5'VSV_gTdTer_3'VSV_PRM9t; pTDH3_gALD5_HIS5t_pFBA1_gADH_SPG5t	1976	*This study*

H. Constructs for transcript processing studies

Plasmid	Selection / Origin	Promoter	Number	Source
pRS316_TDH3p__TDH3t	Cb, Ura3; pBR322, CEN ARS4	pTDH3	2186	This study
pRS316_TDH3_gTdTerTDH3	Cb, Ura3; pBR322, CEN ARS4	pTDH3	1800	This study

pRS316_TDH3_SSA1_TDH3	Cb, Ura3; pBR322, CEN ARS4	pTDH3	2303	This study
pRS316_SSA1_YDJ1	Cb, Ura3; pBR322, CEN ARS4	pTDH3, pTEF1	2304	This study
pESC-Leu_YDJ1_SSA1	Cb, Leu2d; pUC, 2 micron	pTDH3, pTEF1	2326	This study

pESC_URA_ANB1	Cb, Ura3; pUC, 2 micron	pGAL10	2590	This study
pESC_URA_RPS14B	Cb, Ura3; pUC, 2 micron	pGAL10	2591	This study

pESC_URA_TMA10	Cb, Ura3; pUC, 2 micron	pGAL10	2592	This study
pESC_URA_DBP2	Cb, Ura3; pUC, 2 micron	pGAL10	2599	This study
pESC_URA_RLI1	Cb, Ura3; pUC, 2 micron	pGAL10	2600	This study

I. Constructs for CRISPR-Cas9 genome editing

pCas-Pphe-BsaI_NAT (2046) was constructed based on the plasmid template pCAS_Pphe_BASI (1943) from the J. Cate lab, where the original G418 selection marker was replaced by NAT selection through Gibson reaction. pCas-Pphe-BsaI_NAT was used as the template to construct all the following plasmids for genome editing experiments. All plasmids were constructed by digested by BsaI to allow the insertion of guide sequence. All guide sequences were generated using the CRISPR function on Benchling. Two 60 bp single stranded oligoes (forward and reverse) that contained the 20 bp guide sequence plus 20 bp upstream and downstream homology sequence were ordered from IDT. These two oligoes were then Gibson were with the BsaI digested 2046 to generated the desired plasmids. All constructs were confirmed with sanger sequencing (Quintara Bioscience or UC Berkeley Barker Sequencing Facility).

Plasmid	Selection/ Origin	Description	Number	Source
pCAS_Pphe_BSAI	Km, G418; pUC, 2 micron	Cas9; BsaI cutting site for guide sequence cloning	1943	*J. Cate lab*
pCAS_Pphe-BsaI_NAT	NAT; pUC, 2 micron	Cas9; BsaI cutting site for guide sequence cloning	2046	*This study*
pCAS_Pphe-_NAT_g3GCY1	NAT; pUC, 2 micron	Guide targeting for the GCY1 locus	2523	*This study*
pCAS_Pphe-_NAT_g2GCY1	NAT; pUC, 2 micron	Guide targeting for the GCY1 locus	2522	*This study*
pCAS_Pphe-_NAT_g1GCY1	NAT; pUC, 2 micron	Guide targeting for the GCY1 locus	2521	*This study*
pCAS_Pphe-_NAT_g3ADH6	NAT; pUC, 2 micron	Guide targeting for the ADH6 locus	2520	*This study*
pCAS_Pphe-_NAT_g2ADH6	NAT; pUC, 2 micron	Guide targeting for the ADH6 locus	2519	*This study*
pCAS_Pphe-_NAT_g1ADH6	NAT; pUC, 2 micron	Guide targeting for the ADH6 locus	2518	*This study*
pCAS_Pphe-_NAT_g3ADH5	NAT; pUC, 2 micron	Guide targeting for the ADH5 locus	2517	*This study*
pCAS_Pphe-_NAT_g2ADH5	NAT; pUC, 2 micron	Guide targeting for the ADH5 locus	2516	*This study*
pCAS_Pphe-_NAT_g1ADH5	NAT; pUC, 2 micron	Guide targeting for the ADH5 locus	2515	*This study*
pCAS_Pphe-_NAT_g1ADH3	NAT; pUC, 2 micron	Guide targeting for the ADH3 locus	2783	*This study*
pCAS_Pphe-_NAT_g1ADH4	NAT; pUC, 2 micron	Guide targeting for the ADH4 locus	2782	*This study*
pCAS_Pphe-_NAT_g1GPD1	NAT; pUC, 2 micron	Guide targeting for the GPD1 locus	2307	*This study*
pCAS_Pphe-_NAT_g4ADH1	NAT; pUC, 2 micron	Guide targeting for the ADH1 locus	2236	*This study*

pCAS_Pphe-_NAT_PEP4(g1)	NAT; pUC, 2 micron	Guide targeting for the PEP4 locus	2048	*This study*
pCAS_Pphe-_NAT_PBR1(g2)	NAT; pUC, 2 micron	Guide targeting for the PBR1 locus	2047	*This study*
pCAS_Pphe-_NAT_g1HIS3	NAT; pUC, 2 micron	Guide targeting for the HIS3 locus	2608	*This study*
pCAS_Pphe-_NAT_g1LEU2	NAT; pUC, 2 micron	Guide targeting for the LEU2 locus	2607	*This study*
pCAS_Pphe-_NAT_g1COS12	NAT; pUC, 2 micron	Guide targeting for the COS12 locus	2606	*This study*
pCAS_Pphe-_NAT_g2DHH1	NAT; pUC, 2 micron	Guide targeting for the DHH1 locus	2605	*This study*
pCAS_Pphe-_NAT_g1DHH1	NAT; pUC, 2 micron	Guide targeting for the DHH1 locus	2604	*This study*
pCAS_Pphe-_NAT_g1GPD2	NAT; pUC, 2 micron	Guide targeting for the GPD2 locus	2603	*This study*
CAS_Pphe-_NAT_g1YPRC_Tau3	NAT; pUC, 2 micron	Guide targeting for the YPRC_Tau3 locus	3046	*This study*
pCAS_Pphe-_NAT_g1YPRC_Delta15	NAT; pUC, 2 micron	Guide targeting for the YPRC_Delta15 locus	3045	*This study*

Appendix 4.3: Oligonucleotides used for plasmid and strain construction

A. Repair fragments and primers that were used to generated host strains

Repair fragments for knockouts: These sequences were ordered as a single stranded ultramer from IDT. It contained 50 bp homology upstream and downstream sequences for recombination. A stop codon TAA was added after the upstream homology sequence. These sequences were in upper case letters. A random 20 bp bar code sequence was added between the homology sequences, which represented by the lower case letters below. These single stranded DNA sequences were than amplified by the corresponding primers (Primer used to amplify repair fragment for knockouts) to generate a double stranded DNA, which were co-transformed with the corresponding Cas9 plasmid to generate specific knockout strains. Primers used to amplify integration fragments: These primers were used to amplify corresponding fragments for genome integration. Homology sequences to the genome integration site were in upper case letters and the lower case letters represented the annealing sequences for the amplicons.

Repair fragments for knockouts	
P357_IF_g1PEP4	ACTTGAACGCACAATATTACACTGACATTACTTTGGGTACTCCACCTCAAAACTT CAAGGTAAtcacccacaaggttgtaagaTAACGAATGTGGTTCCTTGGCTTGTTTCCTACA TTCTAAATACGATCATGAAGCTTCATC
P352_IF_g2PBR1	CCACAGAGAGCGCCTCAACCTGGGGTCCTTCAACAAGTATCTCTACGATGATGA TGCCGGTAAgttaagccaatggttagaaaATCAACCACAAGGACTTCGAAAAGAGAGCCA TTTGGGGGAAAACCATCCCACTTAACGAC
P617_RF_GPD1 locus	TATATTGTACACCCCCCCCCTCCACAAACACAAATATTGATAATATAAAGgcagacat ctTAAATTTATTGGAGAAAGATAACATATCATACTTTCCCCCACTTTTTTCGAGG
P687_GCY1_RF	TTAGCAAGCTAAAATTTGGACAGCTCTCATTACTAAATTAAGATAGAAAAagctgcga caTAATTGTTTTTGCGTGTTTCTCGTATGATTGTAATATGTAGATAAATTAAACA
P684_ADH6_RF	ATCCACATTCGAGGAAGAAATTCAACACAACAACAAGAAAAGCCAAAATCgccgtct ggaTAAGTTGTCAAGCTCTTGATAAATGTAGCTCCTTTCTTTTTAACTGCTCCATG
P681_ADH5_RF	AAGATACCTAAGAAAATTATTTAACTACATATCTACAAAATCAAAGCATCggccgtaa atTAATCTTTTGTAACGAATTTGATGAATATATTTTTACTTTTTATATAAGCTAT
P809_GPD2_RF	AGATTCAATTCTCTTTCCCTTTCCTTTTCCTTCGCTCCCCTTCCTTATCActaacgctc gACACTCTCCCCCCCCCTCCCCCTCTGATCTTTCCTGTTGCCTCTTTTTCC
P367_PEP4_IF2	ACTTGAACGCACAATATTACACTGACATTACTTTGGGTACTCCACCTCAAAACTT CAAGGTTATTTTGGATACTGGTTCTtaatcacccacaaggttgtaagaCGAATGTGGTTCCT TGGCTTGTTTCCTACATTCTAAATACGATCATGAAGCTTCATC
P366_PBR1_IF2	CCACAGAGAGCGCCTCAACCTGGGGTCCTTCAACAAGTATCTCTACGATGATGA TGCCGGTCGCGGTGTCACGTCCTATGtaagttaagccaatggttagaaaAACCACAAGGAC TTCGAAAAGAGAGCCATTTGGGGGAAAACCATCCCACTTAACGAC
P838_COS12_RF_2	AGGACGTAATAACTGCAAAATAATGTCTCCTGAACTACATCGCCATAGGCggtaggt atgTAATTGGTAAAGATATTGATATACTATTCTTAAAGACCAAAAAAAAGCTGTTA
P837_GPD2_RF2	AGATTCAATTCTCTTTCCCTTTCCTTTTCCTTCGCTCCCCTTCCTTATCActaacgctc gTAAACACTCTCCCCCCCCCCTCCCCCTCTGATCTTTCCTGTTGCCTCTTTTTCC
P838_COS12_RF_2	AGGACGTAATAACTGCAAAATAATGTCTCCTGAACTACATCGCCATAGGCggtaggt atgTAATTGGTAAAGATATTGATATACTATTCTTAAAGACCAAAAAAAAGCTGTTA
P837_GPD2_RF2	AGATTCAATTCTCTTTCCCTTTCCTTTTCCTTCGCTCCCCTTCCTTATCActaacgctc gTAAACACTCTCCCCCCCCCCTCCCCCTCTGATCTTTCCTGTTGCCTCTTTTTCC
P841_ADH1_RF_Full	GCACAATATTTCAAGCTATACCAAGCATACAATCAACTATCTCATATACAtgaacaag gtTAAGCGAATTTCTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAA
P531_g4ADH1 IF2	gttaagggctggaagatcggtgactacgccggtatcaaatggttgaacggagttatcctgTAAaactgtcctcacgct gacttgtctggttacacccacgacggttctttcca

Primer used to amplify repair fragment for knockouts	
P359_IF_g1PEP4_R	GATGAAGCTTCATGATCGTATTTAGAATGTAGG
P358_IF_g1PEP4_F	ACTTGAACGCACAATATTACACTGACAT
P830_COS12_RF_R	TAACAGCTTTTTTTTGGTCTTTAAGAATAGTATATC

P829_COS12_RF_F	AGGACGTAATAACTGCAAAATAATGTCTC
P843_ADH1_RF_Full_R	TTTTTTATAACTTATTTAATAATAAAAATCATAAATCATAAGAAATTCGCTTAac
P842_ADH1_RF_Full_F	GCACAATATTTCAAGCTATACCAAGCATAC
P811_GPD2_RF_R	GGAAAAAGAGGCAACAGGAAAGATC
P810_GPD2_RF_F	AGATTCAATTCTCTTTCCCTTTCCTTTTC
P689_GCY1_RF_R	TGTTTAATTTATCTACATATTACAATCATACGAGAAACACG
P688_GCY1_RF_F	TTAGCAAGCTAAAATTTGGACAGCTCTC
P686_ADH6_RF_R	CATGGAGCAGTTAAAAAGAAAGGAGCTA
P685_ADH6_RF_F	ATCCACATTCGAGGAAGAAATTCAACAC
P683_ADH5_RF_R	ATAGCTTATATAAAAAGTAAAAATATATTCATCAAATTCGTTACAAAAGA
P682_ADH5_RF_F	AAGATACCTAAGAAAATTATTTAACTACATATCTACAAAATCAAAGC
P517_g4ADH1_IF_R	tggaaagaaccgtcgtgggt
P516_g4ADH1_IF_F	gttaagggctggaagatcggtga
P354_IF_g2PBR1_R	GTCGTTAAGTGGGATGGTTTTCC
P353_IF_g2PBR1_F	CCACAGAGAGCGCCTCAAC

Primers used to amplify integration fragments	
P1173_ADH4_eutE_IG_F	CCATCAACAACAAGTTTACATTTGCAACAACTAATAGTCAAATAAGAAAAgaatgctactattttggagattaatctcag
P1174_ADH4_eutE_IG_R	AATAAATAAGGCACACGCATAATTGACGTTTATGAGTTCGTTCGATTTTTttaaacaatgcgaaacgcatcg
P1189_ADH3_pdc_integrate_F	TCTGTTCACAGTTAAAACTAGGAATAGTATAGTCATAAGTTAACACCATCccaactggcaccgctggc
P1190_ADH3_pdc_integrate_R	ATCATTATAAACAAAGACTTTCATAAAAAGTTTGGGTGCGTAACACGCTActagaggagcttgccccatttgacc
P1301_Tau3_2799_Int_F	GAGATATCTGCAATAAAAGCAAAAGTAAGTTTGATAGCAAGAGGTTGTTGagcgactcatgctatacctgag
P1302_Tau3_2799_Int_R	ACTCGGCATACCATATTGGTAACGCTGTATTGGAGAGATATATTCTAAAActtcgagcgtcccaaaaccttc
P1293_YPRC_D_15_800Intergration_F	AAAATTAACTATCATCTATTGACTAGTATTCATATATGACGTAATAAAATagcgacctcatgctatacctgag
P1294_YPRC_D_15_800Intergration_R	TTACAAGTTACGGTAAACATTTCAACACACCGTTATTTAACGAATTTATTtcttcgagcgtcccaaaaccttc
P801_Ter_AdhE2_ADH6_F	ATCCACATTCGAGGAAGAAATTCAACACAACAACAAGAAAAGCCAAAATCagcgactcatgctatacctgag
P802_Ter_AdhE2_ADH6_R	CATGGAGCAGTTAAAAAGAAAGGAGCTACATTTATCAAGAGCTTGACAACcttcgagcgtcccaaaaccttc
P799_PhaA_hbd_Crt_ADH5_F	AAGATACCTAAGAAAATTATTTAACTACATATCTACAAAATCAAAGCATCagcgacctcatgctatacctgag
P800_PhaA_hbd_Crt_ADH5_R	ATAGCTTATATAAAAAGTAAAAATATATTCATCAAATTCGTTACAAAAGAtcttcgagcgtcccaaaacc
P545_GPD1_AdhE2_InF	TATATTGTACACCCCCCCCCTCCACAAACACAAATATTGATAATATAAAGatgaaagtcacgaaccagaaggaac
P546_GPD1_AdhE2_InR	CCTCGAAAAAAGTGGGGGAAAGTATGATATGTTATCTTTCTCCAATAAATTtaaaaagatttgatataaatgtctttcagctcagagatc
P540_ADH1_gTer_ADH1_In_F	GCACAATATTTCAAGCTATACCAAGCATACAATCAACTATCTCATATACAatgattgttaagccaatggttagaaacaacattt
P541_ADH1_gTer_ADH1_In_R	TTTTTTATAACTTATTTAATAATAAAAATCATAAATCATAAGAAATTCGCttaaattctgtcgaatctttcaacttcagcttc

B. Primers used to construct Cas9 plasmids to target specific sites

All specific guide sequences were in upper case letters and overlap with backbone plasmids for Gibson cloning were in lower case letters.

Primer name	Sequences
P513_g4ADH1_NAT_F	gcaacaccttcgggtggcgaatgggactttGCCTGTGAATACTGTGAATTgttttagagctagaaatagcaagttaaaat
P514_g4ADH1_NAT_R	attttaacttgctatttctagctctaaaacAATTCACAGTATTCACAGGCaaagtcccattcgccacccgaaggtgttgc
P542_g1GPD1_NAT_F	accttcgggtggcgaatgggactttAGAGCTATCTCCTGTCTAAAgttttagagctagaaatagcaagttaaaat
P543_g1GPD1_NAT_R	attttaacttgctatttctagctctaaaacTTTAGACAGGAGATAGCTCTaaagtcccattcgccacccgaaggt
P679_Cas9_g3GCY1_F	accttcgggtggcgaatgggactttAGTGTGCCAACAAAGAAGGAgttttagagctagaaatagcaagtt
P680_Cas9_g3GCY1_R	aacttgctatttctagctctaaaacTCCTTCTTTGTTGGCACACTaaagtcccattcgccacccgaaggt
P677_Cas9_g2GCY1_F	accttcgggtggcgaatgggactttGGTTTTGATGAAATTCCAATgttttagagctagaaatagcaagtt
P678_Cas9_g2GCY1_R	aacttgctatttctagctctaaaacATTGGAATTTCATCAAAACCaaagtcccattcgccacccgaaggt
P675_Cas9_g1GCY1_F	accttcgggtggcgaatgggactttGGAGCATCGGTACTACCTAAgttttagagctagaaatagcaagtt
P676_Cas9_g1GCY1_R	aacttgctatttctagctctaaaacTTAGGTAGTACCGATGCTCCaaagtcccattcgccacccgaaggt
P673_Cas9_g3ADH6_F	accttcgggtggcgaatgggactttGCGTCCATGAAGCCTTCGAAgttttagagctagaaatagcaagtt
P674_Cas9_g3ADH6_R	aacttgctatttctagctctaaaacTTCGAAGGCTTCATGGACGCaaagtcccattcgccacccgaaggt
P671_Cas9_g2ADH6_F	accttcgggtggcgaatgggactttATTTCATGACCAACGACTAGgttttagagctagaaatagcaagtt
P672_Cas9_g2ADH6_R	aacttgctatttctagctctaaaacCTAGTCGTTGGTCATGAAATaaagtcccattcgccacccgaaggt
P669_Cas9_g1ADH6_F	accttcgggtggcgaatgggactttGCTGCTCCACTATTATGTGGgttttagagctagaaatagcaagtt
P670_Cas9_g1ADH6_R	aacttgctatttctagctctaaaacCCACATAATAGTGGAGCAGCaaagtcccattcgccacccgaaggt
P667_Cas9_g3ADH5_F	accttcgggtggcgaatgggactttAAGTTATTTGAACAATTAGGgttttagagctagaaatagcaagtt
P668_Cas9_g3ADH5_R	aacttgctatttctagctctaaaacCCTAATTGTTCAAATAACTTaaagtcccattcgccacccgaaggt
P665_Cas9_g2ADH5_F	accttcgggtggcgaatgggactttCAGCTATCGAGGCTTCTACGgttttagagctagaaatagcaagtt
P666_Cas9_g2ADH5_R	aacttgctatttctagctctaaaacCGTAGAAGCCTCGATAGCTGaaagtcccattcgccacccgaaggt
P663_Cas9_g1ADH5_F	accttcgggtggcgaatgggactttGACCCTGTAACCCATAGCAAgttttagagctagaaatagcaagtt
P664_Cas9_g1ADH5_R	aacttgctatttctagctctaaaacTTGCTATGGGTTACAGGGTCaaagtcccattcgccacccgaaggt
P779_g1COS12_F	accttcgggtggcgaatgggactttGCTAATGCCAAGGTACCTGAgttttagagctagaaatagcaagtt
P780_g1COS12_R	aacttgctatttctagctctaaaacTCAGGTACCTTGGCATTAGCaaagtcccattcgccacccgaaggt
P807_g1GPD2_F	accttcgggtggcgaatgggactttTTAACGGTCAATCCGCCCAAgttttagagctagaaatagcaagtt
P808_g1GPD2_R	aacttgctatttctagctctaaaacTTGGGCGGATTGACCGTTAAaaagtcccattcgccacccgaaggt
P846_NAT_g2DHH1_F	accttcgggtggcgaatgggactttGATGATGTCTTAAATACAAAgttttagagctagaaatagcaagttaaaat
P847_NAT_g2DHH1_R	attttaacttgctatttctagctctaaaacTTTGTATTTAAGACATCATCaaagtcccattcgccacccgaaggt
P844_NAT_g1DHH1_F	accttcgggtggcgaatgggactttTCTTGGCTAGTAATTCGACAgttttagagctagaaatagcaagtt
P845 _NAT_g1DHH1_R	aacttgctatttctagctctaaaacTGTCGAATTACTAGCCAAGAaaagtcccattcgccacccgaaggt
P1171_g1ADH4_F	cgggtggcgaatgggactttTTAGTCGCTGCATACAAAGAgttttagagctagaaatagc
P1172_g1ADH4_R	gctatttctagctctaaaacTCTTTGTATGCAGCGACTAAaaagtcccattcgccacccg
P1187_g1ADH3_F	cgggtggcgaatgggactttGGGCAAACCAACCAAAACGAgttttagagctagaaatagc
P1188_g1ADH3_R	attttaacttgctatttctagctctaaaacTCGTTTTGGTTGGTTTGCCCaaagtcccattcgccacccg

P1299_g1YPRC_Tau3_F	cgggtggcgaatgggactttATAATTAATGTTGAACCAATgttttagagctagaaatagc
P1300_g1YPRC_Tau3_R	gctatttctagctctaaaacATTGGTTCAACATTAATTATaaagtcccattcgccacccg
P1295_g1YPRC_D_15F	cgggtggcgaatgggactttATATCCTCAGAGAGAATTTTgttttagagctagaaatagc
P1296_g1YPRC_D_15R	gctatttctagctctaaaacAAAATTCTCTCTGAGGATATaaagtcccattcgccacccg

C. Primers used to genotype knockout and integrated strains after CRISPR-Cas9 editing

Name	Sequence
P362_PBR1_colony_F	GAAGACGCTTTCTTCATTTCTACTAAAGACACCTC
P363_PBR1_colony_R	CCCTTTTTCTTTTCTTGGGCTTCTTTTTGGTG
P364_PEP4_Colony_F	AAAATTTATAAACACGAGTTGTCCGATGAGATGAAAGAAG
P365_PEP4_Colony_R	CAAACCCAAAATACCATCGAACTTGCCAAATG
P567_GPD1locusColonyPCR_F	CTTACTCTCCTACATAAGACATCAAGAAACAATTG
P568_GPD1locusColonyPCR_R	CCTCGAAAAAAGTGGGGGAAAGTATG
P690_ADH5_knockout_ColonyPCR_F	AATCAAATTGTGACATCTGCTGACGC
P691_ADH5_knockout_ColonyPCR_R	GTAAGGCAAAATACCAAATGTCCACC
P692_ADH6_knockout_ColonyPCR_F	GCCAATTTTTCACATCTGGAAGCG
P693_ADH6_knockout_ColonyPCR_R	TTAAAGGTGCTTAGCAAGGAGAAAAAGAG
P694_GCY1_knockout_ColonyPCR_F	CGCTGCTCTCCTTAATTCCCTAGAG
P695_GCY1_knockout_ColonyPCR_F	GCAGGTAAAGTTTTCTTGCCTTATACACC
P696_COS12p_F	AGAAGTGCTGTAGGGCTAAAGAACAG
P697_COS12p_R	ATTTCAGATGGTAAAAAAGCTACAGTATTTTCAAATTTG
P812_GPD2_ColonyPCR_F	CACTAAGCTTTTTCCTTGATTTATCCTTGGG
P813_GPD2_ColonyPCR_R	TGTAAACGATAATAGCGTGTATAATGGTAGTTATGTATATATAG
P875_COS12ko_colonyPCRF	GATAGATGATGGTTTTAGGAAACATATGAACGAAC
P876_COS12ko_colonyPCRR	CAGACTTCATCAAATATTGCATTATCTACTAATGTTTAAAC
P877_DHH1ko_colonyPCRF	CCTTTTATTTCTTCATAACCGCATCGCC
P878_DHH1ko_colonyPCRR	CACAATGGAGATTTGAAAAAAGATAAAAATAATCGACG
P1184_ADH4_ColonyPCR_F	CATTTCTGGTTTATTAAAGACTGGAGTCAAACG
P1185_ADH4_ColonyPCR_R	CTCGAAATTAACGTAATTATATAGATCGTGAAAAGTTAAAAAAATC
P1191_ADH3_integration_colonyPCR_F	GCTTTATCTCTTCGACCGAATTTACTATACATGG
P1192_ADH3_integration_colonyPCR_R	GATATAGAAAAAATACTGGTACTGCTTCTTGATTTAGTG
P1239_ADH3_integration_colonyPCR_F2	GATAATGGCTAAGGCAAGCAGTCCG
P1240_ADH3_integration_colonyPCR_R2	TTGATGGTGATAATGTCTCTCAAACGTTCTATGTG
P1297_YPRC_ColonyPCR _F	CAGAGCATAGGGTTTCGCAAACAAAC
P1298_YPRC_ColonyPCR _R	CTTGTATATGCTCATCCCGACCTTCC

D. qPCR primers.

Name	Sequence
P436_qgTdTerR	tgggtagcttctggaccaat
P435_qgTdTerF	aagaccgttgacccattcac
P434_qgTdTerR	cggtagcttccaagtgttcc
P433_qgTdTerF	ccagctaacgacgaagaagc
P432_qgTdTerR	tttcgtcagagaaagcgtca
P431_qgTdTerF	cggttacggtttggcttcta
P430_qgTdTerR	ctgggtcggttctaactgga
P429_qgTdTerF	tacggtaccccaggttggta
P428_qTDH3R	caacagcgtcttcggtgtaa
P427_qTDH3F	aggctgtcggtaaggtcttg
P426_qTDH3R	agtggagtcaatggcgatgt
P425_qTDH3F	tttgaacgacccattcatca
P424_qTDH3_R	cgatgtcaacgttggaagaa
P423_qTDH3F	gttgctttgaacgacccatt
P422_qTDH3R	aacaaccttcttggcaccag
P421_qTDH3F	ctggtgaagtttcccacgat

E. Primers used for plasmid construction

Name	Sequence
P51_5'UTR CDC19_F	TCGAATTCAACCCTCACTAAAGGGCGGCCGCCCAATCAAAACAAATAAAACATCATCACAATGATCGTCAAGCCAATGGTGC
P37_5'UTR R	GATCTTATCGTCGTCATCCTTGTAATCCATCGATACTAGTTTAAATACGATCGAAACGTTCAACTTCTG
P36_5'UTRYJL177W_F	GAATTCAACCCTCACTAAAGGGCGGCCGCTTGGTGAATCAAAAAATTAACGAAACGAACAAATTTAAAATGATCGTCAAGCCAATGGTGC
P37_5'UTR R	GATCTTATCGTCGTCATCCTTGTAATCCATCGATACTAGTTTAAATACGATCGAAACGTTCAACTTCTG
P32_5'UTRFBA1_F	GAAAATTCGAATTCAACCCTCACTAAAGGGCGGCCGCACATATTCAAAATGATCGTCAAGCCAATGGTGC
P37_5'UTR R	GATCTTATCGTCGTCATCCTTGTAATCCATCGATACTAGTTTAAATACGATCGAAACGTTCAACTTCTG
P36_5'UTRYJL177W_F	GAATTCAACCCTCACTAAAGGGCGGCCGCTTGGTGAATCAAAAAATTAACGAAACGAACAAATTTAAAATGATCGTCAAGCCAATGGTGC
P37_5'UTR R	GATCTTATCGTCGTCATCCTTGTAATCCATCGATACTAGTTTAAATACGATCGAAACGTTCAACTTCTG
P38_3'UTR FBA1R	cgtcgtcatccttgtaatccatcgatactagtaaaactatatcaattaatttgaattaacttaaatacgatcgaaacgttcaacttctgc
P37_5'UTR R	GATCTTATCGTCGTCATCCTTGTAATCCATCGATACTAGTTTAAATACGATCGAAACGTTCAACTTCTG
P36_5'UTRYJL177W_F	GAATTCAACCCTCACTAAAGGGCGGCCGCTTGGTGAATCAAAAAATTAACGAAACGAACAAATTTAAAATGATCGTCAAGCCAATGGTGC

P37_5'UTR R	GATCTTATCGTCGTCATCCTTGTAATCCATCGATACTAGT TTAAATACGATCGAAACGTTCAACTTCTG
P35_5'UTRYHL001W_F	GAAAATTCGAATTCAACCCTCACTAAAGGGCGGCCGCG CGCAAATAAACCAAAAATGATCGTCAAGCCAATGGTGC
P37_5'UTR R	GATCTTATCGTCGTCATCCTTGTAATCCATCGATACTAGT TTAAATACGATCGAAACGTTCAACTTCTG
P34_5'UTRYLRO75W_F	AATTCGAATTCAACCCTCACTAAAGGGCGGCCGCGTACA GTATATCAAATAACTAATTCAAGATGATCGTCAAGCCAAT GGTGC
P37_5'UTR R	GATCTTATCGTCGTCATCCTTGTAATCCATCGATACTAGT TTAAATACGATCGAAACGTTCAACTTCTG
P33_5'UTRGPM1_F	CGAATTCAACCCTCACTAAAGGGCGGCCGCATATTACAA TAATGATCGTCAAGCCAATGGTGC
P37_5'UTR R	GATCTTATCGTCGTCATCCTTGTAATCCATCGATACTAGT TTAAATACGATCGAAACGTTCAACTTCTG
P32_5'UTRFBA1_F	GAAAATTCGAATTCAACCCTCACTAAAGGGCGGCCGCAC ATATTCAAAATGATCGTCAAGCCAATGGTGC
P37_5'UTR R	GATCTTATCGTCGTCATCCTTGTAATCCATCGATACTAGT TTAAATACGATCGAAACGTTCAACTTCTG
P31_5'UTRTDH2_F	CGAATTCAACCCTCACTAAAGGGCGGCCGCAATTAAATT CATCACACAAACAAACAAAACAAAATGATCGTCAAGCCA ATGGTGC
P37_5'UTR R	GATCTTATCGTCGTCATCCTTGTAATCCATCGATACTAGT TTAAATACGATCGAAACGTTCAACTTCTG
P30_5'UTRTPI1_F	CGAATTCAACCCTCACTAAAGGGCGGCCGCTAACTACAA AAAACACATACATAAACTAAAAATGATCGTCAAGCCAATG GTGC
P37_5'UTR R	GATCTTATCGTCGTCATCCTTGTAATCCATCGATACTAGT TTAAATACGATCGAAACGTTCAACTTCTG
P52_5'UTR TDH3_F	TTCGAATTCAACCCTCACTAAAGGGCGGCCGCCCAAGA ACTTAGTTTCGAATAAACACACATAAACAAACAAAATGAT CGTCAAGCCAATGGTGC
P37_5'UTR R	GATCTTATCGTCGTCATCCTTGTAATCCATCGATACTAGT TTAAATACGATCGAAACGTTCAACTTCTG
P39_3'UTR F	GAAAATTCGAATTCAACCCTCACTAAAGGGCGGCCGCAT GATCGTCAAGCCAATGGTGC
P38_3'UTR FBA1R	cgtcgtcatccttgtaatccatcgatactagtaaaactatatcaattaatttgaattaact taaatacgatcgaaacgttcaacttctgc
P39_3'UTR F	GAAAATTCGAATTCAACCCTCACTAAAGGGCGGCCGCAT GATCGTCAAGCCAATGGTGC
P44_3'UTR YJL177WgDNA_TerR	TAAAAGATTTTAAAATTAAAAAAGCATTTAAATACGATCG AAACGTTCAACTTCTGC
P43_3'UTR YJL177WgDNA_R	TCATCCTTGTAATCCATCGATACTAGTTTGATTTTGATTC TGTGTATTGGCCTAAAC
P42_3'UTR YJL177WgDNA_F	GGCAGAAGTTGAACGTTTCGATCGTATTTAAATGCTTTTT TAATTTTAAAATCTTTTAAAGTGAATATTTGATTT
P38_3'UTR FBA1R	cgtcgtcatccttgtaatccatcgatactagtaaaactatatcaattaatttgaattaact taaatacgatcgaaacgttcaacttctgc
P32_5'UTRFBA1_F	GAAAATTCGAATTCAACCCTCACTAAAGGGCGGCCGCAC ATATTCAAAATGATCGTCAAGCCAATGGTGC
P32_5'UTRFBA1_F	GAAAATTCGAATTCAACCCTCACTAAAGGGCGGCCGCAC ATATTCAAAATGATCGTCAAGCCAATGGTGC
P44_3'UTR YJL177WgDNA_TerR	TAAAAGATTTTAAAATTAAAAAAGCATTTAAATACGATCG AAACGTTCAACTTCTGC

P43_3'UTR YJL177WgDNA_R	TCATCCTTGTAATCCATCGATACTAGTTTGATTTTGATTCTGTGTATTGGCCTAAAC
P42_3'UTR YJL177WgDNA_F	GGCAGAAGTTGAACGTTTCGATCGTATTTAAATGCTTTTTTAATTTTAAAATCTTTTAAAGTGAATATTTGATTT
P657_YPK2_AdhE2_R	gttccttctggttcgtgactttcatcgatagtgcttttgttgtaatcttacaatgg
P656_YPK2_AdhE2_F	atccgtaatacgactcactatagggcccgggcaaagaacataaaacattttgaagcagagcg
P3_P(Tef1)F1	AATGCCTATTATGCAGATGTTATAATATCTGTGCGTGGATCCCAAAATGTTTCTACTCCTTTTTTACTCTTC
P4_P(Tef1)R1	CACAATGTAAACCTCTGTGTTAACCATTTTGTAATTAAAACTTAGATTAGATTGCTATGCTTTC
P1_Erg10F1	GAAAGCATAGCAATCTAATCTAAGTTTTAATTACAAAATGGTTAACACAGAGGTTTACATTGTGTCTGC
P2_Erg10R1	CTGGCGAAGAATTGTTAATTAAGAGCTCATGATGATGATGATGATGATGATGATGATGAACTCTTTCGATAACAATGGATGAAGCACC
P23_Erg10_HisR4	cttgaccaaacctctggcgaagaattgttaattaagagctcctaaactctttcgataacaatggatgaagcacc
P22_Erg10_His_R3	gaccaaacctctggcgaagaattgttaattaagagctcctaatgatgatgatgatgatgatgatgatgatgaactctttcgataacaatggatgaagcac
P46_MECR1 His_F	ATTCGAATTCAACCCTCACTAAAGGGCGGCCGCATGCATCATCATCATCATCATCATCATCACATGTCTCTGCCGGCGAAGC
P45_MECR1_F	AATTCGAATTCAACCCTCACTAAAGGGCGGCCGCATGTCTCTGCCGGCGAAGC
P49_MECR1_R	gtcgtcatccttgtaatccatcgatactagtttacagggtgaagagaaccttgc
P761_TPS3t_Adh22_Aldh12_F	TTCCATCTATGAAGCTGCCCTGTAAgcgatcattttccctcctgtacttttc
P760_Adh22_Aldh12_R	aaagtacaggagggaaaatgatcgcTTACAGGGCAGCTTCATAGATGGAAAC
P759_Adh22_Aldh12_F	ctttaacgtcaaggagaaaaaaccccggatccATGAATAACTTCACCTACAGCATCCC
P758_TPS3t_Adh8_Aldh12_F	GAAATCTACCGTGCTGCTCTGTAAgcgatcattttccctcctgtacttttc
P757_Adh8_Aldh12_R	gaaagtacaggagggaaaatgatcgcTTACAGAGCAGCACGGTAGATTTC
P756_Adh8_Aldh12_F	ctttaacgtcaaggagaaaaaaccccggatccATGTACGACTTCATGTTCCACGTAC
P755_TPS3t_Adh2_Aldh12_F	CGTTCTGGAAGAAATTCTGAACCTGGCTTACTAAgcgatcatttttccctcctgtactttc
P754_Adh2_Aldh12_R	aaagtacaggagggaaaatgatcgcTTAGTAAGCCAGGTTCAGAATTTCTTCCAG
P753_Adh2_Aldh12_F	ctttaacgtcaaggagaaaaaaccccggatccATGGTCAACTTTTCCTACTGCAATCC
P752_TPS3t_Adh22_Aldh10_F	TTCCATCTATGAAGCTGCCCTGTAAgggcccggggcgatcattttcctcctgtactttc
P751_Adh22_Aldh10_R	aagtacaggagggaaaatgatcgccccgggcccTTACAGGGCAGCTTCATAGATGGAAAC
P750_Adh22_Aldh10_F	ctttaacgtcaaggagaaaaaaccccggatccATGAATAACTTCACCTACAGCATCCC
P749_TPS3t_Adh8_Aldh10_F	AGAAATCTACCGTGCTGCTCTGTAAgggcccggggcgatcattttcctcctgtactttc
P748_Adh8_Aldh10_R	aagtacaggagggaaaatgatcgccccgggcccTTACAGAGCAGCACGGTAGATTTCTAC

P747_Adh8_Aldh10_F	ctttaacgtcaaggagaaaaaaccccggatccATGTACGACTTCATGTTCCACGTAC
P746_TPS3t_Adh2_Aldh10_F	AGAAATTCTGAACCTGGCTTACTAAgggcccggggcgatcattttcctcctgtactttc
P745_Adh2_Aldh10_R	agtacaggagggaaaatgatcgccccgggcccTTAGTAAGCCAGGTTCAGAATTTCTTCC
P744_Adh2_Aldh10_F	ctttaacgtcaaggagaaaaaaccccggatccATGGTCAACTTTTCCTACTGCAATCC
P743_TPS3t_Adh22_Aldh7_F	TTCCATCTATGAAGCTGCCCTGTAAgggcccggggcgatcattttcctcctgtactttc
P742_Adh22_Aldh7_R	aagtacaggagggaaaatgatcgccccgggcccTTACAGGGCAGCTTCATAGATGGAAAC
P741_Adh22_Aldh7_F	ctttaacgtcaaggagaaaaaaccccggatccATGAATAACTTCACCTACAGCATCCCG
P740_TPS3t_Adh8_Aldh7_F	AGAAATCTACCGTGCTGCTCTGTAAgggcccggggcgatcattttcctcctgtactttc
P739_Adh8_Aldh7_R	gaaagtacaggagggaaaatgatcgccccgggcccTTACAGAGCAGCACGGTAGATTTC
P738_Adh8_Aldh7_F	ctttaacgtcaaggagaaaaaaccccggatccATGTACGACTTCATGTTCCACGTAC
P737_TPS3t_Adh2_Aldh7_F	AGAAATTCTGAACCTGGCTTACTAAgggcccggggcgatcattttcctcctgtactttc
P736_Adh2_Aldh7_R	agtacaggagggaaaatgatcgccccgggcccTTAGTAAGCCAGGTTCAGAATTTCTTCC
P735_Adh2_Aldh7_F	tttaacgtcaaggagaaaaaaccccggatccATGGTCAACTTTTCCTACTGCAATCCAAC
P734_TPS3t_Adh22_Aldh6_F	TTCCATCTATGAAGCTGCCCTGTAAgggcccggggcgatcattttcctcctgtactttc
P733_Adh22_Aldh6_R	aagtacaggagggaaaatgatcgccccgggcccTTACAGGGCAGCTTCATAGATGGAAAC
P732_Adh22_Aldh6_F	ctttaacgtcaaggagaaaaaaccccggatccATGAATAACTTCACCTACAGCATCCCG
P731_TPS3t_Adh8_Aldh6_F	AGAAATCTACCGTGCTGCTCTGTAAgggcccggggcgatcattttcctcctgtactttc
P730_Adh8_Aldh6_R	aagtacaggagggaaaatgatcgccccgggcccTTACAGAGCAGCACGGTAGATTTCTAC
P729_Adh8_Aldh6_F	ctttaacgtcaaggagaaaaaaccccggatccATGTACGACTTCATGTTCCACGTAC
P728_TPS3t_Adh2_Aldh6_F	AGAAATTCTGAACCTGGCTTACTAAgggcccggggcgatcattttcctcctgtactttc
P727_Adh2_Aldh6_R	agtacaggagggaaaatgatcgccccgggcccTTAGTAAGCCAGGTTCAGAATTTCTTCC
P726_Adh2_Aldh6_F	ctttaacgtcaaggagaaaaaaccccggatccATGGTCAACTTTTCCTACTGCAATCC
P725_TPS3t_Adh22_Aldh5_F	TTCCATCTATGAAGCTGCCCTGTAAgtcgacgcgatcattttccctcctgtactttc
P724_Adh22_Aldh5_R	aaagtacaggagggaaaatgatcgcgtcgacTTACAGGGCAGCTTCATAGATGGAAACG
P723_Adh22_Aldh5_F	tttaacgtcaaggagaaaaaaccccggatccATGAATAACTTCACCTACAGCATCCCGAC
P722_TPS3t_Adh8_Aldh5_F	AGAAATCTACCGTGCTGCTCTGTAAgtcgacgcgatcattttccctcctgtactttc
P721_Adh8_Aldh5_R	ttgaaagtacaggagggaaaatgatcgcgtcgacTTACAGAGCAGCACGGTAGATTTC

P720_Adh8_Aldh5_F	ctttaacgtcaaggagaaaaaaccccggatccATGTACGACTTCATGTTCCACGTACC
P719_TPS3t_Adh2_Aldh5_R	gtgctatattcttgttgctaccgtcctcgagtccattacccaacatcttggtgatttg
P718_TPS3t_Adh2_Aldh5_F	AGAAATTCTGAACCTGGCTTACTAAgtcgacgcgatcattttccctcctgtactttc
P717_Adh2_Aldh5_R	aagtacaggagggaaaatgatcgcgtcgacTTAGTAAGCCAGGTTCAGAATTTCTTCCAG
P716_Adh2_Aldh5_F	ctttaacgtcaaggagaaaaaaccccggatccATGGTCAACTTTTCCTACTGCAATCC
P1128_Gal7_Aldh21_adh2_R	GGGTTTCCAGTTCCGCGGTGTTCATccatggtttgagggaatattcaactgttttttttatcatgttg
P1127_Gal7_Aldh21_adh2_F	aatcaccaagatgttgggtaatggactcgaggacggtagcaacaagaatatagcacg
P1130_Aldh21_R	tcttagctagccgcggtaccaagcttacatatgTTAACGAATAGAGAAGCCGTTGGTCAG
P1129_Aldh21_F	aaaaacagttgaatattccctcaaaccatggATGAACACCGCGGAACTGGAAAC
P1206_ADH3_aldh21_F	ctttaacgtcaaggagaaaaaaccccggatccATGATTAACTTCGACTATTGCGTGCC
P1207_ADH3_aldh21_R	aagtacaggagggaaaatgatcgccccgggcccTTAGCGGGCCATTTCCAGAATTTTTAC
P1208_ADH4_aldh21_F	ttaacgtcaaggagaaaaaaccccggatccATGCAGAAATTTGACTACTATACTCCGACC
P1209_ADH4_aldh21_R	agtacaggagggaaaatgatcgccccgggcccTTATTTGTTCGCAGATACATAGATCGGG
P1210_ADH5_aldh21_F	tactttaacgtcaaggagaaaaaaccccggatccATGGAAAACTTCACCTACTACAACCC
P1211_ADH5_aldh21_R	aaagtacaggagggaaaatgatcgccccgggcccTTACAGAGATGCACGCAGGATC
P1212_ADH6_aldh21_F	tttaacgtcaaggagaaaaaaccccggatccATGAACAACTTCCTGTTCGAAAACAAAAC
P1213_ADH6_aldh21_R	agtacaggagggaaaatgatcgccccgggcccTTATTTACATTCGTTCAGGATGTCCAGC
P1214_ADH7_aldh21_F	atactttaacgtcaaggagaaaaaaccccggatccATGCGTAACTTTACCTACCACAACC
P1215_ADH7_aldh21_R	aaagtacaggagggaaaatgatcgccccgggcccTTACAGGGCCATGTGCAGAATGTC
P1216_ADH9_aldh21_F	tttaacgtcaaggagaaaaaaccccggatccATGAATGATTTCCAGTTTCAGAACACTAC
P1217_ADH9_aldh21_R	aagtacaggagggaaaatgatcgccccgggcccTTACAGGCACATTTTATAAATGGCCAG
P1218_ADH10_aldh21_F	tactttaacgtcaaggagaaaaaaccccggatccATGCAGAATTTCGTTTTTCACAACCC
P1219_ADH10_aldh21_R	aaagtacaggagggaaaatgatcgccccgggcccTTAGCGAGAAGCGCGACGC
P1220_ADH12_aldh21_F	ctttaacgtcaaggagaaaaaaccccggatccATGCTGGGCGACTTTACCTACTC
P1221_ADH12_aldh21_R	agtacaggagggaaaatgatcgccccgggcccTTATTTCATGGATTGTTTCAGGATGGTG
P1222_ADH13_aldh21_F	tttaacgtcaaggagaaaaaaccccggatccATGGAAAATTTCGATTTCCACGTTACTAC
P1223_ADH13_aldh21_R	aaagtacaggagggaaaatgatcgccccgggcccTTAGACGAAAGAAGACTCGGACATG
P1224_ADH14_aldh21_F	tactttaacgtcaaggagaaaaaaccccggatccATGGAATCTTTCGATTTTTTCCGTCG

P1225_ADH14_aldh21_R	aaagtacaggagggaaaatgatcgccccgggcccTTAGAAAATATCGCGACCTTCGCAG
P359_IF_g1PEP4_R	GATGAAGCTTCATGATCGTATTTAGAATGTAGG
P358_IF_g1PEP4_F	ACTTGAACGCACAATATTACACTGACAT
P830_COS12_RF_R	TAACAGCTTTTTTTTGGTCTTTAAGAATAGTATATC
P829_COS12_RF_F	AGGACGTAATAACTGCAAAATAATGTCTC
P843_ADH1_RF_Full_R	TTTTTTATAACTTATTTAATAATAAAAATCATAAATCATAAGAAATTCGCTTAac
P842_ADH1_RF_Full_F	GCACAATATTTCAAGCTATACCAAGCATAC
P811_GPD2_RF_R	GGAAAAAGAGGCAACAGGAAAGATC
P810_GPD2_RF_F	AGATTCAATTCTCTTTCCCTTTCCTTTTC
P689_GCY1_RF_R	TGTTTAATTTATCTACATATTACAATCATACGAGAAACACG
P688_GCY1_RF_F	TTAGCAAGCTAAAATTTGGACAGCTCTC
P686_ADH6_RF_R	CATGGAGCAGTTAAAAAGAAAGGAGCTA
P685_ADH6_RF_F	ATCCACATTCGAGGAAGAAATTCAACAC
P683_ADH5_RF_R	ATAGCTTATATAAAAAGTAAAAATATATTCATCAAATTCGTTACAAAAGA
P682_ADH5_RF_F	AAGATACCTAAGAAAATTATTTAACTACATATCTACAAAATCAAAGC
P517_g4ADH1_IF_R	tggaaagaaccgtcgtgggt
P516_g4ADH1_IF_F	gttaagggctggaagatcggtga
P354_IF_g2PBR1_R	GTCGTTAAGTGGGATGGTTTTCC
P353_IF_g2PBR1_F	CCACAGAGAGCGCCTCAAC
P1173_ADH4_eutE_IG_F	CCATCAACAACAAGTTTACATTTGCAACAACTAATAGTCAAATAAGAAAAgaatgctactattttggagattaatctcag
P1174_ADH4_eutE_IG_R	AATAAATAAGGCACACGCATAATTGACGTTTATGAGTTCGTTCGATTTTTttaaacaatgcgaaacgcatcg
P1189_ADH3_pdc_integrate_F	TCTGTTCACAGTTAAAACTAGGAATAGTATAGTCATAAGTTAACACCATCccaactggcaccgctggc
P1190_ADH3_pdc_integrate_R	ATCATTATAAACAAAGACTTTCATAAAAAGTTTGGGTGCGTAACACGCTActagaggagcttgccccatttgacc
P1301_Tau3_2799_Int_F	GAGATATCTGCAATAAAAGCAAAAGTAAGTTTGATAGCAAGAGGTTGTTGagcgacctcatgctatacctgag
P1302_Tau3_2799_Int_R	ACTCGGCATACCATATTGGTAACGCTGTATTGGAGAGATATATTCTAAAActtcgagcgtcccaaaaccttc
P1293_YPRC_D_15_800Intergration_F	AAAATTAACTATCATCTATTGACTAGTATTCATATATGACGTAATAAAATagcgacctcatgctatacctgag
P1294_YPRC_D_15_800Intergration_R	TTACAAGTTACGGTAAACATTTCAACACACCGTTATTTAACGAATTTATTtcttcgagcgtcccaaaaccttc
P801_Ter_AdhE2_ADH6_F	ATCCACATTCGAGGAAGAAATTCAACACAACAACAAGAAAAGCCAAAATCagcgacctcatgctatacctgag
P802_Ter_AdhE2_ADH6_R	CATGGAGCAGTTAAAAAGAAAGGAGCTACATTTATCAAGAGCTTGACAACcttcgagcgtcccaaaaccttc
P799_PhaA_hbd_Crt_ADH5_F	AAGATACCTAAGAAAATTATTTAACTACATATCTACAAAATCAAAGCATCagcgacctcatgctatacctgag
P800_PhaA_hbd_Crt_ADH5_R	ATAGCTTATATAAAAAGTAAAAATATATTCATCAAATTCGTTACAAAAGAtcttcgagcgtcccaaaacc
P545_GPD1_AdhE2_InF	TATATTGTACACCCCCCCCCTCCACAAACACAAATATTGATAATATAAAGatgaaagtcacgaaccagaaggaac
P546_GPD1_AdhE2_InR	CCTCGAAAAAAGTGGGGGAAAGTATGATATGTTATCTTTCTCCAATAAATTtaaaaagatttgatataaatgtctttcagctcagagatc

P540_ADH1_gTer_ADH1_In_F	GCACAATATTTCAAGCTATACCAAGCATACAATCAACTATCTCATATACAatgattgttaagccaatggttagaaacaacattt
P541_ADH1_gTer_ADH1_In_R	TTTTTTATAACTTATTTAATAATAAAAATCATAAATCATAAGAAATTCGCttaaattctgtcgaatctttcaacttcagcttc
P208_gBlock32_SPG5R	TGGTAATAGCGCGATGAAACAACGTCTTTGTTAGAAAGACTTAATGTAAATGTCCTTCAATTCAGAAAT
P207_gBlock32_SPG5F	attttcaagtacttgccaagagcttacaag
P152_CCW12P_F	AGTATTGATAATGATAAACTCGAACTGCCGCGGTACCCAAAGCAAAATAAAAGAAACTTAATACGTTATGCCG
P151_CCW12P_R	tgtgctagtgtctcccgtcttctgtctcgagtattgatatagtgtttaagcgaatgacagaagattaatttc
P442_1.4a.1_PRM9R	cgtaaccaccacacccgccgcgcttaatgcggatccattttcaacatcgtattttccgaagcgtt
P441_1.4a.1_PRM9F	acagaagacgggagacactagcacacaactttaccaggcaaggtatttgacgc
P194_pVYY100_2SPG5R	CGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCAAAGCTTGCTTATTTTCTGCCGAATTTTCATGAAGTTTT
P193_pVYY100_2SPG5F	agaacttagtttcgaataaacacacataaacaaacaaacccgggcaaagacgttgtttcatcgcgc
P196_pVYY100_3TDH3R	ATTATTAACTCTTTTGTTTTTCTCGAGAAGCCCCGGGTTTGTTTGTTTATGTGTGTTTATTCGAAACTAAGTTC
P153_TDH3F	gtttcttttattttgctttgggtaccgcggcagttcgagtttatcattatcaatactgccatttc
P247_HIS2	GTAATAGCGCGATGAAACAACGTCTTTGAGATCTGTAACAATATCATGAGACCTTTTATAGAAGTGGCGCCAAAACTAAATGTATTTGAAAATACAAAAAACGCAC
P246_HIS1	atgaccgactctttgtgtattagataaatagattaatttaaacagtatatgtacagttttatatatatatatatatatatatacatatataaagaaacctgtgcgttttttgtattttcaaatac
P172_1C0_gTdTer	GTCATTCGCTTAAACACTATATCAATAATGATTGTTAAGCCAATGGTTAGAAACAAC
P161_110_PYK2R	tgtgctagtgtctcccgtcttctgtttaaattctgtcgaatctttcaacttcagc
P170_160_TDH3F	CATTCGCTTAAACACTATATCAATACCAAGAACTTAGTTTCGAATAAACACACATAAACAAACAAAATGATTGTTAAGCCAATGGTTAGAAACAAC
P160_110_PYK2F	TGTCATTCGCTTAAACACTATATCAATACAAAGAACATAAAACATTTTGAAGCAGAGCG
P173_180_3'VSVR	tgtgctagtgtctcccgtcttctgtacgaagacaaacaaaccattattaccattaaaaggctcaggagaaacttttaaattctgtcgaatctttcaacttcagc
P172_1C0_gTdTer	GTCATTCGCTTAAACACTATATCAATAATGATTGTTAAGCCAATGGTTAGAAACAAC
P171_170_VSVF	CATTCGCTTAAACACTATATCAATAACGAAGACCACAAAACCAGATAAAAAATAAAAACCACAAGAGGGTCTTAAATGATTGTTAAGCCAATGGTTAGAAACAAC
P170_160_TDH3F	CATTCGCTTAAACACTATATCAATACCAAGAACTTAGTTTCGAATAAACACACATAAACAAACAAAATGATTGTTAAGCCAATGGTTAGAAACAAC
P169_150_TDH2F	CATTCGCTTAAACACTATATCAATAAATTAAATTCATCACACAAACAAACAAAACAAAATGATTGTTAAGCCAATGGTTAGAAACAAC
P168_140_YHL001WF	GTCATTCGCTTAAACACTATATCAATAGCGCAAATAAACCAAAAATGATTGTTAAGCCAATGGTTAGAAACAAC
P167_130_gTdTer F	CATTGGCCAAGAACTAACCATACGCAATGATTGTTAAGCCAATGGTTAGAAACAAC
P166_130_PFK2R	ttctaaccattggcttaacaatcattgcgtatggttagttcttggcc

P165_130_PFK2F	TCATTCGCTTAAACACTATATCAATATCATTTGAACAATAGAACTAGATTTAGAGACTAGTTTAG
P164_120_gTdTerF	AAAATCTGAAACAAAATCATATCAAAGATGATTGTTAAGCCAATGGTTAGAAACAAC
P163_120_PFK1R	ttctaaccattggcttaacaatcatctttgatatgattttgtttcagatttttttatataaaagctttc
P162_120_PFK1F	GTCATTCGCTTAAACACTATATCAATATATTGCTTTCTACCAATAAAATCTGTTAATTCTATTTGG
P161_110_PYK2R	tgtgctagtgtctcccgtcttctgtttaaattctgtcgaatctttcaacttcagc
P160_110_PYK2F	TGTCATTCGCTTAAACACTATATCAATACAAAGAACATAAAACATTTTGAAGCAGAGCG
P639_903_eutE_Seq	ggaccaccatgaccatcacc
P638-Leu_BackbondR	aaaatacgatgttgaaaatggatccgcattaagcgcggcggg
63_gal1454_TDH3_R	CATCCCGGGCCCTATAGTGAGTCGTATTACGGATCCGGGGTTTTTTCTCCTTGACGTTAAAGTATAGAGG
84_pCCW12 for 1558 F	ctctgcttcaaaatgttttatgttctttggcggccgccctttagtga
P361_TDH3t_F	gaataaacacacataaacaaacaaaGGATCCGCTAGCgtgaatttacttaaatcttgcatttaaataaattttctt
P360_TDH3p_R	aatgcaagatttaaagtaaattcacGCTAGCGGATCCtttgtttgtttatgtgtgtttattcgaaactaagttc

Appendix 4.4: gBlocks

#	Name	Sequences
VYG1	EgTer_Yeast_G1	actgtacctaaatctctcttaatggcctcgaccgtccttgccttggttgtactgtcaaaagcatcacc atttaagcttcgggcgtatagtccagcctctagcgctgccttctcaaatgcaacggtgttataccatc cagctgctgctggtctacctttagttggtggaccagcaaggaagacccccagtgtagccgcttga tacccgaaagcggcggtgattcttgtagacaatccgtaccctgtagagcaaccaataactaaaa ccctttttaggacctggcgatgtaggaggatgggctctagcataagcaatctcttcttgtactctttttc acaacctattgggtgagtagttgtacatataaaacctctaatctttggttgaatcaccttagctgtagt tgtgaacattgccatgcggccgccctttagtgagggttgaattcgaattttcaaaa
VYG2	EgTer_Yeast_G2	gtaaccaaagcttttgcaacaacagggtatgcgggacatccatattgttgtgtgatgcgttttgcag ctttttctacatccttcttagcttcgccgatagttccagaccagtacactggccaagtcatttctgggc caatatatgaatatgcaactgttttagccccttctgcgagtacgccagcctcagacagggcttgga tccatagctcccaatcttcgccacccatcaccttgaccgtatcagcaatttcttcaggagaagccg gttctatactcacatcagtcacttcggctttgtctgtgttaactgtacgatttgtatatgtagcccctattg gtttaaggcaagccttatggagaacgcctgtggctggatccgttctctttggtgcagcaatgctgtat accaccaaatcaactgtacctaaatctctcttaatggcctcgaccgtccttg
VYG3	EgTer_Yeast_G3	cagatcttatcgtcgtcatccttgtaatccatcgatactagtttactgttgtgctgctgaaggaagatc agcttcaacgtcaacaggttgatcgtagtcgactccgtcaatgccgaacccaaacagcctaagg aattctgtctgataacctgcgaaatcacttatatcttttaagttggcagtggacacttgtgaccacaa atctttaacagcctgctggacatcttcagccatttcccaatcatcaactctaacacgtcccgcttcat caacaataggagcaccgttctctgggtacaatttagtggtaagtaaacgcaccatctgttcgatac aaccctcatgagtacctttttccttcattactctgtacagtaagcaaatataaagagggaccactgg gatagctgaagaagcttgtgtaaccaaagcttttgcaacaacagggtatgcgggacatc
VYG4	Adhe2_YCO_G1	ggagaaaaaaccccggatccgtaatacgactcactatagggcccgggatggcaagctggag ccacccgcagttcgaaaagggtgcaggtatgaaagtcacaaatcagaaggagttgaaacag aagttaaatgaactacgagaggcacaaaagaaattcgcaacttacacacaagaacaggttga taaaatttttaaacagtgtgctatagccgcggccaaggaacgcattaacttagcaaaattagcag ttgaggagacgggtataggtttagtcgaagacaagataattaagaatcacttcgcggccgaata catttacaataagtataagaatgaaaaaacttgcggcatcattgatcatgatgattctctcggaatc actaaggttgcggaaccaatcggaatagttgctgcaattgtcccaaccactaatcctacgtccact gcaatatttaagtctctaatatcacttaaaaccagaaacgcgattttcttcagt
VYG5	Adhe2_YCO_G2	tctaatatcacttaaaaccagaaacgcgattttcttcagtccacacccacgtgcaaaaaaatctac cattgcagccgctaagttgatcttggatgcggctgtcaaagctggtgcacctaagaacatcatag ggtggattgatgagccttccatcgagttgagccaagacctcatgtccgaagccgatatcatcttgg ccacgggtgggccatcaatggtgaaagcagcatactcttcaggtaagcctgctatagggtaggg tgcaggtaatactccagctattatagatgaaagtgcagatatagacatggctgtctcctctattattc tgagtaaaacttatgacaacggtgttatatgtgcatcagaacaatccattttagttatgaacagtattt acgaaaaagtgaaagaagaatttgttaagcgaggatcttacatcttgaatcagaatgaaatagc caaaatcaaagagacaatgttcaaaaacggcgctataa
VYG6	Adhe2_YCO_G3	gccaaaatcaaagagacaatgttcaaaaacggcgctataaacgccgatatagttggtaagtca gcgtatatcattgccaaaatggctggcattgaagttccacaaactacaaaaattttgatagggga agtccagtctgtggagaaatctgaactattctcgcatgaaaagttgtcacccgtattggcgatgtac aaggttaaggattttgatgaagctctgaagaaagcacagagacttatagaattgggaggctcag gacatacaagctcactatacatcgattcccaaaacaataaggacaaggtaaaagaatttggtct agctatgaaaactagtcgaacatttattaatatgccaagctctcagggtgccagtggtgatctttac aattttgcgatcgctccatcctttactctaggatgcggtacttgggggggtaactcggtgtcacaaa atgttgaacccaagcacctttaaatatcaagtctgttgcaga

VYG7	Adhe2_YCO_G4	ttgaacccaagcacctttaaatatcaagtctgttgcagaaaggcgtgagaacatgctgtggttta aggttcctcagaaaatttactttaaatatggttgtttgcgttttgccctaaaggagctgaaagatatga acaagaagagggccttcatagtgactgacaaagacttgtttaaactaggttacgtcaacaagatt acaaaagtcttggatgaaatagacataaaatactcaatcttcaccgacattaagtcagatcccac catagatagtgttaagaagggtgcaaaggaaatgctcaacttcgagccggatacaattatcagc attggtggtggctccccaatggatgccgctaaagtgatgcacttattatatgaatatccagaagcg gaaattgaaaatctagccattaactttatggatattaggaaaagaatctgtaatttccccaagttag ggaccaaagctatttctgtcgcaattccgactactgct
VYG8	Adhe2_YCO_G5	agggaccaaagctatttctgtcgcaattccgactactgctggtactggttccgaagcaacaccatt tgcagttattacaaatgatgaaactggtatgaaatatccactaacttcatacgaattgactccaaa catggcaataattgacacagaattaatgttaaacatgccccggaaattaaccgctgctacaggc atagacgccctcgttcatgccattgaagcttacgtttcagtcatggcaactgactatacagacgag ttggctttacgcgcaattaaaatgatcttcaagtacctacccagagcttacaaaaacggaacaaa tgacatcgaagctcgggagaagatggcccatgcgtccaatatagcaggaatggcgtttgctaac gctttcttgggtgtttgtcactccatggctcataagttgggggctatgcaccacgttccacacggtatt gcttgtgctgtcttaattgaagaagtgattaaatataatg
VYG9	Adhe2_YCO_G6	gcttgtgctgtcttaattgaagaagtgattaaatataatgctactgattgccctactaagcaaacag catttccacaatacaaatccccaaacgctaagagaaaatacgccgagatcgccgagtatctga atcttaaaggcacgtcggatactgagaaagttactgcccttattgaagccatcagcaaactgaag atcgacctttcaattcctcaaaacatcagtgcagctggaattaacaagaaagatttttacaacacc ttagataagatgtcggagttggcattcgatgaccaatgtacaaccgcgaaccctagatatccact gatctcggaattaaaggacatctacatcaaatcattctaagtcgacatggaacagaagttgatttc cgaagaagacctcgagtaagcttggtaccgcggctagctaagatccgctctaaccgaa
VYG10	Erg10 g1	GAAAGCATAGCAATCTAATCTAAGTTTTAATTACAAAATGGTTAAC ACAGAGGTTTACATTGTGTCTGCTGTTCGCACACCCATGGGCAG CTTTGGCGGCAGTTTCGCTTCTCTTCCTGCCACCAAATTAGGCTC TATTGCCATCAAGGGCGCTCTTGAAAGGGTCAATATTAAGCCTTC TGATGTTGATGAGGTGTTTATGGGTAATGTCGTCTCAGCCAATCT TGGCCAAAACCCTGCTCGTCAATGCGCCTTAGGTGCTGGACTCC CTCGTTCCATCGTTTGTACTACTGTTAACAAGGTTTGTGCCTCTG GTATGAAGGCAACAATTCTCGGTGCTCAAACTATCATGACTGGAA ATGCTGAGATTGTCGTTGCTGGTGGTACCGAGAGCATGTCAAAT GCTCCTTACTATGCCCCAAAGAACCGTTTTGGTGCCAAATACGGT AACGTTGAGTTGGTTGACGGTCTCTTACGTGATGGTTTATCTGAT GCATATG
VYG11	Erg10 g2	GTTGACGGTCTCTTACGTGATGGTTTATCTGATGCATATGATGGA CTTCCCATGGGTAATGCTGCCGAGCTTTGCGCTGAGGAACACAG CATTGACCGTGCCAGTCAAGACGCCTTTGCCATTTCTTCTTACAA ACGTGCTCAAAACGCTCAAGCTACAAAGGCATTTGAACAAGAAAT CGTCCCCGTCGAGGTCCCTGTTGGTCGTGGTAAGCCCAACAAGC TTGTTACTGAGGACGAGGAGCCCAAGAACCTTAACGAAGACAAA CTCAAATCCGTACGTGCTGTCTTCAAGAGTAACGGTACCGTCACT GCTGCTAATGCTTCTACCTTGAACGATGGTGCTAGCGCTCTTGTT CTCATGTCGGCTGCCAAAGTGAAAGAACTTGGTCTTAAGCCTCTT GCTAAAATTATTGGTTGGGGTGAAGCCGCTCAAGATCCTGAGCG TTTTACCACTTCTCCTTCCCTTGCTATTCCCAAA

VYG12	Erg10 g3	AAGATCCTGAGCGTTTTACCACTTCTCCTTCCCTTGCTATTCCCAAAGCTCTCAAGCATGCAGGTATTGAAGCTTCCCAAGTTGACTATTACGAGATCAATGAAGCCTTTAGTGTTGTTGCTGTTGCCAACACCAAAATTCTTGGATTAGATCCCGAGCGTGTTAACATTAACGGAGGTGGTGTTGCCATGGGTCATCCCTTGGGTAGCAGTGGTTCTCGTATCATTTGTACTTTGGCATATATTCTTGCTCAGAAGGATGCCAAGATTGGTGTAGCTGCTGTTTGCAATGGTGGTGGTGGTGCTTCATCCATTGTTATCGAAAGAGTTTAGGAGCTCTTAATTAACAATTCTTCGCCAGAGGGTTTGGTCAAGTCTCCAA
VYG13	5'UTR_PGI1TdTer	cgtaatgcggcttgccagaccgtagccgttgctgcagcccagcaccaggacgtttttcggtgctttggcacctgctttaacttctgcggtgatgcgtttcttagtgtattcaatctggtcttctacacccttttacaaccctgcgggtgagcgttcagacagatattattgcgcaccattggcttgacgatcatttttaggctggtatcttgattctaaatcgattttgcaagactagcggccgccctttagtgagggttgaattcgaatttttcaaaaa
VYG14	5'UTR_TDH3TdTer	cgtaatgcggcttgccagaccgtagccgttgctgcagcccagcaccaggacgtttttcggtgctttggcacctgctttaacttctgcggtgatgcgtttcttagtgtattcaatctggtcttctacacccttttacaaccctgcgggtgagcgttcagacagatattattgcgcaccattggcttgacgatcattttgtttgtttatgtgtgtttattcgaaactaagttcttgggcggccgccctttagtgagggttgaattcgaattttcaaaaaatt
VYG15	5'UTR_PFK1TdTer	cgtaatgcggcttgccagaccgtagccgttgctgcagcccagcaccaggacgtttttcggtgctttggcacctgctttaacttctgcggtgatgcgtttcttagtgtattcaatctggtcttctacacccttttacaaccctgcgggtgagcgttcagacagatattattgcgcaccattggcttgacgatcatctttgatatgattttgtttcagatttttatataaaagctttcccaaatagtgctaaagtgaacttagatttttggtacctgtttcgaaattaaaaatagaaaaatttctctccctatattgttattcttacttcaaatttgtttatcgtttattactaggcgagacttgagtagacgacaatccaaatagaattaacagattttattggtagaaagcaatagcggccgccctttagtgagggttgaattcgaattttcaaaa
VYG16	5'UTR_ENO1TdTer	cgtaatgcggcttgccagaccgtagccgttgctgcagcccagcaccaggacgtttttcggtgctttggcacctgctttaacttctgcggtgatgcgtttcttagtgtattcaatctggtcttctacacccttttacaaccctgcgggtgagcgttcagacagatattattgcgcaccattggcttgacgatcattttgatttagtgtttgtgtgttgataagcagttgcttgggcggccgccctttagtgagggttgaattcgaatttt
VYG17	5'UTR_ENO2TdTer	cgtaatgcggcttgccagaccgtagccgttgctgcagcccagcaccaggacgtttttcggtgctttggcacctgctttaacttctgcggtgatgcgtttcttagtgtattcaatctggtcttctacacccttttacaaccctgcgggtgagcgttcagacagatattattgcgcaccattggcttgacgatcattattattgtatgttatagtattagttgcttggtgttatgaaagaaactaagaaaagaaaaataaaataaaaataaaagattgagacaagggaagaaaagatacaaaataagaattaattacagcggccgccctttagtgagggttgaattcgaattttcaaaaat
VYG18	5'UTR_TDH1TdTer	cgtaatgcggcttgccagaccgtagccgttgctgcagcccagcaccaggacgtttttcggtgctttggcacctgctttaacttctgcggtgatgcgtttcttagtgtattcaatctggtcttctacacccttttacaaccctgcgggtgagcgttcagacagatattattgcgcaccattggcttgacgatcattttgttttgtgtgtaaatttagtgaagtactgttttttgtgtgtgttggtgaaatatcaaaccaagttcttgatggcggccgcccttagtgagggttgaattcgaattttcaaaaatt
VYG19	5'UTR_PFK2TdTer	cgtaatgcggcttgccagaccgtagccgttgctgcagcccagcaccaggacgtttttcggtgctttggcacctgctttaacttctgcggtgatgcgtttcttagtgtattcaatctggtcttctacacccttttacaaccctgcgggtgagcgttcagacagatattattgcgcaccattggcttgacgatcattgcgtatggttagttcttggccaatgctaaactagtctctaaatctagttctattgttcaaatgagcggccgcccttagtgagggttgaattcgaattttcaaaaattc
VYG20	5'UTR_PYK2TdTer	cgtaatgcggcttgccagaccgtagccgttgctgcagcccagcaccaggacgtttttcggtgctttggcacctgctttaacttctgcggtgatgcgtttcttagtgtattcaatctggtcttctacacccttttacaaccctgcgggtgagcgttcagacagatattattgcgcaccattggcttgacgatcatcgatagtgcttttgttgtaatcttacaatggacgtagaggatgaaagtaaaatatagttgcgtttcaccgctctgcttcaaaatgttttatgttctttggcggccgccctttagtgagggttgaattcgaatttt

VYG21	g21_TdTer (S.cgly) with 5'UTR PYK2 gBlock 1	aaaattcgaattcaaccctcactaaagggcggccgccaaagaacataaaacattttgaagcagagcggtgaaacgcaactatattttactttcatcctctacgtccattgtaagattacaacaaaagcactatcgatgattgttaagccaatggttagaaacaacatttgtttgaacgctcacccacaaggttgtaagaagggtgttgaagaccaaattgaatacaccaagaagagaattaccgctgaagttaaggctggtgctaaggctccaaagaacgttttggttttgggttgttctaacggttacggtttggcttctagaattaccgctgctttcggttacggtgctgctaccattggtgtttctttcgaaaaggctggttctgaaaccaagtacggtaccccaggttggtacaacaacttggctttcgacgaagctgctaagagagaaggtttgtactctgttaccattgacggtgacgctttctctgacgaaattaaggctcaagttattgaagaagctaagaagaagggtattaagttcgacttgattgtttactctttggcttctccagttagaaccgacccagacaccggtattatgcacaagtctgttttgaagccattcggtaagaccttcaccggtaagaccgttgacccattcaccggtgaattgaaggaaatttctgctgaaccagctaacgacgaagaagctgctgctaccgttaaggttatgggtggtgaagac
VYG22	g22_TdTer (S.cgly) with 5'UTR PYK2 gBlock 2	gctaccgttaaggttatgggtggtgaagactgggaaagatggattaagcaattgtctaaggaaggtttgttggaagaaggttgtattaccttggcttactcttacattggtccagaagctacccaagctttgtacagaaagggtaccattggtaaggctaaggaacacttggaagctaccgctcacagattgaacaaggaaaacccatctattagagctttcgtttctgttaacaagggtttggttaccagagcttctgctgttattccagttattccattgtacttggcttctttgttcaaggttatgaaggaaaagggtaaccacgaaggttgtattgaacaaattaccagattgtacgctgaaagattgtacagaaaggacggtaccattccagttgacgaagaaaacagaattagaattgacgactgggaattggaagaagacgttcaaaaggctgtttctgctttgatggaaaaggttaccggtgaaaacgctgaatctttgaccgacttggctggttacagacacgacttcttggcttctaacggtttcgacgttgaaggtattaactacgaagctgaagttgaaagattcgacagaatttaaactagtatcgatggattacaaggatgacgacgataagatct
VYG23	g23_TdTer (S.c) with 5'UTR PYK2 gBlock 1	aaaattcgaattcaaccctcactaaagggcggccgccaaagaacataaaacattttgaagcagagcggtgaaacgcaactatattttactttcatcctctacgtccattgtaagattacaacaaaagcactatcgatgatcgtaaaaccaatggttagaaacaacatttgccttaatgcacatccacagggctgtaagaagggagtcgaagatcaaattgaatacactaaaaaaaggattactgcagaggtgaaagcaggcgctaaggctcccaagaacgtcctggtactaggatgttctaacggttatggactggcaagcaggatcacagctgcgttcggatacggagctgcaacgatcggtgtatcgtttgaaaaggctggttcagagactaaatacggaacacccggatggtataataatttggcatttgatgaggctgctaaaagagaaggtctatattctgttacaattgatggagatgccttttccgatgagataaaagcgcaagtaattgaagaagctaaaaaaaagggcattaaatttgacttaattgtctattccttggcttccccagttagaaccgatccagatacagggataatgcacaaatccgtgttaaaacctttcggtaaaacattcacaggtaagacggttgatccatttacgggtgaacttaaggagatttccgcagaacccgctaatgatgaagaggctgcagctacggtgaaggtgatgggaggtgaggat
VYG24	g24_TdTer (S.c) with 5'UTR PYK2 gBlock 2	gctacggtgaaggtgatgggaggtgaggattgggaaagatggatcaaacaattgagtaaagaaggtcttctagaggaaggctgcataactttagcttattcatacatagggccagaagccacacaagctctttacagaaagggtactatcggcaaagctaaggaacacttggaagcgacggctcacagacttaataaggagaatccaagtatcagagcttttgttagtgtcaacaaaggtctggttacacgtgcctctgccgtaatccccgtcattccactttacttggcttctctgttcaaggttatgaaggaaaaagggaatcatgaaggctgtatagaacaaatcacgagattatatgcagaaagactgtaccgaaaagatgggactatcccagttgatgaagaaaacagaatacggatcgatgattgggaactggaggaagatgtgcaaaaggcagtttccgcactaatggagaaagtgactggtgaaaacgctgaaagtttaacagatttagcaggttatagacatgactttctagcatcaaacggcttcgacgtagaaggaattaattacgaagctgaggttgagcgtttcgatagaatttaaactagtatcgatggattacaaggatgacgacgataagatct

VYG25	g25_AdhE2 (gS.c) gBlock 1	cgtaatacgactcactatagggcccgggatgaaagtcacgaaccagaaggaactgaagcag aaactgaacgaactgcgcgaagcacaaaagaaattcgctacctacacccaggaacaggtgg acaaaattttcaagcaatgcgcaatcgcggctgcaaaagaacgtatcaacctggcaaaactgg cggtggaagagactggtattggtctggttgaagataaaatcatcaaaaaccacttcgcggctga gtacatctacaacaaatacaaaaacgaaaagacttgtggtatcatcgatcacgatgactccctg ggtattaccaaagtagctgaaccgatcggcatcgttgctgcgatcgtaccgaccaccaacccga cttccactgctatcttcaaatccctgatttccctgaaaacgcgcaacgcaatctttttcagccctcac ccgcgtgctaaaaagagcactatcgcagccgccaaactgattctggacgccgcagtcaaagc aggtgcgccgaaaaatatcatcggctggatcgatgaaccttctatcgaactgtcccaggatctga tgtccgaagctgatatcattctggctaccggtggtccgagcatggttaaggcggcttacagcagc ggtaaacctgccatcggcgtgggtgccggtaacaccccggcgatcatcgatgagtctgctgaca tcgatatggcagtatcttccattattctgtccaagacttacgataacgg
VYG26	g26_AdhE2 (gS.c) gBlock 2	cattattctgtccaagacttacgataacggtgttatctgcgcaagcgaacagtccatcctggttatg aactccatctacgaaaaagtaaaggaggaatttgtcaagcgtggtagctatatcctgaaccaga acgaaatcgcgaagatcaaagagacgatgttcaagaacggcgcgatcaacgccgacatcgt gggcaaatccgcctacatcattgcgaagatggcaggtatcgaagttccgcagacgactaaaat cctgatcggtgaagtacagtctgttgaaaagtccgaactgttcagccatgagaaactgagcccg gtcctggccatgtataaagttaaagacttcgatgaagctctgaaaaaggcgcaacgtctgatcg agctgggtggttctggtcacacctctagcctgtacatcgactctcaaaataacaaggacaaggta aaagaatttggtctggctatgaaaacctcccgcaccttcatcaacatgccaagctcccagggtgc cagcggtgacctgtacaactttgcaattgcgccgtccttcaccctgggttgcggcacctggggtgg caacagcgtttcccaaaacgtggagccgaagcatctgctgaacatcaaatctgttgcagaacg ccgtgaaaacatgctgtggttcaaagtcccacagaaaatttacttcaaatacggctgcctgcgttt cgcgctgaaagaactgaaagacatgaacaaaaagcgtgcgtt
VYG27	g27_AdhE2 (gS.c) gBlock 3	actgaaagacatgaacaaaaagcgtgcgttcattgttaccgacaaagacctgttcaaactgggt tacgtgaacaaaatcaccaaagttctggatgaaattgacatcaagtactccatcttcactgatatc aaatccgacccaacgattgatagcgtgaaaaagggcgctaaagaaatgctgaactttgaacc ggacaccatcatcagcatcggtggtggctctcctatggatgctgcgaaggtcatgcacctgctgt acgaatacccggaagcggaaatcgaaaacctggctatcaacttcatggacatccgcaaacgt atctgcaacttcccgaagctgggcactaaagctatttccgttgccatcccgactaccgcgggcact ggttccgaagccacgccgttcgccgtgatcaccaacgatgaaaccggtatgaaatacccgctg acctcttacgaactgaccccgaacatggcaattatcgacaccgagctgatgctgaacatgccgc gcaagctgaccgctgctaccggcatcgacgctctggtacatgctattgaggcgtacgtttccgtga tggctaccgattacaccgacgaactggccctgcgtgcgatcaaaatgattttcaagtacctgcctc gcgcttacaaaaacggcacgaatgacatcgaggcgcgtgagaaaatggcccatgcaagcaa catcgcgggcatggccttcgccaacgcgttcctgggcgtgtgcca
VYG28	g38_AdhE2 (gS.c) gBlock 4	cttcgccaacgcgttcctgggcgtgtgccactctatggctcacaaactgggtgctatgcaccacgt gccgcacggtatcgcgtgtgctgtcctgatcgaagaagtaattaagtacaacgctactgattgcc cgactaaacagaccgccttcccacagtacaaatctcctaacgctaaacgtaagtacgctgagat cgccgaatacctgaacctgaagggtacgagcgacactgagaaagttactgcgctgatcgaag ctatctctaaactgaaaattgacctgtccatcccgcagaacatcagcgccgcaggcatcaacaa aaaggacttttacaacacgctggacaaaatgagcgaactggcttttgacgaccagtgcaccact gcaaacccgcgttacccgctgatctctgagctgaaagacatttatatcaaatctttttaagtcgaca tggaacagaagttgatttccgaagaagacctcgagtaagcttggtaccgcggctagctaagatc cgct

VYG29	g29_TDH3_ALD5-1_His5	ttagtttcgaataaacacacataaacaaacaaaatgtctgttaacgaaaagatggttcaagacat tgttcaagaagttgttgctaagatgcaaatttcttctgacgtttctggtaagaagggtgttttctctgac atgaacgaagctattgaagcttctaagaaggctcaaaagattgttgctaagatgtctatggacca aagagaagctattatttctaagattagagaaaagattaaggaaaacgctgaaattttggctagaa tgggtgttgaagaaaccggtatgggtaacgttggtcacaagattttgaagcaccaattggttgctg aaaagaccccaggtaccgaagacattaccaccaccgcttggtctggtgacagaggtttgaccttt gattgaaatgggtccattcggtgttattggtgctattaccccatgtaccaacccatctgaaaccgtttt gtgtaacaccattggtatgttggctggtggtaacaccgttgttttcaacccacacccagctgctatta agacctctatttacgctgttaacttgttgaacgaagcttctgttgaagttggtggtccagaaaacatt gctgttaccgttgaacacccaaccatggaa
VYG30	g30_TDH3ALD5-2_His5	gctgttaccgttgaacacccaaccatggaaacctctgacattatgatgaagcacaaggacattc acttgattgctgctaccggtggtccaggtgttgttaccgctgttttgtcttctggtaagagaggtattgg tgctggtgctggtaacccaccagctttggttgacgaaaccgctgacattagaaaggctgctgaag acattgttaacggttgtaccttcgacaacaacttgccatgtattgctgaaaaggaaattgttgctgtt gactctattgctgacgaattgttgcactacatggtttctgaacaaggttgttacatgatttctaaggaa gaacaagacgctttgaccgaagttgttttgaagggtggtagattgaacagaaagtgtgttggtag agacgctaagaccttgttgggtatgattggtattaccgttccagacaacattagatgtattaccttcg aaggtccaaaggaacacccattgattgctgaagaattgatgatgccaattttgggtgttgttagag ctaaggacttcgacgacgctgttgaacaagctgtttggttggaacacggtaacagacactctgct cacattcactctaagaacgttgacaacattaccaagtacgctaaggctattgacaccgctattttg gttaagaacggtccatcttacgctgctttgggtttcggtggtgaaggttactgtaccttcaccattgctt ctagaaccggtgaaggtttgacctctgcttctaccttcaccaagagaagaagatgtgttatgaccg actctttgtgtattagataaatagattaatttaaacagtatatgtaca
VYG31	g31_FBA11_ADH-1_CPS1	accataaccaagtaatacatattcaaaatgttgtggttcaaggttccacaaaagatttacttcaagt acggttgtttgagattcgctttgaaggaattgaaggacatgaacaagaagagagctttcattgtta ccgacaaggacttgttcaagttgggttacgttaacaagattaccaaggttttggacgaaattgaca ttaagtactctattttcaccgacattaagtctgacccaaccattgactctgttaagaagggtgctaag gaaatgttgaacttcgaaccagacaccattatttctattggtggtggttctccaatggacgctgctaa ggttatgcacttgttgtacgaatacccagaagctgaaattgaaaacttggctattaacttcatggac attagaaagagaatttgtaacttcccaaagttgggtaccaaggctatttctgttgctattccaaccac cgctggtaccggttctgaagctaccccattcgctgttattaccaacgacgaaaccggtatgaagt acccattgacctcttacgaattgaccccaaacatggctattattgacaccgaattgatgttgaacat gccaagaaagttgaccgctgctaccggtattgacgctttggttcacgctattgaagcttacgtttctg ttatggctaccgactacaccgacgaattggctttgagagctattaagatgattttcaagtacttgcca agagcttacaag
VYG32	g32_FBA11_ADH-2_CPS1	attttcaagtacttgccaagagcttacaagaacggtaccaacgacattgaagctagagaaaag atggctcacgcttctaacattgctggtatggctttcgctaacgctttcttgggtgtttgtcactctatggc tcacaagttgggtgctatgcaccacgttccacacggtattgcttgtgctgttttgattgaagaagttat taagtacaacgctaccgactgtccaaccaagcaaaccgctttcccacaatacaagtctccaaa cgctaagagaaagtacgctgaaattgctgaatacttgaacttgaagggtacctctgacaccgaa aaggttaccgctttgattgaagctatttctaagttgaagattgacttgtctattccacaaaacatttctg ctgctggtattaacaagaaggacttctacaacaccttggacaagatgtctgaattggctttcgacg accaatgtaccaccgctaacccaagatacccattgatttctgaattgaaggacatttacattaagt ctttctaagcgcaatgattgaatagtcaaagatttttt

Appendix 4.5. UTR sequences

Systematic Names	Gene Name	5'UTR	3'UTR
YAL038W	CDC19	CCAATCAAAACAAATAAAACATCATCACA	AAAAAGAATCATGATTGAATGAAGATATTATTTTTTTGAATTATATTTTTTAAATTTTATATAAAGACATGGTTTTTCTTTTCAACTCAAATAAAGATTTATAAGTTACTTAAATAACATACATTTTATAAGGTATTCTATAAAAAGAGTATTATGTTATTGTTAA
YBR196C	PGI1	TAGTCTTGCAAAATCGATTTAGAATCAAGATACCAGCCTAAAA	ACAAATCGCTCTTAAATATATACCTAAAGAACATTAAAGCTATATTATAAGCAAAGATACGTAAATTTTGCTTATATTATTATACACATATCATATTTCTATATTTTTAAGATTTGGTTATATAATGTACGTAATGCA
YCR012W	PGK1	TCAAGGAAGTAATTATCTACTTTTTACAACAAATATAAAACA	ATTGAATTGAATTGAAATCGATAGATCAATTTTTTTCTTTTCTCTTTCCCCATCCTTTACGCTAAAATAATAGTTTATTTTATTTTTTGAATATTTTTTATTTATATACGTATATATAGACTATTATTTATCTTTTAATGATTATTAAGATTTTTATTAAAAAAAAA
YDR050C	TPI1	TAACTACAAAAAACACATACATAAACTAAAA	GATTAATATAATTATATAAAAATATTATCTTCTTTTCTTTATATCTAGTGTTATGT
YGR192C	TDH3	CCAAGAACTTAGTTTCGAATAAACACACATAAACAAACAAA	GTGAATTTACTTTAAATCTTGCATTTAAATAAATTTTCTTTTTATAGCTTTATGACTTAGTTTCAATTTATATACTATTTTAATGACAT
YGR240C	PFK1	TATTGCTTTCTACCAATAAAATCTGTTAATTCTATTTGGATTGTCGTCTACTCAAGTCTCGCCTAGTAAATAAACGATAAACAAATTTGAAGTAAGAATAACAATATAGGGAGAGAAATTTTTCTATTTTTAATTTCGAAACAGGTACCAAAAAATCTAAGTTCACTTTAGCACTATTTGGGAAAGCTTTTATATAAAAAATCTGAAACAAAATCATATCAAAG	ATGATTGCAATGAAAAGTTTAAGTTAAGCAAAAGGAGGTAAAAATGGCATGCACTTTAATTTTTATACAATCGTTTTTTTGTCATAAGACTTATTTATGTATCTGTTGTTTTTCTTTTTCTATCCTCTATTTTTGTCTATTTGTCTTTGTTTTACTCTTTTTCATTATTATTTCTTTATATAATTTTTGTACGATATGATACACA
YGR254W	ENO1	CCAAGCAACTGCTTATCAACACACAAACACTAAATCAAA	AGCTTTTGATTAAGCCTTCTAGTCCAAAAAACACGTTTTTTTGTCATTTATTTCATTTTCTTAGAATAGTTTAGTTTATTCATTTTATAGTCACGAATGTTTTATGATTCTATATAGGGTTGCAAACAAGCATTTTTCATTTTATGTTAAAA
YHR174W	ENO2	TGTAATTAATTCTTATTTTGTATCTTTTCTTCCCTTGTCTCAATCTTTTATTTTTATTTTATTTTTCTTTTCTTAGTTTCTTTCATAACACCAAGCAACTAATACTATAACATACAATAATA	AGTGCTTTTAACTAAGAATTATTAGTCTTTTCTGCTTATTTTTTCATCATAGTTTAGAACACTTTATATTAACGAATAGTTTATGAATCTATTTAGGTTTAAAAATTGATACAGTTTTA
YJL052W	TDH1	CATCAAGAACTTGGTTTGATATTTCACCAACACACACAAAAAACAGTACTTCACTAAATTTACACACAAAACAAA	ATAAAGCAATCTTGATGAGGATAATGATTTTTTTTTGAATATACATAAATACTACCGTTTTTCTGCTAGATTTTGTGAAGACGTAAATAAGTACATATTACTTTTTAAGCCAAGACAAGATTAAGCATTAA
YJR009C	TDH2	AATTAAATTCATCACACAAACAAACAAAACAAA	ATTTAACTCCTTAAGTTACTTTAATGATTTAGTTTTTATTATTAATAATTCATGCTCATGACATCTCATATACACGTTTATAAAACTTAAATAGATT
YKL060C	FBA1	ACATATTCAAA	GTTAATTCAAATTAATTGATATAGTTTT
YKL152C	GPM1	ATATTACAATA	GTCTGAAGAATGAATGATTTGATGATTTCTTTTTCCCTCCATTTTTCTTACTGAATATATCAATGATATAGACTTGTATAGTTTATTATTTCAAATTAAGTAGCTATATATAGTCAA
YMR205C	PFK2	TCATTTGAACAATAGAACTAGATTTAGAGACTAGTTTAGCATTGGCCAAGAACTAACCATACGCA	AAGAAAATGACCTTTTATTACACTTTCTATTATTAATGTCAATTAATGTTAACCCATGTTTTTCTTTTGTGTCTATAATTCTTTTTTTTTATCTCTAAGCTTTTGAACAATGAATTTTTTGTTCCTTTCTTTTAATAATACAAGTACTACCCCATGAAACCAATATTATCATGCATTTTTATGAATGTCAAGAATAAAGATACTGTTATTTTTTGTGTCTTATTTTTTTTCTCTTTGTTTATTTAAACGTTTTCTAAAATTAAAACTTATGTATACTGGAATATGTGATATA

YOR347C	PYK2	CAAAGAACATAAAACATTTTGAAGCAG AGCGGTGAAACGCAACTATATTTTACT TTCATCCTCTACGTCCATTGTAAGATTA CAACAAAAGCACTATCG	TAAAAATTAAAGTCCTTATTTTTTTTACTTAA
YLR075W	RPL10	GTACAGTATATCAAATAACTAATTCAAG	GTTCTTTTCAAACATTTGAACTAACTTAAAAGAGAAA TTTTTTTGATTTAAATTCTAGTTTATTAAACAGTAAAT ATATTTATACATTATTGTAATACATATAATTATGTGTT TTTTTAA
YPR102C	RPL11A	CCAAGAACATACAAACATAGCCAAAG	TTTAATTAGTTTGTGAAGAAATATAATACATTTATATA CTCATATCTATGTTTTTTTTGTAACCA
YGR085C	RPL11B	CAACATATACAAAAATACGCGTCCAAG	TTTGGTCTCGGTATAGTCAGTGACAACATCAACTAC TTAATATATAAGAACAAATAAAATATCCCA
YKL006W	RPL14A	AAGAGAGTCGTGAAAAATAAAATAAACA	AATATGTACATGATCTTTAATTCTGATATATTTCGTAT GTAATTTTATCTTTAACTGGTGATTCTTTTAATAAATA AACTACAATATTATAACTATTGAAAGCCTCTGTTA
YHL001W	RPL14B	GCGCAAATAAACCAAAA	AAAAAGAAGAATAATTCTAAAATCCATAGGTAAGTAC TGAAAGCAATTTTGCGTTCCGTCAATGCATATTATAT ATATTAATCTTAACCATTTATGTAAACAACATATCATT TCATTTTGTTCTGGCCA
YKL180W	RPL17A	CCTCAAGAACTACTAATAGATTAAA	ATAGAAGATGAAAAATAATGATAATATAATTCTCTGT TAATTTTAATCTTTTATCTATGTAATTTTATCTCCCCG TCATATCACATAACTCTAAAATAAGTTTTTTATTAA
YJL177W	RPL17B	TTGGTGAATCAAAAAATTAACGAAACGAA CAAATTTAAA	ATGCTTTTTTAATTTTAAAATCTTTTAAAGTGAATATT TGATTTATATACTACTATATACTAATTTGTTTAGGCC AATACACAGAATCAAAATCAA
YPL220W	RPL18A	TAGAACATGTTTCATTGATATTGGACGTT ACTATTTCAATTTAACAGTCAACCAGTCG TCCAAAA	GCATAATTACGTGTTTTCATAGTTTAACGCTTTCAGA ACTACTTATTTAATTTTGTAAGAAGTAATTTGAGTCA CATTTGTATTTAGTAAAAGATTAAGAGTATTTC
YGL135W	RPL1B	CATAGAACTAGTCGCAAGCCTCACGGAC CACCAAATACTTTGGAAGACTAATTACAT ATCATAAA	TCACTTCCGAGCGATTAATACATATCTCCATCTTTTT AAATACCTTTTTTAATACGTATGACTCTAAGTAGTAA AAGTATTATGCATAGTTTTA
N/A	VSV	ACGAAGACCACAAAACCAGATAAAAAATA AAAACCACAAGAGGGTCTTAA	AAAGTTTCTCCTGAGCCTTTTAATGGTAATAATGGTT TGTTTGTCTTCGT

Appendix 4.6: Codon optimized sequences

DNA sequences was optimized for *E. coli*, standard *S. cerevisiae*, or *S. cerevisiae* glycolytic genes only codon usage.

TdTer

atgatcgtcaagccaatggtgcgcaataatatctgtctgaacgctcacccgcagggttgtaaaaagggtgtagaagaccagattgaataca
ctaagaaacgcatcaccgcagaagttaaagcaggtgccaaagcaccgaaaaacgtcctggtgctgggctgcagcaacggctacggtct
ggcaagccgcattacggctgcattcggttacggcgctgctactattggtgttagcttcgaaaaggcgggttctgaaaccaaatacggcactc
caggctggtacaacaacctggcattcgacgaagcagcgaagcgtgagggtctgtactctgttaccatcgacggtgacgcgttctctgacg
agatcaaagctcaggttatcgaggaagctaaaaagaaaggtatcaaattcgacctgattgtgtactccctggcctctccggttcgtaccgac
ccggataccggcatcatgcacaaaagcgtactgaagccgtttggcaaaaccttcactggtaaaaccgttgatcctttcaccggcgagctga
aggaaatctccgccgagccagctaacgatgaggaggctgctgcgaccgttaaagtgatgggtggcgaagactgggaacgttggatcaa
acaactgtccaaggaaggtctgctggaggagggctgtattactctggcatattcttacatcggcccggaggcgactcaggcactgtatcgt
aagggcaccatcggtaaagcgaaagaacatctggaggccaccgctcaccgtctgaacaaggaaaacccgagcatccgtgctttcgtgtc
cgttaacaagggcctggttacgcgcgcttccgcagtaattccggtcattccgctgtacctggcttccctgtttaaagtcatgaaagaaaaagg
caaccacgaaggttgtatcgaacaaattactcgcctgtatgcggagcgcctgtaccgtaaggatggcactatcccggttgatgaagagaac
cgcatccgcattgacgattgggaactggaagaggatgtacagaaagcggtttccgcgctgatggaaaaagtgacgggcgaaaacgcgg
aatccctgacggatctggcaggttaccgtcacgactttctggcgtctaatggtttcgacgttgagggtattaactacgaggcagaagttgaac
gtttcgatcgtatttaa

TdTer codon optimized with *S. cerevisiae* codon

atgattgttaagccaatggttagaaacaacatttgtttgaacgctcacccacaaggttgtaagaagggtgttgaagaccaaattgaatacacc
aagaagagaattaccgctgaagttaaggctggtgctaaggctccaaagaacgttttggttttgggttgttctaacggttacggtttggcttcta
gaattaccgctgctttcggttacggtgctgctaccattggtgtttctttcgaaaaggctggttctgaaaccaagtacggtaccccaggttggta
caacaacttggctttcgacgaagctgctaagagagaaggtttgtactctgttaccattgacggtgacgctttctctgacgaaattaaggctcaa
gttattgaagaagctaagaagaagggtattaagttcgacttgattgtttactctttggcttctccagttagaaccgacccagacaccggtattat
gcacaagtctgttttgaagccattcggtaagaccttcaccggtaagaccgttgacccattcaccggtgaattgaaggaaatttctgctgaacc
agctaacgacgaagaagctgctgctaccgttaaggttatgggtggtgaagactgggaaagatggattaagcaattgtctaaggaaggtttg
ttggaagaaggttgtattaccttggcttactcttacattggtccagaagctacccaagctttgtacagaaagggtaccattggtaaggctaagg
aacacttggaagctaccgctcacagattgaacaaggaaaacccatctattagagctttcgtttctgttaacaagggtttggttaccagagcttc
tgctgttattccagttattccattgtacttggcttctttgttcaaggttatgaaggaaaagggtaaccacgaaggttgtattgaacaaattaccag
attgtacgctgaaagattgtacagaaaggacggtaccattccagttgacgaagaaaacagaattagaattgacgactgggaattggaagaa
gacgttcaaaaggctgtttctgctttgatggaaaaggttaccggtgaaaacgctgaatctttgaccgacttggctggttacagacacgacttct
tggcttctaacggtttcgacgttgaaggtattaactacgaagctgaagttgaaagattcgacagaatttaa

TdTer codon optimized with S. cerevisiae glycolytic genes codon

atgattgttaagccaatggttagaaacaacatttgtttgaacgctcacccacaaggttgtaagaagggtgttgaagaccaaattgaatacacc
aagaagagaattaccgctgaagttaaggctggtgctaaggctccaaagaacgttttggttttgggttgttctaacggttacggtttggcttcta
gaattaccgctgctttcggttacggtgctgctaccattggtgtttctttcgaaaaggctggttctgaaaccaagtacggtaccccaggttggta
caacaacttggctttcgacgaagctgctaagagagaaggtttgtactctgttaccattgacggtgacgctttctctgacgaaattaaggctcaa
gttattgaagaagctaagaagaagggtattaagttcgacttgattgtttactctttggcttctccagttagaaccgacccagacaccggtattat
gcacaagtctgttttgaagccattcggtaagaccttcaccggtaagaccgttgacccattcaccggtgaattgaaggaaatttctgctgaacc
agctaacgacgaagaagctgctgctaccgttaaggttatgggtggtgaagactgggaaagatggattaagcaattgtctaaggaaggtttg
ttggaagaaggttgtattaccttggcttactcttacattggtccagaagctacccaagctttgtacagaaagggtaccattggtaaggctaagg

aacacttggaagctaccgctcacagattgaacaaggaaaacccatctattagagctttcgtttctgttaacaagggtttggttaccagagcttc
tgctgttattccagttattccattgtacttggcttctttgttcaaggttatgaaggaaaagggtaaccacgaaggttgtattgaacaaattaccag
attgtacgctgaaagattgtacagaaaggacggtaccattccagttgacgaagaaaacagaattagaattgacgactgggaattggaagaa
gacgttcaaaaggctgtttctgctttgatggaaaaggttaccggtgaaaacgctgaatctttgaccgacttggctggttacagacacgacttct
tggcttctaacggtttcgacgttgaaggtattaactacgaagctgaagttgaaagattcgacagaatttaa

EgTer (natives sequence)

atgtcgtgccccgcctcgccgtctgctgccgtggtgtctgccggcgccctctgcctgtgcgtggcaacggtattgttggcgactggatcca
accccaccgccctgtccactgcttccactcgctctccgacctcactggtccgtggggtggacaggggcttgatgaggccaaccactgcag
cggctctgacgacaatgagagaggtgccccagatggctgagggattttcaggcgaagccacgtctgcatgggccgccgcggggccgc
agtgggcggcgccgctcgtggccgcggcctcctccgcactggcgctgtggtggtgggccgcccggcgcagcgtgcggcggccgctg
gcagcgctggcggagctgcccaccgcggtcacccacctggcccccccgatggcgatgttcaccaccacagcgaaggtcatccagccc
aagattcgtggcttcatctgcacgaccacccacccgatcggctgtgagaagcgggtccaggaggagatcgcgtacgcccgtgcccaccc
gcccaccagccctggcccgaagagggtgctggtcatcggctgcagtaccggctacgggctctccacccgcatcaccgctgccttcggct
accaggccgccacgctgggcgtgttcctggcgggccccccgacgaagggccgccccgccgcggcgggctggtacaacaccgtggc
gttcgagaaggccgccctggaggccgggctgtacgcccggagccttaatggcgacgccttcgactccacaacgaaggcgcggacggt
cgaggcgatcaagcgggacctcggcacggtggacctcgtggtgtacagcatcgccgccccgaagcggacggaccctgccaccggcg
tcctccacaaggcctgcctgaagcccatcggcgccacgtacaccaaccgcactgtgaacaccgacaaggcggaggtgaccgacgtca
gcattgagccggcctcccccgaagagatcgcggacacggtgaaggtgatgggcggggaggactgggagctctggatccaggcgctgt
cggaggccggcgtgctggcggaggggggccaagacggtggcgtactcctacatcggccccgagatgacgtggcctgtctactggtccg
gcaccatcggggaggccaagaaggacgtggagaaggctgccaagcgcatcacgcagcagtacggctgcccggcgtacccggtggtg
gccaaggccttggtcacccaggccagctccgccatcccggtggtgccgctctacatctgcctgctgtaccgcgttatgaaggagaagggc
acccacgagggctgcatcgagcagatggtgcggctgctcaccacgaagctgtaccccgagaacggggcccccatcgtcgatgaggcc
ggacgtgtgcgggtggatgactgggagatggcggaggatgtgcagcaggctgttaaggacctctggagccaggtgagcactgccaac
ctcaaggacatctccgacttcgctgggtatcaaactgagttcctgcggctgttcgggttcggcattgacggcgtggactacgaccagcccg
tggacgtggaggcggacctccccagtgctgcccagcagtag

EgTer codon optimized with *E. coli* codon

atggctatgtttaccactaccgcgaaagttatccagccgaaaatccgtggttttatctgcactaccactcacccaattggctgcgaaaaacgc
gtccaggaagaaattgcttacgctcgtgcgcacccgccaaccagccctggccctaagcgtgtactggtcatcggttgtagcacgggttac
ggtctgtctacccgtatcactgctgcgttcggctaccaggcggcgaccctgggcgttttcctggcgggtccaccgaccaaaggtcgcccg
gcagctgcgggttggtacaacactgttgccttcgagaaagcagcgctggaggcgggcctgtatgcccgttctctgaacggcgacgctttt
gattccactacgaaagcgcgcactgttgaagctatcaaacgtgacctgggcaccgtagacctggtagtgtactctatcgctgccccgaagc
gtaccgatccggcgaccggcgttctgcacaaggcttgtctgaaaccaatcggcgcgacttacaccaaccgtaccgtcaacaccgacaaa
gcggaggtgaccgatgttagcatcgaacctgcctccccggaagagatcgcggacacggttaaagtgatgggtggtgaagactgggagc
tgtggattcaggcgctgagcgaagccggtgttctggcggagggtgcgaaaaccgtggcgtactcctacattggccctgagatgacctgg
ccggtatattggtctggtactattggcgaagccaaaaaggatgttgaaaaggcggctaaacgtatcacccagcagtatggttgcccagcat
acccggtagtcgctaaagcgctggtcacccaggccagctccgcaattccggtagttccactgtacatttgcctgctgtaccgtgtgatgaaa
gaaaaaggtactcatgaaggttgcattgaacagatggttcgtctgctgaccactaaactgtaccctgagaacggtgctccgatcgtggacg
aagcgggccgtgttcgtgttgatgactgggaaatggctgaagacgtgcagcaagctgttaaagacctgtggtcccaggtgtctacggcta
acctgaaagacatcagcgacttcgctggctaccaaactgagttcctgcgtctgtttggttttggtatcgacggtgtagactacgaccagccg
gttgacgttgaagcggacctgccgagcgcagcgcagcaataa

EgTer codon optimized with *S. cerevisiae* codon

atggcaatgttcacaactacagctaaggtgattcaaccaaagattagaggttttatatgtacaactactcacccaataggttgtgaaaaaaga
gtacaagaagagattgcttatgctagagcccatcctcctacatcgccaggtcctaaaagggttttagttattggttgctctacagggtacggat
tgtctacaagaatcaccgccgctttcgggtatcaagcggctacactgggggtcttccttgctggtccaccaactaaaggtagaccagcagc
agctggatggtataacaccgttgcatttgagaaggcagcgctagaggctggactatacgcccgaagcttaaatggtgatgcttttgacagta
caaccaaggcaaggacggtcgaggccattaagagagatttaggtacagttgatttggtggtatacagcattgctgcaccaaagagaacgg
atccagccacaggcgttctccataaggcttgccttaaaccaataggggctacatatacaaatcgtacagttaacacagacaaagccgaagt
gactgatgtgagtatagaaccggcttctcctgaagaaattgctgatacggtcaaggtgatgggtggcgaagattgggagctatggatccaa
gccctgtctgaggctggcgtactcgcagaaggggctaaaacagttgcatattcatatattggcccagaaatgacttggccagtgtactggtc
tggaactatcggcgaagctaagaaggatgtagaaaaagctgcaaaacgcatcacacaacaatatggatgtcccgcataccctgttgttgca
aaagctttggttacacaagcttcttcagctatcccagtggtccctctttatatttgcttactgtacagagtaatgaaggaaaaaggtactcatga
gggttgtatcgaacagatggtgcgtttacttaccactaaattgtacccagagaacggtgctcctattgttgatgaagcgggacgtgttagagtt
gatgattgggaaatggctgaagatgtccagcaggctgttaaagatttgtggtcacaagtgtccactgccaacttaaaagatataagtgatttc
gcaggttatcagacagaattccttaggctgtttgggttcggcattgacggagtcgactacgatcaacctgttgacgttgaagctgatcttcctt
cagcagcacaacagtaa

AdhE2

atgaaagtcacgaaccagaaggaactgaagcagaaactgaacgaactgcgcgaagcacaaaagaaattcgctacctacacccaggaac
aggtggacaaaattttcaagcaatgcgcaatcgcggctgcaaaagaacgtatcaacctggcaaaactggcggtggaagagactggtattg
gtctggttgaagataaaatcatcaaaaaccacttcgcggctgagtacatctacaacaaatacaaaaacgaaaagacttgtggtatcatcgat
cacgatgactccctgggtattaccaaagtagctgaaccgatcggcatcgttgctgcgatcgtaccgaccaccaacccgacttccactgctat
cttcaaatccctgatttccctgaaaacgcgcaacgcaatctttttcagccctcacccgcgtgctaaaaagagcactatcgcagccgccaaac
tgattctggacgccgcagtcaaagcaggtgcgccgaaaaatatcatcggctggatcgatgaaccttctatcgaactgtcccaggatctgat
gtccgaagctgatatcattctggctaccggtggtccgagcatggttaaggcggcttacagcagcggtaaacctgccatcggcgtgggtgc
cggtaacaccccggcgatcatcgatgagtctgctgacatcgatatggcagtatcttccattattctgtccaagacttacgataacggtgttatct
gcgcaagcgaacagtccatcctggttatgaactccatctacgaaaaagtaaaggaggaatttgtcaagcgtggtagctatatcctgaacca
gaacgaaatcgcgaagatcaaagagacgatgttcaagaacggcgcgatcaacgccgacatcgtgggcaaatccgcctacatcattgcga
agatggcaggtatcgaagttccgcagacgactaaaatcctgatcggtgaagtacagtctgttgaaaagtccgaactgttcagccatgagaa
actgagcccggtcctggccatgtataaagttaaagacttcgatgaagctctgaaaaaggcgcaacgtctgatcgagctgggtggttctggt
cacacctctagcctgtacatcgactctcaaaataacaaggacaaggtaaaagaatttggtctggctatgaaaacctcccgcaccttcatcaa
catgccaagctcccagggtgccagcggtgacctgtacaactttgcaattgcgccgtccttcaccctgggttgcggcacctggggtggcaa
cagcgtttcccaaaacgtggagccgaagcatctgctgaacatcaaatctgttgcagaacgccgtgaaaacatgctgtggttcaaagtccca
cagaaaatttacttcaaatacggctgcctgcgtttcgcgctgaaagaactgaaagacatgaacaaaaagcgtgcgttcattgttaccgacaa
agacctgttcaaactgggttacgtgaacaaaatcaccaaagttctggatgaaattgacatcaagtactccatcttcactgatatcaaatccgac
ccaacgattgatagcgtgaaaaagggcgctaaagaaatgctgaactttgaaccggacaccatcatcagcatcggtggtggctctcctatgg
atgctgcgaaggtcatgcacctgctgtacgaatacccggaagcggaaatcgaaaacctggctatcaacttcatggacatccgcaaacgtat
ctgcaacttcccgaagctgggcactaaagctatttccgttgccatcccgactaccgcgggcactggttccgaagccacgccgttcgccgtg
atcaccaacgatgaaaccggtatgaaatacccgctgacctcttacgaactgaccccgaacatggcaattatcgacaccgagctgatgctga
acatgccgcgcaagctgaccgctgctaccggcatcgacgctctggtacatgctattgaggcgtacgtttccgtgatggctaccgattacac
cgacgaactggccctgcgtgcgatcaaaatgattttcaagtacctgcctcgcgcttacaaaaacggcacgaatgacatcgaggcgcgtga
gaaaatggcccatgcaagcaacatcgcgggcatggccttcgccaacgcgttcctgggcgtgtgccactctatggctcacaaactgggtgc
tatgcaccacgtgccgcacggtatcgcgtgtgctgtcctgatcgaagaagtaattaagtacaacgctactgattgcccgactaaacagacc
gccttcccacagtacaaatctcctaacgctaaacgtaagtacgctgagatcgccgaatacctgaacctgaagggtacgagcgacactgag
aaagttactgcgctgatcgaagctatctctaaactgaaaattgacctgtccatcccgcagaacatcagcgccgcaggcatcaacaaaaagg
acttttacaacacgctggacaaaatgagcgaactggcttttgacgaccagtgcaccactgcaaacccgcgttacccgctgatctctgagctg
aaagacatttatatcaaatctttttaa

AdhE2 codon optimized with *S. cerevisiae* codon

atgaaagtcacaaatcagaaggagttgaaacagaagttaaatgaactacgagaggcacaaaagaaattcgcaacttacacacaagaaca
ggttgataaaatttttaaacagtgtgctatagccgcggccaaggaacgcattaacttagcaaaattagcagttgaggagacgggtataggttt
agtcgaagacaagataattaagaatcacttcgcggccgaatacatttacaataagtataagaatgaaaaaacttgcggcatcattgatcatga
tgattctctcggaatcactaaggttgcggaaccaatcggaatagttgctgcaattgtcccaaccactaatcctacgtccactgcaatatttaagt
ctctaatatcacttaaaaccagaaacgcgattttcttcagtccacacccacgtgcaaaaaaatctaccattgcagccgctaagttgatcttgga
tgcggctgtcaaagctggtgcacctaagaacatcatagggtggattgatgagccttccatcgagttgagccaagacctcatgtccgaagcc
gatatcatcttggccacgggtgggccatcaatggtgaaagcagcatactcttcaggtaagcctgctatagggggtaggtgcaggtaatactc
cagctattatagatgaaagtgcagatatagacatggctgtctcctctattattctgagtaaaacttatgacaacggtgttatatgtgcatcagaa
caatccattttagttatgaacagtatttacgaaaaagtgaaagaagaatttgttaagcgaggatcttacatcttgaatcagaatgaaatagccaa
aatcaaagagacaatgttcaaaaacggcgctataaacgccgatatagttggtaagtcagcgtatatcattgccaaaatggctggcattgaag
ttccacaaactacaaaaattttgataggggaagtccagtctgtggagaaatctgaactattctcgcatgaaaagttgtcacccgtattggcgat
gtacaaggttaaggattttgatgaagctctgaagaaagcacagagacttatagaattgggaggctcaggacatacaagctcactatacatcg
attcccaaaacaataaggacaaggtaaaagaatttggtctagctatgaaaactagtcgaacatttattaatatgccaagctctcagggtgcca
gtggtgatctttacaattttgcgatcgctccatcctttactctaggatgcggtacttggggtggtaactcggtgtcacaaaatgttgaacccaag
caccttttaaatatcaagtctgttgcagaaaggcgtgagaacatgctgtggtttaaggttcctcagaaaatttactttaaatatggttgtttgcgtt
ttgccctaaaggagctgaaagatatgaacaagaagagggccttcatagtgactgacaaagacttgtttaaactaggttacgtcaacaagatt
acaaaagtcttggatgaaatagacataaaatactcaatcttcaccgacattaagtcagatcccaccatagatagtgttaagaagggtgcaaa
ggaaatgctcaacttcgagccggatacaattatcagcattggtggtggctccccaatggatgccgctaaagtgatgcacttattatatgaatat
ccagaagcggaaattgaaaatctagccattaactttatggatattaggaaaagaatctgtaatttccccaagttagggaccaaagctatttctg
tcgcaattccgactactgctggtactggttccgaagcaacaccatttgcagttattacaaatgatgaaactggtatgaaatatccactaacttca
tacgaattgactccaaacatggcaataattgacacagaattaatgttaaacatgccccggaaattaaccgctgctacaggcatagacgccct
cgttcatgccattgaagcttacgtttcagtcatggcaactgactatacagacgagttggctttacgcgcaattaaaatgatcttcaagtacctac
ccagagcttacaaaaacggaacaaatgacatcgaagctcgggagaagatggcccatgcgtccaatatagcaggaatggcgtttgctaac
gctttcttgggtgtttgtcactccatggctcataagttgggggctatgcaccacgttccacacggtattgcttgtgctgtcttaattgaagaagtg
attaaatataatgctactgattgccctactaagcaaacagcatttccacaatacaaatccccaaacgctaagagaaaatacgccgagatcgc
cgagtatctgaatcttaaaggcacgtcggatactgagaaagttactgcccttattgaagccatcagcaaactgaagatcgacctttcaattcct
caaaacatcagtgcagctggaattaacaagaaagatttttacaacaccttagataagatgtcggagttggcattcgatgaccaatgtacaacc
gcgaaccctagatatccactgatctcggaattaaaggacatctacatcaaatcattctaa

AdhE2 codon optimized with *S. cerevisiae* glycolytic genes codon

atgaaagtcacgaaccagaaggaactgaagcagaaactgaacgaactgcgcgaagcacaaaagaaattcgctacctacacccaggaac
aggtggacaaaattttcaagcaatgcgcaatcgcggctgcaaaagaacgtatcaacctggcaaaactggcggtggaagagactggtattg
gtctggttgaagataaaatcatcaaaaaccacttcgcggctgagtacatctacaacaaatacaaaaacgaaaagacttgtggtatcatcgat
cacgatgactccctgggtattaccaaagtagctgaaccgatcggcatcgttgctgcgatcgtaccgaccaccaacccgacttccactgctat
cttcaaatccctgatttccctgaaaacgcgcaacgcaatctttttcagccctcacccgcgtgctaaaaagagcactatcgcagccgccaaac
tgattctggacgccgcagtcaaagcaggtgcgccgaaaaatatcatcggctggatcgatgaaccttctatcgaactgtcccaggatctgat
gtccgaagctgatatcattctggctaccggtggtccgagcatggttaaggcggcttacagcagcggtaaacctgccatcggcgtgggtgc
cggtaacaccccggcgatcatcgatgagtctgctgacatcgatatggcagtatcttccattattctgtccaagacttacgataacggtgttatct
gcgcaagcgaacagtccatcctggttatgaactccatctacgaaaaagtaaaggaggaatttgtcaagcgtggtagctatatcctgaacca
gaacgaaatcgcgaagatcaaagagacgatgttcaagaacggcgcgatcaacgccgacatcgtgggcaaatccgcctacatcattgcga
agatggcaggtatcgaagttccgcagacgactaaaatcctgatcggtgaagtacagtctgttgaaaagtccgaactgttcagccatgagaa
actgagcccggtcctggccatgtataaagttaaagacttcgatgaagctctgaaaaaggcgcaacgtctgatcgagctgggtggttctggt
cacacctctagcctgtacatcgactctcaaaataacaaggacaaggtaaaagaatttggtctggctatgaaaacctcccgcaccttcatcaa

catgccaagctcccagggtgccagcggtgacctgtacaactttgcaattgcgccgtccttcaccctgggttgcggcacctggggtggcaa
cagcgtttcccaaaacgtggagccgaagcatctgctgaacatcaaatctgttgcagaacgccgtgaaaacatgctgtggttcaaagtccca
cagaaaatttacttcaaatacggctgcctgcgtttcgcgctgaaagaactgaaagacatgaacaaaaagcgtgcgttcattgttaccgacaa
agacctgttcaaactgggttacgtgaacaaaatcaccaaagttctggatgaaattgacatcaagtactccatcttcactgatatcaaatccgac
ccaacgattgatagcgtgaaaaagggcgctaaagaaatgctgaactttgaaccggacaccatcatcagcatcggtggtggctctcctatgg
atgctgcgaaggtcatgcacctgctgtacgaatacccggaagcggaaatcgaaaacctggctatcaacttcatggacatccgcaaacgtat
ctgcaacttcccgaagctgggcactaaagctatttccgttgccatcccgactaccgcgggcactggttccgaagccacgccgttcgccgtg
atcaccaacgatgaaaccggtatgaaatacccgctgacctcttacgaactgaccccgaacatggcaattatcgacaccgagctgatgctga
acatgccgcgcaagctgaccgctgctaccggcatcgacgctctggtacatgctattgaggcgtacgtttccgtgatggctaccgattacac
cgacgaactggccctgcgtgcgatcaaaatgattttcaagtacctgcctcgcgcttacaaaaacggcacgaatgacatcgaggcgcgtga
gaaaatggcccatgcaagcaacatcgcgggcatggccttcgccaacgcgttcctgggcgtgtgccactctatggctcacaaactgggtgc
tatgcaccacgtgccgcacggtatcgcgtgtgctgtcctgatcgaagaagtaattaagtacaacgctactgattgcccgactaaacagacc
gccttcccacagtacaaatctcctaacgctaaacgtaagtacgctgagatcgccgaatacctgaacctgaagggtacgagcgacactgag
aaagttactgcgctgatcgaagctatctctaaactgaaaattgacctgtccatcccgcagaacatcagcgccgcaggcatcaacaaaaagg
acttttacaacacgctggacaaaatgagcgaactggcttttgacgaccagtgcaccactgcaaacccgcgttacccgctgatctctgagctg
aaagacatttatatcaaatctttttaa

Ald5 codon optimized with *S. cerevisiae* glycolytic genes codon

atgtctgttaacgaaaagatggttcaagacattgttcaagaagttgttgctaagatgcaaatttcttctgacgtttctggtaagaagggtgttttct
ctgacatgaacgaagctattgaagcttctaagaaggctcaaaagattgttgctaagatgtctatggaccaaagagaagctattatttctaagat
tagagaaaagattaaggaaaacgctgaaatttttggctagaatgggtgttgaagaaaccggtatgggtaacgttggtcacaagattttgaagc
accaattggttgctgaaaagaccccaggtaccgaagacattaccaccaccgcttggtctggtgacagaggtttgaccttgattgaaatgggt
ccattcggtgttattggtgctattaccccatgtaccaacccatctgaaaccgttttgtgtaacaccattggtatgttggctggtggtaacaccgtt
gttttcaacccacacccagctgctattaagacctctatttacgctgttaacttgttgaacgaagcttctgttgaagttggtggtccagaaaacatt
gctgttaccgttgaacacccaaccatggaaacctctgacattatgatgaagcacaaggacattcacttgattgctgctaccggtggtccaggt
gttgttaccgctgttttgtcttctggtaagagaggtattggtgctggtgctggtaacccaccagctttggttgacgaaaccgctgacattagaa
aggctgctgaagacattgttaacggttgtaccttcgacaacaacttgccatgtattgctgaaaaggaaattgttgctgttgactctattgctgac
gaattgttgcactacatggtttctgaacaaggttgttacatgatttctaaggaagaacaagacgctttgaccgaagttgttttgaagggtggta
gattgaacagaaagtgtgttggtagagacgctaagaccttgttgggtatgattggtattaccgttccagacaacattagatgtattaccttcga
aggtccaaaggaacacccattgattgctgaagaattgatgatgccaattttgggtgttgttagagctaaggacttcgacgacgctgttgaaca
agctgtttggttggaacacggtaacagacactctgctcacattcactctaagaacgttgacaacattaccaagtacgctaaggctattgacac
cgctattttggttaagaacggtccatcttacgctgctttgggtttcggtggtgaaggttactgtaccttcaccattgcttctagaaccggtgaag
gtttgacctctgcttctaccttcaccaagagaagaagatgtgttatgaccgactctttgtgtattagataa

Adh (Adh domain from AdhE2) codon optimized with *S. cerevisiae* glycolytic genes codon

atgttgtggttcaaggttccacaaaagatttacttcaagtacggttgtttgagattcgctttgaaggaattgaaggacatgaacaagaagaga
gctttcattgttaccgacaaggacttgttcaagttgggttacgttaacaagattaccaaggttttggacgaaattgacattaagtactctattttca
ccgacattaagtctgacccaaccattgactctgttaagaagggtgctaaggaaatgttgaacttcgaaccagacaccattatttctattggtgg
tggttctccaatggacgctgctaaggttatgcacttgttgtacgaatacccagaagctgaaattgaaaacttggctattaacttcatggacatta
gaaagagaatttgtaacttcccaaagttgggtaccaaggctatttctgttgctattccaaccaccgctggtaccggttctgaagctaccccatt
cgctgttattaccaacgacgaaaccggtatgaagtacccattgacctcttacgaattgaccccaaacatggctattattgacaccgaattgat
gttgaacatgccaagaaagttgaccgctgctaccggtattgacgctttggttcacgctattgaagcttacgtttctgttatggctaccgactaca
ccgacgaattggctttgagagctattaagatgattttcaagtacttgccaagagcttacaagaacggtaccaacgacattgaagctagagaa
aagatggctcacgcttctaacattgctggtatggctttcgctaacgctttcttgggtgtttgtcactctatggctcacaagttgggtgctatgcac
cacgttccacacggtattgcttgtgctgttttgattgaagaagttattaagtacaacgctaccgactgtccaaccaagcaaaccgctttcccac
aatacaagtctccaaacgctaagagaaagtacgctgaaattgctgaatacttgaacttgaagggtacctctgacaccgaaaaggttaccgct
ttgattgaagctatttctaagttgaagattgacttgtctattccacaaaacatttctgctgctggtattaacaagaaggacttctacaacaccttgg
acaagatgtctgaattggctttcgacgaccaatgtaccaccgctaacccaagatacccattgatttctgaattgaaggacatttacattaagtct
ttctaa

Appendix 4.7: RNA-Sequencing

The following data were filtered with the normalized fold change value ≤ 2 and ≥ 2 after all the statically analysis on the CLC Genomics Workbench. n=3.

A. BY4741*adh1*Δ and BY4741*adh1*Δ_#68-69-70

Feature ID	Experiment - Fold Change (normalized values)	Baggerley's test: Host_EmptyVector vs Host normalized values - Test statistic	Baggerley's test: Host_EmptyVector vs Host normalized values - P-value	Baggerley's test: Host_Empty Vector vs Host normalized values - FDR p-value correction	Annotations - Transcript ID	Annotations - Gene title
ADH1_1	696.37	27.06	0	0	YOL086C	Alcohol dehydrogenase; fermentative isozyme active as homo- or heterotetramers; required for the reduction of acetaldehyde to ethanol, the last step in the glycolytic pathway; ADH1 has a paralog, ADH5, that arose from the whole genome duplication
AGA2_1	6.72	7.01	2.39E-12	6.86E-11	YGL032C	Adhesion subunit of a-agglutinin of a-cells; C-terminal sequence acts as a ligand for alpha-agglutinin (Sag1p) during agglutination, modified with O-linked oligomannosyl chains, linked to anchorage subunit Aga1p via two disulfide bonds
AHT1_1	âˆž	6.13	8.54E-10	1.80E-08		
CMK2_1	2.05	11.69	0	0	YOL016C	Calmodulin-dependent protein kinase; may play a role in stress response, many CA++/calmodulan dependent phosphorylation substrates demonstrated in vitro, amino acid sequence similar to mammalian Cam Kinase II; CMK2 has a paralog, CMK1, that arose from the whole genome duplication

DSF1_1	4.46	4.64	3.42E-06	4.30E-05	YEL070W	Putative mannitol dehydrogenase; YNR073C has a paralog, DSF1, that arose from a segmental duplication /// Putative mannitol dehydrogenase; YNR073C has a paralog, DSF1, that arose from a segmental duplication
FIT2_1	3.58	6.64	3.24E-11	8.15E-10	YOR382W	Mannoprotein that is incorporated into the cell wall; incorporated via a glycosylphosphatidylinositol (GPI) anchor; involved in the retention of siderophore-iron in the cell wall
FIT3_1	2.31	8.29	0	0	YOR383C	Mannoprotein that is incorporated into the cell wall; incorporated via a glycosylphosphatidylinositol (GPI) anchor; involved in the retention of siderophore-iron in the cell wall
HEM13_1	2.16	6.07	1.27E-09	2.63E-08	YDR044W	Coproporphyrinogen III oxidase; an oxygen requiring enzyme that catalyzes the sixth step in the heme biosynthetic pathway; transcription is repressed by oxygen and heme (via Rox1p and Hap1p)
HIS3_1	110.27	11.10	0	0	YOR202W	Imidazoleglycerol-phosphate dehydratase; catalyzes the sixth step in histidine biosynthesis; mutations cause histidine auxotrophy and sensitivity to Cu, Co, and Ni salts; transcription is regulated by general amino acid control via Gcn4p
HSP26_1	2.49	6.37	1.91E-10	4.34E-09	YBR072W	Small heat shock protein (sHSP) with chaperone activity; forms hollow, sphere-shaped oligomers that suppress unfolded proteins aggregation; oligomer activation requires heat-induced conformational change; also has mRNA binding activity
HXK1_1	2.14	24.22	0	0	YFR053C	Hexokinase isoenzyme 1; a cytosolic protein that catalyzes phosphorylation of glucose during glucose metabolism; expression is highest during growth on non-glucose carbon sources; glucose-induced repression involves hexokinase Hxk2p; HXK1 has a paralog, HXK2, that arose from the whole genome duplication
HXT4_1	42.76	12.35	0	0	YHR092C	High-affinity glucose transporter; member of the major facilitator superfamily, expression is induced by low levels of glucose and repressed by high levels of glucose; HXT4 has a paralog, HXT7, that arose from the whole genome duplication

HXT6_1	2.93	8.34	0	0		
INA1_1	3.07	13.46	0	0	YLR413W	Putative protein of unknown function; not an essential gene; YLR413W has a paralog, FAT3, that arose from the whole genome duplication
LEU2_1	1095.56	17.88	0	0	YCL018W	Beta-isopropylmalate dehydrogenase (IMDH); catalyzes the third step in the leucine biosynthesis pathway; can additionally catalyze the conversion of β-ethylmalate into α-ketovalerate
MF(ALPHA)1_1	2.12	8.43	0	0	YPL187W	Mating pheromone alpha-factor, made by alpha cells; interacts with mating type a cells to induce cell cycle arrest and other responses leading to mating; also encoded by MF(ALPHA)2, although MF(ALPHA)1 produces most alpha-factor; MF(ALPHA)1 has a paralog, MF(ALPHA)2, that arose from the whole genome duplication
MFA1_1	10.08	8.95	0	0	YDR461W	Mating pheromone a-factor; made by a cells; interacts with alpha cells to induce cell cycle arrest and other responses leading to mating; biogenesis involves C-terminal modification, N-terminal proteolysis, and export; also encoded by MFA2
MHF1_1	26.09	13.00	0	0	YOL086W-A	Component of the heterotetrameric MHF histone-fold complex; in humans the MMF complex interacts with both DNA and Mph1p ortholog FANCM, a Fanconi anemia complementation group protein, to stabilize and remodel blocked replication forks and repair damaged DNA; mhf1 srs2 double mutants are MMS hypersensitive; ortholog of human centromere constitutive-associated network (CCAN) subunit CENP-S, also known as MHF1
NDI1_1	2.05	14.35	0	0	YML120C	NADH:ubiquinone oxidoreductase; transfers electrons from NADH to ubiquinone in the respiratory chain but does not pump protons, in contrast to the higher eukaryotic multisubunit respiratory complex I; phosphorylated; involved in Mn and H2O2 induced apoptosis; upon apoptotic stress, Ndip is activated in the mitochondria by N-terminal cleavage, and the truncated protein translocates to the cytoplasm to induce apoptosis; homolog of human AMID

PHO5_1	3.06	9.46	0	0	YBR093C	Repressible acid phosphatase; 1 of 3 repressible acid phosphatases that also mediates extracellular nucleotide-derived phosphate hydrolysis; secretory pathway derived cell surface glycoprotein; induced by phosphate starvation and coordinately regulated by PHO4 and PHO2
PHO84_1	12.01	31.12	0	0	YML123C	High-affinity inorganic phosphate (Pi) transporter; also low-affinity manganese transporter; regulated by Pho4p and Spt7p; mutation confers resistance to arsenate; exit from the ER during maturation requires Pho86p; cells overexpressing Pho84p accumulate heavy metals but do not develop symptoms of metal toxicity
PHO89_1	9.07	12.55	0	0	YBR296C	Plasma membrane Na+/Pi cotransporter; active in early growth phase; similar to phosphate transporters of Neurospora crassa; transcription regulated by inorganic phosphate concentrations and Pho4p; mutations in related human transporter genes hPit1 and hPit2 are associated with hyperphosphatemia-induced calcification of vascular tissue and familial idiopathic basal ganglia calcification
REG2_1	2.19	5.11	3.24E-07	4.83E-06	YBR050C	Regulatory subunit of the Glc7p type-1 protein phosphatase; involved with Reg1p, Glc7p, and Snf1p in regulation of glucose-repressible genes, also involved in glucose-induced proteolysis of maltose permease; REG2 has a paralog, REG1, that arose from the whole genome duplication
RPM2_1	2.17	5.63	1.80E-08	3.16E-07	YML091C	Protein subunit of mitochondrial RNase P; has roles in nuclear transcription, cytoplasmic and mitochondrial RNA processing, and mitochondrial translation; distributed to mitochondria, cytoplasmic processing bodies, and the nucleus
RPS14B_1	3.00	7.69	1.53E-14	5.60E-13	YJL191W	Protein component of the small (40S) ribosomal subunit; required for ribosome assembly and 20S pre-rRNA processing; mutations confer cryptopleurine resistance; homologous to mammalian ribosomal protein S14 and bacterial S11; RPS14B has a paralog, RPS14A, that arose from the whole genome duplication

RRG8_1	2.30	3.69	2.27E-04	1.84E-03	YPR116W	Putative protein of unknown function; required for mitochondrial genome maintenance; null mutation results in a decrease in plasma membrane electron transport
SAG1_1	2.10	9.22	0	0	YJR004C	Alpha-agglutinin of alpha-cells; binds to Aga1p during agglutination, N-terminal half is homologous to the immunoglobulin superfamily and contains binding site for a-agglutinin, C-terminal half is highly glycosylated and contains GPI anchor
SIT1_1	4.37	11.30	0	0	YEL065W	Ferrioxamine B transporter; member of the ARN family of transporters that specifically recognize siderophore-iron chelates; transcription is induced during iron deprivation and diauxic shift; potentially phosphorylated by Cdc28p
SPL2_1	22.40	14.70	0	0	YHR136C	Protein with similarity to cyclin-dependent kinase inhibitors; downregulates low-affinity phosphate transport during phosphate limitation by targeting Pho87p to the vacuole; upstream region harbors putative hypoxia response element (HRE) cluster; overproduction suppresses a plc1 null mutation; promoter shows an increase in Snf2p occupancy after heat shock; GFP-fusion protein localizes to the cytoplasm
TIP1_1	2.01	27.97	0	0	YBR067C	Major cell wall mannoprotein with possible lipase activity; transcription is induced by heat- and cold-shock; member of the Srp1p/Tip1p family of serine-alanine-rich proteins
TIS11_1	2.20	3.90	9.56E-05	8.56E-04	YLR136C	mRNA-binding protein expressed during iron starvation; binds to a sequence element in the 3'-untranslated regions of specific mRNAs to mediate their degradation; involved in iron homeostasis; protein increases in abundance and relative distribution to the nucleus increases upon DNA replication stress; TIS11 has a paralog, CTH1, that arose from the whole genome duplication
TRA1_1	2.29	5.15	2.54E-07	3.87E-06	YHR099W	Subunit of SAGA and NuA4 histone acetyltransferase complexes; interacts with acidic activators (e.g., Gal4p) which leads to transcription activation; similar to human TRRAP, which is a

						cofactor for c-Myc mediated oncogenic transformation
URA3_1	1340.69	4.95	7.36E-07	1.03E-05	YEL021W	Orotidine-5'-phosphate (OMP) decarboxylase; catalyzes the sixth enzymatic step in the de novo biosynthesis of pyrimidines, converting OMP into uridine monophosphate (UMP); converts 5-FOA into 5-fluorouracil, a toxic compound
VTC3_1	2.04	6.62	3.53E-11	8.83E-10	YPL019C	Subunit of vacuolar transporter chaperone (VTC) complex; involved in membrane trafficking, vacuolar polyphosphate accumulation, microautophagy and non-autophagic vacuolar fusion; VTC3 has a paralog, VTC2, that arose from the whole genome duplication
YBR056W-A_1	2.31	4.32	1.55E-05	1.68E-04	YBR056W-A	Protein of unknown function; mRNA identified as translated by ribosome profiling data; partially overlaps dubious ORF YBR056C-B; YBR056W-A has a paralog, YDR034W-B, that arose from the whole genome duplication
YBR230W-A_1	2.22	11.48	0	0	YBR230W-A	Putative protein of unknown function; YBR230W-A has a paralog, COQ8, that arose from the whole genome duplication
YBR296C-A_1	3.89	5.77	7.89E-09	1.48E-07	YBR296C-A	Putative protein of unknown function; identified by gene-trapping, microarray-based expression analysis, and genome-wide homology searching
YFR052C-A_1	2.06	7.79	6.44E-15	2.43E-13		
YML090W_1	2.22	3.66	2.48E-04	1.98E-03		
YMR052C-A_1	2.36	3.44	5.91E-04	4.28E-03		
YOR203W_1	75.79	11.41	0	0		
AAD3_1	-2.71	-4.66	3.16E-06	3.99E-05	YCR107W	Putative aryl-alcohol dehydrogenase; similar to P. chrysosporium aryl-alcohol dehydrogenase; mutational analysis has not yet revealed a physiological role; AAD15 has a paralog, AAD3, that arose from a segmental duplication; members of the AAD gene family comprise three pairs (AAD3 + AAD15, AAD6/AAD16 + AAD4, AAD10 + AAD14)

						whose two genes are more related to one another than to other members of the family /// Putative aryl-alcohol dehydrogenase; similar to P. chrysosporium aryl-alcohol dehydrogenase; mutational analysis has not yet revealed a physiological role; AAD15 has a paralog, AAD3, that arose from a segmental duplication; members of the AAD gene family comprise three pairs (AAD3 + AAD15, AAD6/AAD16 + AAD4, AAD10 + AAD14) whose two genes are more related to one another than to other members of the family
ABP1_1	-2.07	-8.92	4.82E-19	2.45E-17	YCR088W	Actin-binding protein of the cortical actin cytoskeleton; important for activation of the Arp2/3 complex that plays a key role actin in cytoskeleton organization; inhibits barbed-end actin filament elongation; phosphorylation within its Proline-Rich Regio, mediated by Cdc28p and Pho85p, protects Abp1p from proteolysis mediated by its own PEST sequences; mammalian homologue of HIP-55 (hematopoietic progenitor kinase 1 [HPK1]-interacting protein of 55 kDa)
ACO2_1	-2.01	-5.41	6.39E-08	1.05E-06	YJL200C	Putative mitochondrial aconitase isozyme; similarity to Aco1p, an aconitase required for the TCA cycle; expression induced during growth on glucose, by amino acid starvation via Gcn4p, and repressed on ethanol
ADH2_1	-2.53	-6.56	5.44E-11	1.32E-09	YMR303C	Glucose-repressible alcohol dehydrogenase II; catalyzes the conversion of ethanol to acetaldehyde; involved in the production of certain carboxylate esters; regulated by ADR1
ADH5_1	-2.19	-11.84	2.29E-32	1.63E-30	YBR145W	Alcohol dehydrogenase isoenzyme V; involved in ethanol production; ADH5 has a paralog, ADH1, that arose from the whole genome duplication
ADH6_1	-4.44	-43.88	0	0	YMR318C	NADPH-dependent medium chain alcohol dehydrogenase; has broad substrate specificity; member of the cinnamyl family of alcohol dehydrogenases; may be involved in fusel alcohol synthesis or in aldehyde tolerance; protein abundance increases in response to DNA replication stress

ADP1_1	-2.40	-6.06	1.38E-09	2.83E-08	YCR011C	Putative ATP-dependent permease of the ABC transporter family
AGP1_1	-4.68	-12.91	3.99E-38	3.10E-36	YCL025C	Low-affinity amino acid permease with broad substrate range; involved in uptake of asparagine, glutamine, and other amino acids; expression regulated by SPS plasma membrane amino acid sensor system (Ssy1p-Ptr3p-Ssy5p); AGP1 has a paralog, GNP1, that arose from the whole genome duplication
AHC2_1	-2.07	-8.37	5.78E-17	2.57E-15	YCR082W	Component of the ADA histone acetyltransferase complex; Ach2p and Ach1p are unique to the ADA complex and not shared with the related SAGA and SLIK complexes; may tether Ach1p to the complex
AIM20_1	-2.53	-8.95	3.44E-19	1.76E-17	YIL158W	Putative protein of unknown function; overexpression causes cell cycle delay or arrest; green fluorescent protein (GFP)-fusion protein localizes to vacuole; null mutant displays elevated frequency of mitochondrial genome loss; relocalizes from nucleus to cytoplasm upon DNA replication stress; AIM20 has a paralog, SKG1, that arose from the whole genome duplication
ALD5_1	-6.06	-12.19	3.33E-34	2.42E-32	YER073W	Mitochondrial aldehyde dehydrogenase; involved in regulation or biosynthesis of electron transport chain components and acetate formation; activated by K+; utilizes NADP+ as the preferred coenzyme; constitutively expressed
ALD6_1	-2.58	-6.32	2.64E-10	5.94E-09	YPL061W	Cytosolic aldehyde dehydrogenase; activated by Mg2+ and utilizes NADP+ as the preferred coenzyme; required for conversion of acetaldehyde to acetate; constitutively expressed; locates to the mitochondrial outer surface upon oxidative stress
ARE1_1	-2.32	-6.91	4.91E-12	1.38E-10	YCR048W	Acyl-CoA:sterol acyltransferase; endoplasmic reticulum enzyme that contributes the major sterol esterification activity in the absence of oxygen; ARE1 has a paralog, ARE2, that arose from the whole genome duplication
ARG1_1	-6.55	-6.24	4.34E-10	9.55E-09	YOL058W	Arginosuccinate synthetase; catalyzes the formation of L-argininosuccinate from citrulline and L-aspartate in the arginine biosynthesis pathway; potential Cdc28p substrate

ARG3_1	-6.88	-6.17	6.73E-10	1.45E-08	YJL088W	Ornithine carbamoyltransferase; also known as carbamoylphosphate:L-ornithine carbamoyltransferase; catalyzes the biosynthesis of the arginine precursor citrulline
ARG5,6_1	-3.43	-6.42	1.40E-10	3.25E-09	YER069W	Acetylglutamate kinase and N-acetyl-gamma-glutamyl-phosphate reductase; N-acetyl-L-glutamate kinase (NAGK) catalyzes the 2nd and N-acetyl-gamma-glutamyl-phosphate reductase (NAGSA), the 3rd step in arginine biosynthesis; synthesized as a precursor which is processed in the mitochondrion to yield mature NAGK and NAGSA; enzymes form a metabolon complex with Arg2p; NAGK C-terminal domain stabilizes the enzymes, slows catalysis and is involved in feed-back inhibition by arginine
ATG15_1	-3.00	-7.26	3.84E-13	1.22E-11	YCR068W	Lipase required for intravacuolar lysis of autophagic and Cvt bodies; targeted to intravacuolar vesicles during autophagy via the multivesicular body (MVB) pathway
BAP2_1	-3.08	-19.60	1.60E-85	1.83E-83	YBR068C	High-affinity leucine permease; functions as a branched-chain amino acid permease involved in uptake of leucine, isoleucine and valine; contains 12 predicted transmembrane domains; BAP2 has a paralog, BAP3, that arose from the whole genome duplication
BAT1_1	-5.13	-21.03	3.29E-98	3.91E-96	YHR208W	Mitochondrial branched-chain amino acid (BCAA) aminotransferase; preferentially involved in BCAA biosynthesis; homolog of murine ECA39; highly expressed during logarithmic phase and repressed during stationary phase; BAT1 has a paralog, BAT2, that arose from the whole genome duplication
BNA4_1	-2.79	-7.14	9.42E-13	2.83E-11	YBL098W	Kynurenine 3-mono oxygenase; required for the de novo biosynthesis of NAD from tryptophan via kynurenine; expression regulated by Hst1p; putative therapeutic target for Huntington disease
BUD23_1	-2.17	-8.48	2.32E-17	1.07E-15	YCR047C	Methyltransferase; methylates residue G1575 of 18S rRNA; required for rRNA processing and nuclear export of 40S ribosomal subunits independently of methylation activity; diploid mutant displays random budding pattern

CAR1_1	-2.30	-5.17	2.33E-07	3.57E-06	YPL111W	Arginase, catabolizes arginine to ornithine and urea; expression responds to both induction by arginine and nitrogen catabolite repression; disruption decreases production of carcinogen ethyl carbamate during wine fermentation and also enhances freeze tolerance
CAR2_1	-2.63	-6.17	6.74E-10	1.45E-08	YLR438W	L-ornithine transaminase (OTAse); catalyzes the second step of arginine degradation, expression is dually-regulated by allophanate induction and a specific arginine induction process; not nitrogen catabolite repression sensitive; protein abundance increases in response to DNA replication stress
CDA1_1	-2.13	-3.41	6.41E-04	4.60E-03	YLR307W	Chitin deacetylase; together with Cda2p involved in the biosynthesis ascospore wall component, chitosan; required for proper rigidity of the ascospore wall
CDC10_1	-2.12	-7.70	1.41E-14	5.19E-13	YCR002C	Component of the septin ring, required for cytokinesis; septins are GTP-binding proteins that assemble into rod-like hetero-oligomers that can associate to form filaments; septin rings at the mother-bud neck act as scaffolds for recruiting cell division factors and as barriers to prevent diffusion of specific proteins between mother and daughter cells; N-terminus interacts with phosphatidylinositol-4,5-bisphosphate; protein abundance increases under DNA damage stress
CDC50_1	-2.07	-4.41	1.03E-05	1.16E-04	YCR094W	Endosomal protein that interacts with phospholipid flippase Drs2p; interaction with Cdc50p is essential for Drs2p catalytic activity; mutations affect cell polarity and polarized growth; similar to Lem3p; CDC50 has a paralog, YNR048W, that arose from the whole genome duplication
CDC60_1	-2.03	-8.46	2.79E-17	1.28E-15	YPL160W	Cytosolic leucyl tRNA synthetase; ligates leucine to the appropriate tRNA
CIS3_1	-2.22	-9.70	3.01E-22	1.74E-20	YJL158C	Mannose-containing glycoprotein constituent of the cell wall; member of the PIR (proteins with internal repeats) family
COS12_1	-9.54	-3.54	4.00E-04	3.05E-03	YGL263W	Protein of unknown function; member of the DUP380 subfamily of conserved, often subtelomerically-encoded proteins

CPA1_1	-2.02	-17.89	1.34E-71	1.40E-69	YOR303W	Small subunit of carbamoyl phosphate synthetase; carbamoyl phosphate synthetase catalyzes a step in the synthesis of citrulline, an arginine precursor; translationally regulated by an attenuator peptide encoded by YOR302W within the CPA1 mRNA 5'-leader
CPA2_1	-2.32	-5.15	2.65E-07	4.00E-06	YJR109C	Large subunit of carbamoyl phosphate synthetase; carbamoyl phosphate synthetase catalyzes a step in the synthesis of citrulline, an arginine precursor
CPR4_1	-2.70	-12.81	1.46E-37	1.12E-35	YCR069W	Peptidyl-prolyl cis-trans isomerase (cyclophilin); catalyzes the cis-trans isomerization of peptide bonds N-terminal to proline residues; has a potential role in the secretory pathway; CPR4 has a paralog, CPR8, that arose from the whole genome duplication
CSM1_1	-2.22	-4.31	1.61E-05	1.73E-04	YCR086W	Nucleolar protein that mediates homolog segregation during meiosis I; forms a complex with Lrs4p and then Mam1p at kinetochores; required for condensin recruitment to the replication fork barrier site and rDNA repeat segregation
CTR86_1	-2.46	-3.99	6.59E-05	6.11E-04	YCR054C	Essential protein of unknown function; with orthologs in Ashbya gossypii and Candida albicans; similar to human ATXN10, mutations in which cause spinocerebellar ataxia type 10; codon usage corresponds to that observed for yeast genes expressed at low levels; relative distribution to the nucleus increases upon DNA replication stress
CWH43_1	-2.60	-7.05	1.78E-12	5.25E-11	YCR017C	Putative sensor/transporter protein involved in cell wall biogenesis; contains 14-16 transmembrane segments and several putative glycosylation and phosphorylation sites; null mutation is synthetically lethal with pkc1 deletion
CWP1_1	-2.22	-5.70	1.22E-08	2.23E-07	YKL096W	Cell wall mannoprotein that localizes to birth scars of daughter cells; linked to a beta-1,3- and beta-1,6-glucan heteropolymer through a phosphodiester bond; required for propionic acid resistance
CWP2_1	-2.42	-14.08	5.17E-45	4.48E-43	YKL096W-A	Covalently linked cell wall mannoprotein; major constituent of the cell wall; plays a role in stabilizing the cell wall; involved in low pH resistance; precursor is GPI-anchored

DCC1_1	-2.34	-5.01	5.46E-07	7.69E-06	YCL016C	Subunit of a complex with Ctf8p and Ctf18p; shares some components with Replication Factor C; required for sister chromatid cohesion and telomere length maintenance
ELO2_1	-3.06	-7.88	3.33E-15	1.30E-13	YCR034W	Fatty acid elongase, involved in sphingolipid biosynthesis; acts on fatty acids of up to 24 carbons in length; mutations have regulatory effects on 1,3-beta-glucan synthase, vacuolar ATPase, and the secretory pathway; FEN1 has a paralog, ELO1, that arose from the whole genome duplication
EMC1_1	-2.22	-7.02	2.23E-12	6.48E-11	YCL045C	Member of conserved endoplasmic reticulum membrane complex; involved in efficient folding of proteins in the ER; null mutant displays induction of the unfolded protein response; interacts with Gal80p; homologous to worm H17B01.4/EMC-1, fly CG2943, and human KIAA0090
ERS1_1	-2.88	-5.16	2.41E-07	3.67E-06	YCR075C	Protein with similarity to human cystinosin; cystinosin is a H(+)-driven transporter involved in L-cystine export from lysosomes and implicated in the disease cystinosis; contains seven transmembrane domains
FEN2_1	-2.10	-4.15	3.28E-05	3.27E-04	YCR028C	Plasma membrane H+-pantothenate symporter; confers sensitivity to the antifungal agent fenpropimorph; relocalizes from vacuole to cytoplasm upon DNA replication stress
FKH1_1	-2.44	-4.38	1.17E-05	1.29E-04	YIL131C	Forkhead family transcription factor; minor role in expression of G2/M phase genes; negatively regulates transcription elongation; positive role in chromatin silencing at HML, HMR; facilitates clustering and activation of early-firing replication origins; binds to recombination enhancer near HML, regulates donor preference during mating-type switching; relocalizes to cytosol in response to hypoxia; FKH1 has a paralog, FKH2, that arose from the whole genome duplication
FUB1_1	-2.14	-6.77	1.32E-11	3.55E-10	YCR076C	Putative protein of unknown function; interacts physically with multiple subunits of the 20S proteasome and genetically with genes encoding 20S core particle and 19S regulatory particle subunits; exhibits boundary activity which blocks the propagation of heterochromatic silencing; contains a

						PI31 proteasome regulator domain and sequence similarity with human PSMF1, a proteasome inhibitor; not an essential gene
FUI1_1	-2.93	-5.49	3.98E-08	6.76E-07	YBL042C	High affinity uridine permease, localizes to the plasma membrane; also mediates low but significant transport of the cytotoxic nucleoside analog 5-fluorouridine; not involved in uracil transport; relative distribution to the vacuole increases upon DNA replication stress
FUR4_1	-2.51	-4.52	6.10E-06	7.32E-05	YBR021W	Plasma membrane localized uracil permease; expression is tightly regulated by uracil levels and environmental cues; conformational alterations induced by unfolding or substrate binding result in Rsp5p-mediated ubiquitination and degradation
GAS2_1	-2.23	-6.08	1.18E-09	2.45E-08	YLR343W	1,3-beta-glucanosyltransferase; involved with Gas4p in spore wall assembly; has similarity to Gas1p
GAS3_1	-2.39	-5.79	6.94E-09	1.32E-07	YMR215W	Putative 1,3-beta-glucanosyltransferase; has similarity go other GAS family members; low abundance, possibly inactive member of the GAS family of GPI-containing proteins; localizes to the cell wall; mRNA induced during sporulation
GAT1_1	-2.11	-5.03	5.00E-07	7.10E-06	YFL021W	Transcriptional activator of nitrogen catabolite repression genes; contains a GATA-1-type zinc finger DNA-binding motif; activity and localization regulated by nitrogen limitation and Ure2p; different translational starts produce two major and two minor isoforms that are differentially regulated and localized
GBP2_1	-2.14	-8.84	9.36E-19	4.72E-17	YCL011C	Poly(A+) RNA-binding protein; key surveillance factor for the selective export of spliced mRNAs from the nucleus to the cytoplasm; preference for intron-containing genes; similar to Npl3p; also binds single-stranded telomeric repeat sequence in vitro; relocalizes to the cytosol in response to hypoxia; GBP2 has a paralog, HRB1, that arose from the whole genome duplication
GCY1_1	-3.16	-32.79	7.88E-236	1.01E-233	YOR120W	Glycerol dehydrogenase; involved in an alternative pathway for glycerol catabolism used under microaerobic conditions; also has mRNA binding

						activity; member of the aldo-keto reductase (AKR) family; protein abundance increases in response to DNA replication stress; GCY1 has a paralog, YPR1, that arose from the whole genome duplication
GDH1_1	-2.43	-8.55	1.28E-17	6.02E-16	YOR375C	NADP(+)-dependent glutamate dehydrogenase; synthesizes glutamate from ammonia and alpha-ketoglutarate; rate of alpha-ketoglutarate utilization differs from Gdh3p; expression regulated by nitrogen and carbon sources; GDH1 has a paralog, GDH3, that arose from the whole genome duplication
GFD2_1	-2.98	-5.04	4.74E-07	6.78E-06	YCL036W	Protein of unknown function; identified as a high-copy suppressor of a dbp5 mutation; GFD2 has a paralog, YDR514C, that arose from the whole genome duplication
GLK1_1	-2.05	-11.83	2.72E-32	1.92E-30	YCL040W	Glucokinase; catalyzes the phosphorylation of glucose at C6 in the first irreversible step of glucose metabolism; one of three glucose phosphorylating enzymes; expression regulated by non-fermentable carbon sources; GLK1 has a paralog, EMI2, that arose from the whole genome duplication
GRX4_1	-2.75	-8.18	2.95E-16	1.25E-14	YER174C	Glutathione-dependent oxidoreductase; hydroperoxide and superoxide-radical responsive; monothiol glutaredoxin subfamily member along with Grx3p and Grx5p; protects cells from oxidative damage; with Grx3p, binds to Aft1p in iron-replete conditions, promoting its dissociation from promoters; mutant has increased aneuploidy tolerance; transcription regulated by Yap5p; GRX4 has a paralog, GRX3, that arose from the whole genome duplication
HCM1_1	-2.79	-4.01	6.01E-05	5.62E-04	YCR065W	Forkhead transcription factor; drives S-phase specific expression of genes involved in chromosome segregation, spindle dynamics, and budding; suppressor of calmodulin mutants with specific SPB assembly defects; telomere maintenance role
HIS4_1	-4.77	-17.28	7.29E-67	7.24E-65	YCL030C	Multifunctional enzyme containing phosphoribosyl-ATP pyrophosphatase; phosphoribosyl-AMP cyclohydrolase, and histidinol dehydrogenase

						activities; catalyzes the second, third, ninth and tenth steps in histidine biosynthesis
HMS2_1	-2.19	-6.80	1.02E-11	2.79E-10	YJR147W	Protein with similarity to heat shock transcription factors; overexpression suppresses the pseudohyphal filamentation defect of a diploid mep1 mep2 homozygous null mutant; HMS2 has a paralog, SKN7, that arose from the whole genome duplication
HOM2_1	-2.11	-16.79	2.98E-63	2.92E-61	YDR158W	Aspartic beta semi-aldehyde dehydrogenase; catalyzes the second step in the common pathway for methionine and threonine biosynthesis; expression regulated by Gcn4p and the general control of amino acid synthesis
HOT13_1	-2.18	-4.84	1.32E-06	1.76E-05	YKL084W	Zinc-binding mitochondrial intermembrane space (IMS) protein; involved in a disulfide relay system for IMS import of cysteine-containing proteins; binds Mia40p and stimulates its Erv1p-dependent oxidation, probably by sequestering zinc
ILV3_1	-2.61	-12.36	4.57E-35	3.36E-33	YJR016C	Dihydroxyacid dehydratase; catalyzes third step in the common pathway leading to biosynthesis of branched-chain amino acids
ILV5_1	-2.84	-29.39	8.32E-190	1.04E-187	YLR355C	Acetohydroxyacid reductoisomerase and mtDNA binding protein; involved in branched-chain amino acid biosynthesis and maintenance of wild-type mitochondrial DNA; found in mitochondrial nucleoids
ILV6_1	-3.15	-17.71	3.76E-70	3.86E-68	YCL009C	Regulatory subunit of acetolactate synthase; acetolactate synthase catalyzes the first step of branched-chain amino acid biosynthesis; enhances activity of the Ilv2p catalytic subunit, localizes to mitochondria
ISU2_1	-4.33	-23.95	8.63E-127	1.04E-124	YOR226C	Protein required for synthesis of iron-sulfur proteins; localized to the mitochondrial matrix; performs a scaffolding function in mitochondria during Fe/S cluster assembly; involved in Fe-S cluster assembly for both mitochondrial and cytosolic proteins; isu1 isu2 double mutant is inviable; protein abundance increases in response to DNA replication stress; evolutionarily conserved; ISU2 has a paralog, ISU1, that arose from the whole genome duplication

KCC4_1	-3.55	-3.93	8.40E-05	7.65E-04	YCL024W	Protein kinase of the bud neck involved in the septin checkpoint; associates with septin proteins, negatively regulates Swe1p by phosphorylation, shows structural homology to bud neck kinases Gin4p and Hsl1p; KCC4 has a paralog, GIN4, that arose from the whole genome duplication
LAP3_1	-2.30	-18.68	7.08E-78	7.59E-76	YNL239W	Cysteine aminopeptidase with homocysteine-thiolactonase activity; protects cells against homocysteine toxicity; has bleomycin hydrolase activity in vitro; transcription is regulated by galactose via Gal4p; orthologous to human BLMH
LDB16_1	-2.38	-5.75	9.10E-09	1.69E-07	YCL005W	Protein of unknown function; null mutants have decreased net negative cell surface charge; GFP-fusion protein expression is induced in response to the DNA-damaging agent MMS; native protein is detected in purified mitochondria
LEU1_1	-5.38	-15.74	8.17E-56	7.57E-54	YGL009C	Isopropylmalate isomerase; catalyzes the second step in the leucine biosynthesis pathway
LEU9_1	-3.45	-10.86	1.73E-27	1.13E-25	YOR108W	Alpha-isopropylmalate synthase II (2-isopropylmalate synthase); catalyzes the first step in the leucine biosynthesis pathway; the minor isozyme, responsible for the residual alpha-IPMS activity detected in a leu4 null mutant; LEU9 has a paralog, LEU4, that arose from the whole genome duplication
LSB5_1	-2.36	-9.74	2.03E-22	1.19E-20	YCL034W	Protein of unknown function; binds Las17p, which is a homolog of human Wiskott-Aldrich Syndrome protein involved in actin patch assembly and actin polymerization; may mediate disassembly of the Pan1 complex from the endocytic coat
LYS20_1	-2.73	-7.22	5.34E-13	1.66E-11	YDL182W	Homocitrate synthase isozyme; catalyzes the condensation of acetyl-CoA and alpha-ketoglutarate to form homocitrate, which is the first step in the lysine biosynthesis pathway; LYS20 has a paralog, LYS21, that arose from the whole genome duplication
LYS9_1	-2.18	-9.56	1.15E-21	6.47E-20	YNR050C	Saccharopine dehydrogenase (NADP+, L-glutamate-forming); catalyzes the formation of saccharopine from alpha-aminoadipate 6-semialdehyde, the seventh step in lysine

						biosynthesis pathway; exhibits genetic and physical interactions with TRM112
MAE1_1	-2.52	-13.37	8.64E-41	7.21E-39	YKL029C	Mitochondrial malic enzyme; catalyzes the oxidative decarboxylation of malate to pyruvate, which is a key intermediate in sugar metabolism and a precursor for synthesis of several amino acids
MAK31_1	-2.06	-6.09	1.13E-09	2.36E-08	YCR020C-A	Non-catalytic subunit of N-terminal acetyltransferase of the NatC type; required for replication of dsRNA virus; member of the Sm protein family
MEP3_1	-2.07	-6.44	1.17E-10	2.73E-09	YPR138C	Ammonium permease of high capacity and low affinity; belongs to a ubiquitous family of cytoplasmic membrane proteins that transport only ammonium (NH4+); expression is under the nitrogen catabolite repression regulation ammonia permease; MEP3 has a paralog, MEP1, that arose from the whole genome duplication
MGA1_1	-2.21	-4.14	3.46E-05	3.43E-04	YGR249W	Protein similar to heat shock transcription factor; multicopy suppressor of pseudohyphal growth defects of ammonium permease mutants
MSB2_1	-2.24	-5.78	7.28E-09	1.38E-07	YGR014W	Mucin family member involved in various signaling pathways; functions as osmosensor in Sho1p-mediated HOG pathway with Msb2p; functions in Cdc42p- and MAP kinase-dependent filamentous growth signaling pathway; processed into secreted and cell-associated forms by aspartyl protease Yps1p; potential Cdc28p substrate
MSH3_1	-2.80	-3.56	3.68E-04	2.84E-03	YCR092C	Mismatch repair protein; forms dimers with Msh2p that mediate repair of insertion or deletion mutations and removal of nonhomologous DNA ends, contains a PCNA (Pol30p) binding motif required for genome stability
MXR2_1	-2.36	-8.11	5.19E-16	2.15E-14	YCL033C	Methionine-R-sulfoxide reductase; involved in the response to oxidative stress; protects iron-sulfur clusters from oxidative inactivation along with MXR1; involved in the regulation of lifespan

NCA3_1	-2.67	-17.34	2.37E-67	2.39E-65	YJL116C	Protein involved in mitochondrion organization; functions with Nca2p to regulate mitochondrial expression of subunits 6 (Atp6p) and 8 (Atp8p) of the Fo-F1 ATP synthase; member of the SUN family; expression induced in cells treated with the mycotoxin patulin; NCA3 has a paralog, UTH1, that arose from the whole genome duplication
NDJ1_1	-3.28	-7.85	4.22E-15	1.64E-13	YOL104C	Meiosis-specific telomere protein; required for bouquet formation, effective homolog pairing, ordered cross-over distribution, sister chromatid cohesion at meiotic telomeres, chromosomal segregation and telomere-led rapid prophase movement
NEL1_1	-3.08	-3.64	2.70E-04	2.14E-03	YHR035W	Activator of Sar1p GTPase activity; paralog of Sec23 but does not associate with the COPII components; not an essential gene
NPP1_1	-2.45	-6.95	3.71E-12	1.05E-10	YCR026C	Nucleotide pyrophosphatase/phosphodiesterase; mediates extracellular nucleotide phosphate hydrolysis along with Npp2p and Pho5p; activity and expression enhanced during conditions of phosphate starvation; involved in spore wall assembly; NPP1 has a paralog, NPP2, that arose from the whole genome duplication, and an npp1 npp2 double mutant exhibits reduced dityrosine fluorescence relative to the single mutants
NRT1_1	-2.25	-8.33	8.04E-17	3.55E-15	YOR071C	High-affinity nicotinamide riboside transporter; also transports thiamine with low affinity; major transporter for 5-aminoimidazole-4-carboxamide-1-beta-D-ribofuranoside (acadesine) uptake; shares sequence similarity with Thi7p and Thi72p; proposed to be involved in 5-fluorocytosine sensitivity
NSE1_1	-2.08	-6.69	2.25E-11	5.78E-10	YLR007W	Component of the SMC5-SMC6 complex; this complex plays a key role in the removal of X-shaped DNA structures that arise between sister chromatids during DNA replication and repair
OAC1_1	-3.22	-18.31	6.60E-75	6.97E-73	YKL120W	Mitochondrial inner membrane transporter; transports oxaloacetate, sulfate, thiosulfate, and isopropylmalate; member of the mitochondrial carrier family

ODC2_1	-2.67	-10.68	1.32E-26	8.47E-25	YOR222W	Mitochondrial inner membrane transporter; exports 2-oxoadipate and 2-oxoglutarate from the mitochondrial matrix to the cytosol for use in lysine and glutamate biosynthesis and in lysine catabolism; ODC2 has a paralog, ODC1, that arose from the whole genome duplication
OPT2_1	-2.50	-5.25	1.48E-07	2.34E-06	YPR194C	Oligopeptide transporter; member of the OPT family, with potential orthologs in S. pombe and C. albicans; also plays a role in formation of mature vacuoles
PCL10_1	-2.52	-9.42	4.65E-21	2.58E-19	YGL134W	Pho85p cyclin; recruits, activates, and targets Pho85p cyclin-dependent protein kinase to its substrate; PCL10 has a paralog, PCL8, that arose from the whole genome duplication
PCL7_1	-2.39	-8.44	3.16E-17	1.43E-15	YIL050W	Pho85p cyclin of the Pho80p subfamily; forms a functional kinase complex with Pho85p which phosphorylates Mmr1p and is regulated by Pho81p; involved in glycogen metabolism, expression is cell-cycle regulated; PCL7 has a paralog, PCL6, that arose from the whole genome duplication
PDI1_1	-2.39	-19.24	1.82E-82	2.01E-80	YCL043C	Protein disulfide isomerase; multifunctional protein of ER lumen, essential for formation of disulfide bonds in secretory and cell-surface proteins, unscrambles non-native disulfide bonds; key regulator of Ero1p; forms complex with Mnl1p that has exomannosidase activity, processing unfolded protein-bound Man8GlcNAc2 oligosaccharides to Man7GlcNAc2, promoting degradation in unfolded protein response; PDI1 has a paralog, EUG1, that arose from the whole genome duplication
PER1_1	-2.22	-4.76	1.96E-06	2.55E-05	YCR044C	Protein of the endoplasmic reticulum; required for GPI-phospholipase A2 activity that remodels the GPI anchor as a prerequisite for association of GPI-anchored proteins with lipid rafts; functionally complemented by human ortholog PERLD1
PET18_1	-2.85	-5.86	4.65E-09	9.08E-08	YCR020C	Protein of unknown function; has weak similarity to proteins involved in thiamin metabolism; expression is induced in the absence of thiamin

PGK1_1	-2.78	-15.86	1.30E-56	1.22E-54	YCR012W	3-phosphoglycerate kinase; catalyzes transfer of high-energy phosphoryl groups from the acyl phosphate of 1,3-bisphosphoglycerate to ADP to produce ATP; key enzyme in glycolysis and gluconeogenesis
PLB3_1	-2.97	-12.61	1.82E-36	1.37E-34	YOL011W	Phospholipase B (lysophospholipase) involved in lipid metabolism; hydrolyzes phosphatidylinositol and phosphatidylserine and displays transacylase activity in vitro; PLB3 has a paralog, PLB1, that arose from the whole genome duplication
PMA2_1	-3.02	-5.98	2.24E-09	4.51E-08	YPL036W	Plasma membrane H+-ATPase; isoform of Pma1p, involved in pumping protons out of the cell; regulator of cytoplasmic pH and plasma membrane potential
PMP1_1	-2.05	-13.33	1.59E-40	1.31E-38	YCR024C-A	Regulatory subunit for the plasma membrane H(+)-ATPase Pma1p; small single-membrane span proteolipid; forms unique helix and positively charged cytoplasmic domain that is able to specifically segregate phosphatidylserines; PMP1 has a paralog, PMP2, that arose from the whole genome duplication
POF1_1	-2.11	-4.20	2.65E-05	2.70E-04	YCL047C	Nicotinamide mononucleotide-specific adenylyltransferase (NMNAT); catalyzes the conversion of nicotinamide mononucleotide (NMN) to nicotinamide adenine dinucleotide (NAD+); role in the nicotinamide riboside (NR) salvage pathway of NAD+ biosynthesis; involved in NR and NAD+ homeostasis; ATPase involved in protein quality control and filamentation pathways; interacts physically with Kss1p and suppresses the filamentation defect of a kss1 deletion
POL4_1	-2.15	-3.62	2.95E-04	2.31E-03	YCR014C	DNA polymerase IV; undergoes pair-wise interactions with Dnl4p-Lif1p and Rad27p to mediate repair of DNA double-strand breaks by non-homologous end joining (NHEJ); homologous to mammalian DNA polymerase beta
PRD1_1	-2.13	-6.57	5.07E-11	1.23E-09	YCL057W	Zinc metalloendopeptidase; found in the cytoplasm and intermembrane space of mitochondria; with Cym1p, involved in degradation of mitochondrial proteins and of presequence peptides cleaved from imported proteins; protein abundance increases in response to DNA replication stress

PRY2_1	-2.09	-6.18	6.41E-10	1.39E-08	YKR013W	Sterol binding protein involved in the export of acetylated sterols; secreted glycoprotein and member of the CAP protein superfamily (cysteine-rich secretory proteins (CRISP), antigen 5, and pathogenesis related 1 proteins); sterol export function is redundant with that of PRY1; may be involved in detoxification of hydrophobic compounds; PRY2 has a paralog, PRY1, that arose from the whole genome duplication
PTR2_1	-2.23	-3.35	7.98E-04	5.58E-03	YKR093W	Integral membrane peptide transporter; mediates transport of di- and tri-peptides; conserved protein that contains 12 transmembrane domains; PTR2 expression is regulated by the N-end rule pathway via repression by Cup9p
PWP2_1	-2.18	-5.99	2.12E-09	4.29E-08	YCR057C	Conserved 90S pre-ribosomal component; essential for proper endonucleolytic cleavage of the 35 S rRNA precursor at A0, A1, and A2 sites; contains eight WD-repeats; PWP2 deletion leads to defects in cell cycle and bud morphogenesis
QDR2_1	-2.69	-12.40	2.64E-35	1.96E-33	YIL121W	Plasma membrane transporter of the major facilitator superfamily; member of the 12-spanner drug:H(+) antiporter DHA1 family; exports copper; has broad substrate specificity and can transport many mono- and divalent cations; transports a variety of drugs and is required for resistance to quinidine, barban, cisplatin, and bleomycin; contributes to potassium homeostasis; expression is regulated by copper
RER1_1	-2.43	-9.10	8.67E-20	4.54E-18	YCL001W	Protein involved in retention of membrane proteins; including Sec12p, in the ER; localized to Golgi; functions as a retrieval receptor in returning membrane proteins to the ER
RGL1_1	-3.01	-16.52	2.59E-61	2.50E-59	YPL066W	Regulator of Rho1p signaling, cofactor of Tus1p; required for the localization of Tus1p during all phases of cytokinesis; green fluorescent protein (GFP)-fusion protein localizes to the bud neck and cytoplasm; null mutant is viable and exhibits growth defect on a non-fermentable (respiratory) carbon source
RIB4_1	-2.17	-14.24	5.10E-46	4.48E-44	YOL143C	Lumazine synthase (DMRL synthase); catalyzes synthesis of immediate precursor to riboflavin; DMRL

						synthase stands for 6,7-dimethyl-8-ribityllumazine synthase
RNQ1_1	-2.13	-8.96	3.20E-19	1.65E-17	YCL028W	[PIN(+)] prion; an infectious protein conformation that is generally an ordered protein aggregate
RPS14A_1	-2.66	-25.45	6.47E-143	7.95E-141	YCR031C	Protein component of the small (40S) ribosomal subunit; required for ribosome assembly and 20S pre-rRNA processing; mutations confer cryptopleurine resistance; homologous to mammalian ribosomal protein S14 and bacterial S11; RPS14A has a paralog, RPS14B, that arose from the whole genome duplication
RPS9A_1	-2.66	-6.35	2.09E-10	4.75E-09	YPL081W	Protein component of the small (40S) ribosomal subunit; homologous to mammalian ribosomal protein S9 and bacterial S4; RPS9A has a paralog, RPS9B, that arose from the whole genome duplication
RRT12_1	-2.59	-4.50	6.65E-06	7.90E-05	YCR045C	Probable subtilisin-family protease; role in formation of the dityrosine layer of spore walls; localizes to the spore wall and also the nuclear envelope and ER region in mature spores
RUP1_1	-2.04	-9.20	3.66E-20	1.93E-18	YOR138C	Protein that regulates ubiquitination of Rsp5p; has a WW domain consensus motif of PPPSY (residues 131-135) that mediates binding of Rsp5p to Ubp2p; contains an UBA domain; relative distribution to the nucleus increases upon DNA replication stress
SAT4_1	-2.26	-7.28	3.46E-13	1.11E-11	YCR008W	Ser/Thr protein kinase involved in salt tolerance; funtions in regulation of Trk1p-Trk2p potassium transporter; partially redundant with Hal5p; has similarity to Npr1p
SFG1_1	-2.61	-6.50	8.12E-11	1.91E-09	YOR315W	Nuclear protein putative transcription factor; required for growth of superficial pseudohyphae (which do not invade the agar substrate) but not for invasive pseudohyphal growth; may act together with Phd1p; potential Cdc28p substrate
SGF29_1	-2.18	-7.66	1.81E-14	6.48E-13	YCL010C	Component of the HAT/Core module of the SAGA, SLIK, and ADA complexes; HAT/Core module also contains Gcn5p, Ngg1p, and Ada2p; binds methylated histone H3K4; involved in transcriptional

						regulation through SAGA and TBP recruitment to target promoters and H3 acetylation
SOL2_1	-2.27	-8.06	7.41E-16	3.03E-14	YCR073W-A	Protein with a possible role in tRNA export; shows similarity to 6-phosphogluconolactonase non-catalytic domains but does not exhibit this enzymatic activity; homologous to Sol3p and Sol4p; SOL2 has a paralog, SOL1, that arose from the whole genome duplication
SRD1_1	-5.12	-13.77	3.70E-43	3.17E-41	YCR018C	Protein involved in the processing of pre-rRNA to mature rRNA; contains a C2/C2 zinc finger motif; srd1 mutation suppresses defects caused by the rrp1-1 mutation
SSK22_1	-2.45	-3.83	1.30E-04	1.12E-03	YCR073C	MAP kinase kinase kinase of HOG1 mitogen-activated signaling pathway; functionally redundant with Ssk2p; interacts with and is activated by Ssk1p; phosphorylates Pbs2p; SSK22 has a paralog, SSK2, that arose from the whole genome duplication
SSU1_1	-2.16	-6.52	7.11E-11	1.70E-09	YPL092W	Plasma membrane sulfite pump involved in sulfite metabolism; required for efficient sulfite efflux; major facilitator superfamily protein
STP22_1	-2.42	-5.81	6.16E-09	1.18E-07	YCL008C	Component of the ESCRT-I complex; ESCRT-I is involved in ubiquitin-dependent sorting of proteins into the endosome; prevents polyubiquitination of the arrestin-related protein Rim8p, thereby directing its monoubiquitination by Rsp5p; homologous to the mouse and human Tsg101 tumor susceptibility gene; mutants exhibit a Class E Vps phenotype;
STR3_1	-2.64	-7.00	2.57E-12	7.32E-11	YGL184C	Peroxisomal cystathionine beta-lyase; converts cystathionine into homocysteine; may be redox regulated by Gto1p; involved in the release of the aromatic thiol 3-mercaptohexanol during wine fermentation
TAH1_1	-2.26	-7.22	5.06E-13	1.58E-11	YCR060W	Component of conserved R2TP complex (Rvb1-Rvb2-Tah1-Pih1); R2TP complex interacts with Hsp90 (Hsp82p and Hsc82p) to mediate assembly of large protein complexes such as box C/D snoRNPs and RNA polymerase II; contains a single TPR

						domain with at least two TPR motifs; plays a role in determining prion variants
TAT1_1	-2.99	-13.15	1.81E-39	1.46E-37	YBR069C	Amino acid transporter for valine, leucine, isoleucine, and tyrosine; low-affinity tryptophan and histidine transporter; overexpression confers FK506 and FTY720 resistance; protein abundance increases in response to DNA replication stress
THR4_1	-2.07	-13.17	1.31E-39	1.07E-37	YCR053W	Threonine synthase; conserved protein that catalyzes formation of threonine from O-phosphohomoserine; expression is regulated by the GCN4-mediated general amino acid control pathway
TIR1_1	-2.08	-3.46	5.31E-04	3.91E-03	YER011W	Cell wall mannoprotein; expression is downregulated at acidic pH and induced by cold shock and anaerobiosis; abundance is increased in cells cultured without shaking; member of the Srp1p/Tip1p family of serine-alanine-rich proteins
TRX3_1	-2.10	-7.88	3.24E-15	1.27E-13	YCR083W	Mitochondrial thioredoxin; highly conserved oxidoreductase required to maintain the redox homeostasis of the cell, forms the mitochondrial thioredoxin system with Trr2p, redox state is maintained by both Trr2p and Glr1p
TSA2_1	-2.45	-7.98	1.44E-15	5.81E-14	YDR453C	Stress inducible cytoplasmic thioredoxin peroxidase; cooperates with Tsa1p in the removal of reactive oxygen, nitrogen and sulfur species using thioredoxin as hydrogen donor; deletion enhances the mutator phenotype of tsa1 mutants; protein abundance increases in response to DNA replication stress; TSA2 has a paralog, TSA1, that arose from the whole genome duplication
TYW1_1	-2.01	-7.66	1.87E-14	6.64E-13	YPL207W	Iron-sulfer protein required for synthesis of Wybutosine modified tRNA; Wybutosine is a modified guanosine found at the 3'-position adjacent to the anticodon of phenylalanine tRNA which supports reading frame maintenance by stabilizing codon-anticodon interactions; induction by Yap5p in response to iron provides protection from high iron

						toxicity; overexpression results in increased cellular iron
URA1_1	-3.75	-20.52	1.33E-93	1.56E-91	YKL216W	Dihydroorotate dehydrogenase; catalyzes the fourth enzymatic step in the de novo biosynthesis of pyrimidines, converting dihydroorotic acid into orotic acid
VAC17_1	-2.27	-3.48	4.93E-04	3.66E-03	YCL063W	Phosphoprotein involved in vacuole inheritance; degraded in late M phase of the cell cycle; acts as a vacuole-specific receptor for myosin Myo2p
VMA9_1	-2.44	-16.20	5.30E-59	5.05E-57	YCL005W-A	Vacuolar H+ ATPase subunit e of the V-ATPase V0 subcomplex; essential for vacuolar acidification; interacts with the V-ATPase assembly factor Vma21p in the ER; involved in V0 biogenesis
WSC3_1	-2.38	-6.54	6.35E-11	1.53E-09	YOL105C	Sensor-transducer of the stress-activated PKC1-MPK1 signaling pathway; involved in maintenance of cell wall integrity; involved in response to heat shock and other stressors; regulates 1,3-beta-glucan synthesis; WSC3 has a paralog, WSC2, that arose from the whole genome duplication
YAR029W_1	-2.07	-4.03	5.49E-05	5.18E-04	YAR029W	Member of DUP240 gene family but contains no transmembrane domains; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm in a punctate pattern
YBL005W-B_1	-2.61	-6.66	2.83E-11	7.17E-10		
YBL113C_1	-2.05	-3.56	3.73E-04	2.87E-03	YBL113C	Putative Y' element ATP-dependent helicase
YBR012C_1	-2.93	-7.51	5.83E-14	1.99E-12		
YBR012W-B_1	-2.68	-6.68	2.40E-11	6.15E-10		
YCL002C_1	-2.06	-4.40	1.07E-05	1.20E-04	YCL002C	Putative protein of unknown function; YCL002C is not an essential gene
YCL012C_1	-2.29	-5.28	1.26E-07	2.00E-06	YCL012C	Putative protein of unknown function; orthologs are present in S. bayanus, S. paradoxus and Ashbya gossypii; YCL012C is not an essential gene

YCL021W-A_1	-3.16	-4.25	2.13E-05	2.23E-04	YCL021W-A	Putative protein of unknown function
YCL041C_1	-2.32	-5.73	9.89E-09	1.82E-07		
YCL048W-A_1	-2.58	-6.86	6.87E-12	1.91E-10		
YCL049C_1	-2.29	-8.51	1.67E-17	7.78E-16	YCL049C	Protein of unknown function; localizes to membrane fraction; YCL049C is not an essential gene
YCP4_1	-2.37	-13.08	4.08E-39	3.24E-37	YCR004C	Protein of unknown function; has sequence and structural similarity to flavodoxins; predicted to be palmitoylated; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies
YCR007C_1	-2.73	-5.45	5.12E-08	8.58E-07	YCR007C	Putative integral membrane protein; member of DUP240 gene family; YCR007C is not an essential gene
YCR013C_1	-2.82	-14.71	5.42E-49	4.95E-47		
YCR023C_1	-2.43	-5.92	3.23E-09	6.38E-08	YCR023C	Vacuolar membrane protein of unknown function; member of the multidrug resistance family; YCR023C is not an essential gene
YCR024C-B_1	-2.13	-14.60	2.67E-48	2.41E-46	YCR024C-B	Putative protein of unknown function; identified by expression profiling and mass spectrometry
YCR025C_1	-4.48	-3.67	2.47E-04	1.97E-03		
YCR041W_1	-2.03	-4.58	4.59E-06	5.63E-05		
YCR043C_1	-2.46	-5.40	6.51E-08	1.07E-06	YCR043C	Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the Golgi apparatus; YCR043C is not an essential gene
YCR051W_1	-2.24	-10.49	9.73E-26	6.09E-24	YCR051W	Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm and nucleus; contains ankyrin (Ank) repeats; YCR051W is not an essential gene
YCR061W_1	-2.71	-6.06	1.34E-09	2.75E-08	YCR061W	Protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm in a punctate pattern; induced by treatment with 8-methoxypsoralen and UVA irradiation

YCR075W-A_1	-2.20	-5.11	3.27E-07	4.86E-06	YCR075W-A	Putative protein of unknown function; identified by homology to Ashbya gossypii; YCR075W-A has a paralog, YNR034W-A, that arose from the whole genome duplication
YCR087W_1	-2.22	-7.59	3.29E-14	1.15E-12		
YCR090C_1	-2.25	-6.51	7.49E-11	1.78E-09	YCR090C	Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm and nucleus; YCR090C is not an essential gene
YDR210C-D_1	-2.64	-3.49	4.78E-04	3.57E-03		
YDR261C-D_1	-2.60	-5.11	3.20E-07	4.78E-06	YDR261C-D	Retrotransposon TYA Gag and TYB Pol genes; transcribed/translated as one unit; polyprotein is processed to make a nucleocapsid-like protein (Gag), reverse transcriptase (RT), protease (PR), and integrase (IN); similar to retroviral genes
YDR365W-B_1	-4.32	-8.15	3.65E-16	1.53E-14		
YEL057C_1	-2.82	-9.21	3.12E-20	1.66E-18	YEL057C	Protein of unknown function involved in telomere maintenance; target of UME6 regulation
YER068C-A_1	-5.34	-4.51	6.43E-06	7.68E-05		
YER152C_1	-2.30	-7.44	1.01E-13	3.40E-12	YER152C	Protein with 2-aminoadipate transaminase activity; shares amino acid similarity with the aminotransferases Aro8p and Aro9p; YER152C is not an essential gene
YFL067W_1	-2.55	-6.11	9.66E-10	2.03E-08	YFL067W	Protein of unknown function; down-regulated at low calcium levels
YFR020W_1	-7.32	-18.77	1.40E-78	1.52E-76		
YGL081W_1	-2.22	-3.87	1.08E-04	9.45E-04	YGL081W	Putative protein of unknown function; non-essential gene; interacts genetically with CHS5, a gene involved in chitin biosynthesis
YGL262W_1	-6.64	-3.75	1.76E-04	1.46E-03	YGL262W	Putative protein of unknown function; null mutant displays elevated sensitivity to expression of a mutant huntingtin fragment or of alpha-synuclein; YGL262W is not an essential gene
YGR079W_1	-2.12	-6.33	2.50E-10	5.65E-09	YGR079W	Putative protein of unknown function; YGR079W is not an essential gene

YHP1_1	-2.30	-5.94	2.87E-09	5.71E-08	YDR451C	Homeobox transcriptional repressor; binds Mcm1p and early cell cycle box (ECB) elements of cell cycle regulated genes, thereby restricting ECB-mediated transcription to the M/G1 interval; YHP1 has a paralog, YOX1, that arose from the whole genome duplication
YIH1_1	-2.50	-10.01	1.34E-23	8.02E-22	YCR059C	Negative regulator of eIF2 kinase Gcn2p; competes with Gcn2p for binding to Gcn1p; may contribute to regulation of translation in response to starvation via regulation of Gcn2p; binds to monomeric actin and to ribosomes and polyribosomes; ortholog of mammalian IMPACT
YJR027W_1	-2.61	-4.62	3.78E-06	4.70E-05		
YJR115W_1	-4.93	-8.33	8.10E-17	3.56E-15	YJR115W	Putative protein of unknown function; YJR115W has a paralog, ECM13, that arose from the whole genome duplication
YKL030W_1	-2.82	-6.79	1.09E-11	2.97E-10		
YLR035C-A_1	-2.88	-3.36	7.68E-04	5.38E-03		
YLR042C_1	-2.09	-4.41	1.05E-05	1.18E-04	YLR042C	Cell wall protein of unknown function; localizes to the cytoplasm; deletion improves xylose fermentation in industrially engineered strains; YLL042C is not an essential gene
YLR152C_1	-2.20	-7.01	2.30E-12	6.65E-11	YLR152C	Putative protein of unknown function; YLR152C is not an essential gene
YLR342W-A_1	-3.31	-4.39	1.12E-05	1.24E-04	YLR342W-A	Putative protein of unknown function
YLR437C-A_1	-3.95	-4.79	1.68E-06	2.22E-05		
YML122C_1	-4.14	-5.67	1.46E-08	2.62E-07		
YMR045C_1	-2.38	-6.14	8.26E-10	1.76E-08		
YMR244W_1	-2.69	-3.89	1.02E-04	8.98E-04	YMR244W	Putative protein of unknown function
YNL284C-B_1	-2.30	-7.52	5.66E-14	1.94E-12		

YOL103W-B_1	-2.84	-4.09	4.31E-05	4.17E-04		
YOR121C_1	-3.26	-19.39	8.92E-84	1.01E-81		
YOR225W_1	-3.91	-12.15	6.03E-34	4.34E-32		
YPL014W_1	-2.31	-10.87	1.55E-27	1.03E-25	YPL014W	Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm and to the nucleus
YPL067C_1	-3.58	-10.63	2.09E-26	1.32E-24	YPL067C	Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm; YPL067C is not an essential gene
YPR002C-A_1	-2.76	-6.82	9.18E-12	2.53E-10		
YPR158C-D_1	-3.13	-7.04	1.89E-12	5.54E-11		

B. BY4741*adh1*Δ and BY4741*adh1*Δ_#800-1454-903

Feature ID	Experiment - Fold Change (normalized values)	Baggerley's test: Host_Pathway vs Host normalized values - Test statistic	Baggerley's test: Host_Pathway vs Host normalized values - P-value	Annotations - Transcript ID	Annotations - Gene title
AAP1_1	2.08	3.80	1.44E-04	YHR047C	Arginine/alanine amino peptidase; overproduction stimulates glycogen accumulation; AAP1 has a paralog, APE2, that arose from the whole genome duplication
ADH1_1	1007.55	15.21	0	YOL086C	Alcohol dehydrogenase; fermentative isozyme active as homo- or heterotetramers; required for the reduction of acetaldehyde to ethanol, the last step in the glycolytic pathway; ADH1 has a paralog, ADH5, that arose from the whole genome duplication
AFR1_1	2.68	3.75	1.76E-04	YDR085C	Protein required for pheromone-induced projection (shmoo) formation; regulates septin architecture during mating; has an RVXF motif that mediates targeting of Glc7p to mating projections; interacts with Cdc12p; AFR1 has a paralog, YER158C, that arose from the whole genome duplication
AGA2_1	21.01	4.03	5.48E-05	YGL032C	Adhesion subunit of a-agglutinin of a-cells; C-terminal sequence acts as a ligand for alpha-agglutinin (Sag1p) during agglutination, modified with O-linked oligomannosyl chains, linked to anchorage subunit Aga1p via two disulfide bonds
AIM17_1	2.58	3.79	1.49E-04	YHL021C	Putative protein of unknown function; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies; null mutant displays reduced frequency of mitochondrial genome loss
ARO9_1	2.94	4.61	4.02E-06	YHR137W	Aromatic aminotransferase II; catalyzes the first step of tryptophan, phenylalanine, and tyrosine catabolism

BTN2_1	2.87	7.68	1.60E-14	YGR142W	v-SNARE binding protein; facilitates specific protein retrieval from a late endosome to the Golgi; modulates arginine uptake, possible role in mediating pH homeostasis between the vacuole and plasma membrane H(+)-ATPase; contributes to prion curing; BTN2 has a paralog, CUR1, that arose from the whole genome duplication
CSR2_1	2.37	4.50	6.82E-06	YPR030W	Nuclear ubiquitin protein ligase binding protein; may regulate utilization of nonfermentable carbon sources and endocytosis of plasma membrane proteins; overproduction suppresses chs5 spa2 lethality at high temp; ubiquitinated by Rsp5p, deubiquitinated by Ubp2p; CSR2 has a paralog, ECM21, that arose from the whole genome duplication
CUR1_1	3.67	6.08	1.18E-09	YPR158W	Sorting factor, central regulator of spatial protein quality control; physically and functionally interacts with chaperones to promote sorting and deposition of misfolded proteins into cytosolic compartments; involved in destabilization of [URE3] prions; CUR1 has a paralog, BTN2, that arose from the whole genome duplication
CYC7_1	3.02	3.53	4.21E-04	YEL039C	Cytochrome c isoform 2, expressed under hypoxic conditions; electron carrier of the mitochondrial intermembrane space that transfers electrons from ubiquinone-cytochrome c oxidoreductase to cytochrome c oxidase during cellular respiration; protein abundance increases in response to DNA replication stress; CYC7 has a paralog, CYC1, that arose from the whole genome duplication
DAK2_1	5.48	3.67	2.43E-04	YFL053W	Dihydroxyacetone kinase; required for detoxification of dihydroxyacetone (DHA); involved in stress adaptation
DSF1_1	2.74	3.33	8.64E-04	YEL070W	Putative mannitol dehydrogenase; YNR073C has a paralog, DSF1, that arose from a segmental duplication /// Putative mannitol dehydrogenase; YNR073C has a paralog, DSF1, that arose from a segmental duplication

ENB1_1	2.47	4.92	8.51E-07	YOL158C	Endosomal ferric enterobactin transporter; expressed under conditions of iron deprivation; member of the major facilitator superfamily; expression is regulated by Rcs1p and affected by chloroquine treatment
ERG5_1	2.06	4.16	3.19E-05	YMR015C	C-22 sterol desaturase; a cytochrome P450 enzyme that catalyzes the formation of the C-22(23) double bond in the sterol side chain in ergosterol biosynthesis; may be a target of azole antifungal drugs
FET3_1	2.60	4.10	4.19E-05	YMR058W	Ferro-O2-oxidoreductase; multicopper oxidase that oxidizes ferrous (Fe2+) to ferric iron (Fe3+) for subsequent cellular uptake by transmembrane permease Ftr1p; required for high-affinity iron uptake and involved in mediating resistance to copper ion toxicity, belongs to class of integral membrane multicopper oxidases; protein abundance increases in response to DNA replication stress
FMP23_1	2.21	4.55	5.36E-06	YBR047W	Putative protein of unknown function; proposed to be involved in iron or copper homeostasis; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies
FUS2_1	2.78	3.43	6.05E-04	YMR232W	Cell fusion regulator; cytoplasmic protein localized to shmoo tip; required for alignment of parental nuclei before nuclear fusion during mating; contains a Dbl-homology domain; binds specifically with activated Cdc42p
GDB1_1	3.02	3.60	3.16E-04	YPR184W	Glycogen debranching enzyme; contains glucanotranferase and alpha-1,6-amyloglucosidase activities; required for glycogen degradation; phosphorylated in mitochondria; activity is inhibited by Igd1p; protein abundance increases in response to DNA replication stress
GDE1_1	2.07	4.25	2.15E-05	YPL110C	Glycerophosphocholine (GroPCho) phosphodiesterase; hydrolyzes GroPCho to choline and glycerolphosphate, for use as a phosphate source and as a precursor for phosphocholine synthesis; may interact with ribosomes

GOR1_1	2.51	3.51	4.48E-04	YNL274C	Glyoxylate reductase; null mutation results in increased biomass after diauxic shift; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies; protein abundance increases in response to DNA replication stress
GPH1_1	2.92	3.91	9.07E-05	YPR160W	Glycogen phosphorylase required for the mobilization of glycogen; non-essential; regulated by cyclic AMP-mediated phosphorylation; expression is regulated by stress-response elements and by the HOG MAP kinase pathway
HIS3_1	64.34	41.30	0	YOR202W	Imidazoleglycerol-phosphate dehydratase; catalyzes the sixth step in histidine biosynthesis; mutations cause histidine auxotrophy and sensitivity to Cu, Co, and Ni salts; transcription is regulated by general amino acid control via Gcn4p
HSP104_1	2.01	3.72	1.99E-04	YLL026W	Disaggregase; heat shock protein that cooperates with Ydj1p (Hsp40) and Ssa1p (Hsp70) to refold and reactivate previously denatured, aggregated proteins; responsive to stresses including: heat, ethanol, and sodium arsenite; involved in [PSI+] propagation; protein becomes more abundant and forms cytoplasmic foci in response to DNA replication stress; potentiated Hsp104p variants decrease TDP-43 proteotoxicity by eliminating its cytoplasmic aggregation
HSP150_1	2.17	5.37	7.75E-08	YJL159W	O-mannosylated heat shock protein; secreted and covalently attached to the cell wall via beta-1,3-glucan and disulfide bridges; required for cell wall stability; induced by heat shock, oxidative stress, and nitrogen limitation; HSP150 has a paralog, PIR3, that arose from the whole genome duplication
HSP78_1	2.15	4.20	2.63E-05	YDR258C	Oligomeric mitochondrial matrix chaperone; cooperates with Ssc1p in mitochondrial thermotolerance after heat shock; able to prevent the aggregation of misfolded proteins as well as resolubilize protein aggregates

HSP82_1	2.49	5.66	1.56E-08	YPL240C	Hsp90 chaperone; redundant in function with Hsc82p; required for pheromone signaling, negative regulation of Hsf1p; docks with Tom70p for mitochondrial preprotein delivery; promotes telomerase DNA binding, nucleotide addition; protein abundance increases in response to DNA replication stress; contains two acid-rich unstructured regions that promote solubility of chaperone-substrate complexes; HSP82 has a paralog, HSC82, that arose from the whole genome duplication
HXK1_1	3.44	7.19	6.49E-13	YFR053C	Hexokinase isoenzyme 1; a cytosolic protein that catalyzes phosphorylation of glucose during glucose metabolism; expression is highest during growth on non-glucose carbon sources; glucose-induced repression involves hexokinase Hxk2p; HXK1 has a paralog, HXK2, that arose from the whole genome duplication
HXT4_1	52.01	4.29	1.77E-05	YHR092C	High-affinity glucose transporter; member of the major facilitator superfamily, expression is induced by low levels of glucose and repressed by high levels of glucose; HXT4 has a paralog, HXT7, that arose from the whole genome duplication
HXT6_1	2.24	3.96	7.42E-05		
INA1_1	3.09	5.94	2.79E-09	YLR413W	Putative protein of unknown function; not an essential gene; YLR413W has a paralog, FAT3, that arose from the whole genome duplication
LEE1_1	2.01	3.52	4.29E-04	YPL054W	Zinc-finger protein of unknown function
LEU2_1	958.16	5.61	2.00E-08	YCL018W	Beta-isopropylmalate dehydrogenase (IMDH); catalyzes the third step in the leucine biosynthesis pathway; can additionally catalyze the conversion of β-ethylmalate into α-ketovalerate
MET10_1	2.56	6.18	6.42E-10	YFR030W	Subunit alpha of assimilatory sulfite reductase; complex converts sulfite into sulfide
MET14_1	3.63	5.02	5.23E-07	YKL001C	Adenylylsulfate kinase; required for sulfate assimilation and involved in methionine metabolism

MET16_1	2.29	4.27	2.00E-05	YPR167C	3'-phosphoadenylsulfate reductase; reduces 3'-phosphoadenylyl sulfate to adenosine-3',5'-bisphosphate and free sulfite using reduced thioredoxin as cosubstrate, involved in sulfate assimilation and methionine metabolism
MET1_1	2.02	4.04	5.29E-05	YKR069W	S-adenosyl-L-methionine uroporphyrinogen III transmethylase; involved in the biosynthesis of siroheme, a prosthetic group used by sulfite reductase; required for sulfate assimilation and methionine biosynthesis
MET2_1	2.43	4.29	1.77E-05	YNL277W	L-homoserine-O-acetyltransferase; catalyzes the conversion of homoserine to O-acetyl homoserine which is the first step of the methionine biosynthetic pathway
MET32_1	2.06	4.18	2.91E-05	YDR253C	Zinc-finger DNA-binding transcription factor; involved in transcriptional regulation of the methionine biosynthetic genes; targets strong transcriptional activator Met4p to promoters of sulfur metabolic genes; feedforward loop exists in the regulation of genes controlled by Met4p and Met32p; lack of such a loop for MET31 may account for the differential actions of Met32p and Met31p; MET32 has a paralog, MET31, that arose from the whole genome duplication
MET3_1	2.92	4.13	3.68E-05	YJR010W	ATP sulfurylase; catalyzes the primary step of intracellular sulfate activation, essential for assimilatory reduction of sulfate to sulfide, involved in methionine metabolism
MET5_1	3.27	5.90	3.59E-09	YJR137C	Sulfite reductase beta subunit; involved in amino acid biosynthesis, transcription repressed by methionine
MET6_1	2.94	4.96	7.01E-07	YER091C	Cobalamin-independent methionine synthase; involved in methionine biosynthesis and regeneration; requires a minimum of two glutamates on the methyltetrahydrofolate substrate, similar to bacterial metE homologs
MET8_1	2.08	3.87	1.07E-04	YBR213W	Bifunctional dehydrogenase and ferrochelatase; involved in the biosynthesis of siroheme, a prosthetic group used by sulfite reductase; required for sulfate assimilation and methionine biosynthesis

MF(ALPHA)1_1	2.83	6.86	7.09E-12	YPL187W	Mating pheromone alpha-factor, made by alpha cells; interacts with mating type a cells to induce cell cycle arrest and other responses leading to mating; also encoded by MF(ALPHA)2, although MF(ALPHA)1 produces most alpha-factor; MF(ALPHA)1 has a paralog, MF(ALPHA)2, that arose from the whole genome duplication
MFA1_1	18.62	4.53	5.90E-06	YDR461W	Mating pheromone a-factor; made by a cells; interacts with alpha cells to induce cell cycle arrest and other responses leading to mating; biogenesis involves C-terminal modification, N-terminal proteolysis, and export; also encoded by MFA2
MFA2_1	4.76	4.27	1.99E-05	YNL145W	Mating pheromone a-factor; made by a cells; interacts with alpha cells to induce cell cycle arrest and other responses leading to mating; biogenesis involves C-terminal modification, N-terminal proteolysis, and export; also encoded by MFA1
MHF1_1	11.83	19.61	0	YOL086W-A	Component of the heterotetrameric MHF histone-fold complex; in humans the MMF complex interacts with both DNA and Mph1p ortholog FANCM, a Fanconi anemia complementation group protein, to stabilize and remodel blocked replication forks and repair damaged DNA; mhf1 srs2 double mutants are MMS hypersensitive; ortholog of human centromere constitutive-associated network (CCAN) subunit CENP-S, also known as MHF1
MHT1_1	3.35	3.64	2.78E-04	YLL062C	S-methylmethionine-homocysteine methyltransferase; functions along with Sam4p in the conversion of S-adenosylmethionine (AdoMet) to methionine to control the methionine/AdoMet ratio
MMP1_1	3.14	3.58	3.45E-04	YLL061W	High-affinity S-methylmethionine permease; required for utilization of S-methylmethionine as a sulfur source; has similarity to S-adenosylmethionine permease Sam3p

MPE1_1	2.49	5.71	1.12E-08	YKL059C	Essential conserved subunit of CPF cleavage and polyadenylation factor; plays a role in 3' end formation of mRNA via the specific cleavage and polyadenylation of pre-mRNA, contains a putative RNA-binding zinc knuckle motif; relocalizes to the cytosol in response to hypoxia
MSI1_1	3.70	9.57	0	YBR195C	Subunit of chromatin assembly factor I (CAF-1); chromatin assembly by CAF-1 affects multiple processes including silencing at telomeres, mating type loci, and rDNA; maintenance of kinetochore structure; deactivation of DNA damage checkpoint after DNA repair; chromatin dynamics during transcription; and repression of divergent noncoding transcription; Msi1p localizes to nucleus and cytoplasm and independently regulates the RAS/cAMP pathway via sequestration of Npr1p kinase
MXR1_1	2.03	3.85	1.19E-04	YER042W	Methionine-S-sulfoxide reductase; involved in the response to oxidative stress; protects iron-sulfur clusters from oxidative inactivation along with MXR2; involved in the regulation of lifespan; reduced activity of human homolog implicated in Alzheimer disease
PDC6_1	4.76	5.51	3.51E-08	YGR087C	Minor isoform of pyruvate decarboxylase; decarboxylates pyruvate to acetaldehyde, involved in amino acid catabolism; transcription is glucose- and ethanol-dependent, and is strongly induced during sulfur limitation
PHO11_1	6.26	5.99	2.15E-09	YAR071W	One of three repressible acid phosphatases; glycoprotein that is transported to the cell surface by the secretory pathway; induced by phosphate starvation and coordinately regulated by PHO4 and PHO2; PHO11 has a paralog, PHO12, that arose from a segmental duplication
PHO12_1	9.74	5.19	2.06E-07		

PHO5_1	15.27	4.48	7.53E-06	YBR093C	Repressible acid phosphatase; 1 of 3 repressible acid phosphatases that also mediates extracellular nucleotide-derived phosphate hydrolysis; secretory pathway derived cell surface glycoprotein; induced by phosphate starvation and coordinately regulated by PHO4 and PHO2
PHO81_1	2.73	6.47	1.00E-10	YGR233C	Cyclin-dependent kinase (CDK) inhibitor; regulates Pho80p-Pho85p and Pcl7p-Pho85p cyclin-CDK complexes in response to phosphate levels; inhibitory activity for Pho80p-Pho85p requires myo-D-inositol heptakisphosphate (IP7) generated by Vip1p; relative distribution to the nucleus increases upon DNA replication stress
PHO84_1	22.43	6.44	1.19E-10	YML123C	High-affinity inorganic phosphate (Pi) transporter; also low-affinity manganese transporter; regulated by Pho4p and Spt7p; mutation confers resistance to arsenate; exit from the ER during maturation requires Pho86p; cells overexpressing Pho84p accumulate heavy metals but do not develop symptoms of metal toxicity
PHO89_1	18.06	9.76	0	YBR296C	Plasma membrane Na+/Pi cotransporter; active in early growth phase; similar to phosphate transporters of Neurospora crassa; transcription regulated by inorganic phosphate concentrations and Pho4p; mutations in related human transporter genes hPit1 and hPit2 are associated with hyperphosphatemia-induced calcification of vascular tissue and familial idiopathic basal ganglia calcification
PRM5_1	2.27	8.93	0	YIL117C	Pheromone-regulated protein, predicted to have 1 transmembrane segment; induced during cell integrity signaling; PRM5 has a paralog, YNL058C, that arose from the whole genome duplication

PRM6_1	9.15	3.76	1.71E-04	YML047C	Potassium transporter that mediates K+ influx; activates high-affinity Ca2+ influx system (HACS) during mating pheromone response; expression up-regulated in response to alpha factor; regulated by Ste12p during mating; localized to sites of polarized growth; member of a fungal-specific gene family; PRM6 has a paralog, KCH1, that arose from the whole genome duplication
RPS14B_1	3.91	4.17	3.02E-05	YJL191W	Protein component of the small (40S) ribosomal subunit; required for ribosome assembly and 20S pre-rRNA processing; mutations confer cryptopleurine resistance; homologous to mammalian ribosomal protein S14 and bacterial S11; RPS14B has a paralog, RPS14A, that arose from the whole genome duplication
SAG1_1	6.69	3.49	4.85E-04	YJR004C	Alpha-agglutinin of alpha-cells; binds to Aga1p during agglutination, N-terminal half is homologous to the immunoglobulin superfamily and contains binding site for a-agglutinin, C-terminal half is highly glycosylated and contains GPI anchor
SEO1_1	3.50	5.94	2.88E-09	YAL067C	Putative permease; member of the allantoate transporter subfamily of the major facilitator superfamily; mutation confers resistance to ethionine sulfoxide
SER33_1	2.33	14.96	0	YIL074C	3-phosphoglycerate dehydrogenase; catalyzes the first step in serine and glycine biosynthesis; SER33 has a paralog, SER3, that arose from the whole genome duplication
SIT1_1	5.99	7.87	3.33E-15	YEL065W	Ferrioxamine B transporter; member of the ARN family of transporters that specifically recognize siderophore-iron chelates; transcription is induced during iron deprivation and diauxic shift; potentially phosphorylated by Cdc28p

SPL2_1	66.24	7.55	4.26E-14	YHR136C	Protein with similarity to cyclin-dependent kinase inhibitors; downregulates low-affinity phosphate transport during phosphate limitation by targeting Pho87p to the vacuole; upstream region harbors putative hypoxia response element (HRE) cluster; overproduction suppresses a plc1 null mutation; promoter shows an increase in Snf2p occupancy after heat shock; GFP-fusion protein localizes to the cytoplasm
STE2_1	4.50	6.31	2.73E-10	YFL026W	Receptor for alpha-factor pheromone; seven transmembrane-domain GPCR that interacts with both pheromone and a heterotrimeric G protein to initiate the signaling response that leads to mating between haploid a and alpha cells
STI1_1	2.38	8.65	0	YOR027W	Hsp90 cochaperone; interacts with the Ssa group of the cytosolic Hsp70 chaperones and activates Ssa1p ATPase activity; interacts with Hsp90 chaperones and inhibits their ATPase activity; homolog of mammalian Hop
SUL1_1	3.27	4.55	5.31E-06	YBR294W	High affinity sulfate permease of the SulP anion transporter family; sulfate uptake is mediated by specific sulfate transporters Sul1p and Sul2p, which control the concentration of endogenous activated sulfate intermediates
SUL2_1	2.70	3.91	9.14E-05	YLR092W	High affinity sulfate permease; sulfate uptake is mediated by specific sulfate transporters Sul1p and Sul2p, which control the concentration of endogenous activated sulfate intermediates
TDA10_1	2.05	4.63	3.57E-06	YGR205W	ATP-binding protein of unknown function; crystal structure resembles that of E.coli pantothenate kinase and other small kinases; null mutant is sensitive to expression of the top1-T722A allele
TIP1_1	2.17	5.42	5.90E-08	YBR067C	Major cell wall mannoprotein with possible lipase activity; transcription is induced by heat- and cold-shock; member of the Srp1p/Tip1p family of serine-alanine-rich proteins

TMA10_1	5.25	3.62	2.95E-04	YLR327C	Protein of unknown function that associates with ribosomes; protein abundance increases in response to DNA replication stress; TMA10 has a paralog, STF2, that arose from the whole genome duplication
TRA1_1	2.07	4.41	1.05E-05	YHR099W	Subunit of SAGA and NuA4 histone acetyltransferase complexes; interacts with acidic activators (e.g., Gal4p) which leads to transcription activation; similar to human TRRAP, which is a cofactor for c-Myc mediated oncogenic transformation
URA3_1	493.94	14.52	0	YEL021W	Orotidine-5'-phosphate (OMP) decarboxylase; catalyzes the sixth enzymatic step in the de novo biosynthesis of pyrimidines, converting OMP into uridine monophosphate (UMP); converts 5-FOA into 5-fluorouracil, a toxic compound
VMR1_1	2.13	3.43	6.07E-04	YHL035C	Vacuolar membrane protein; involved in multiple drug resistance and metal sensitivity; ATP-binding cassette (ABC) family member involved in drug transport; potential Cdc28p substrate; induced under respiratory conditions; VMR1 has a paralog, YBT1, that arose from the whole genome duplication
VTC1_1	2.65	6.30	3.00E-10	YER072W	Subunit of the vacuolar transporter chaperone (VTC) complex; VTC complex is involved in membrane trafficking, vacuolar polyphosphate accumulation, microautophagy and non-autophagic vacuolar fusion; also has mRNA binding activity; protein abundance increases in response to DNA replication stress
VTC3_1	3.80	5.38	7.29E-08	YPL019C	Subunit of vacuolar transporter chaperone (VTC) complex; involved in membrane trafficking, vacuolar polyphosphate accumulation, microautophagy and non-autophagic vacuolar fusion; VTC3 has a paralog, VTC2, that arose from the whole genome duplication

VTC4_1	3.26	5.77	7.96E-09	YJL012C	Vacuolar membrane polyphosphate polymerase; subunit of the vacuolar transporter chaperone (VTC) complex involved in synthesis and transfer of polyP to the vacuole; regulates membrane trafficking; role in non-autophagic vacuolar fusion; protein abundance increases in response to DNA replication stress
YAP6_1	2.05	3.59	3.37E-04	YDR259C	Basic leucine zipper (bZIP) transcription factor; physically interacts with the Tup1-Cyc8 complex and recruits Tup1p to its targets; overexpression increases sodium and lithium tolerance; computational analysis suggests a role in regulation of expression of genes involved in carbohydrate metabolism; YAP6 has a paralog, CIN5, that arose from the whole genome duplication
YBR296C-A_1	3.44	5.10	3.31E-07	YBR296C-A	Putative protein of unknown function; identified by gene-trapping, microarray-based expression analysis, and genome-wide homology searching
YFL051C_1	2.71	3.51	4.46E-04	YFL051C	Putative protein of unknown function; YFL051C is not an essential gene
YFL052W_1	2.45	3.36	7.67E-04	YFL052W	Putative zinc cluster protein that contains a DNA binding domain; computational analysis suggests a role as a transcription factor; null mutant is sensitive to Calcofluor White, low osmolarity, and heat, suggesting a role for YFL052Wp in cell wall integrity
YFR052C-A_1	3.31	4.64	3.56E-06		
YLR111W_1	4.39	3.53	4.21E-04		
YLR149C_1	2.13	4.07	4.67E-05	YLR149C	Protein of unknown function; overexpression causes a cell cycle delay or arrest; null mutation results in a decrease in plasma membrane electron transport; YLR149C is not an essential gene; protein abundance increases in response to DNA replication stress
YLR460C_1	2.07	4.84	1.33E-06	YLR460C	Member of the quinone oxidoreductase family; up-regulated in response to the fungicide mancozeb; possibly up-regulated by iodine

YML131W_1	2.39	4.25	2.17E-05	YML131W	Protein of unknown function; similar to medium chain dehydrogenase/reductases; expression induced by stresses including osmotic shock, DNA damaging agents, and other chemicals; GFP-fusion protein localizes to the cytoplasm; protein abundance increases in response to DNA replication stress
YMR052C-A_1	2.40	3.52	4.27E-04		
YOL014W_1	2.23	3.75	1.74E-04	YOL014W	Putative protein of unknown function
YOL118C_1	2.26	5.04	4.74E-07		
YOR203W_1	44.52	31.42	0		
AAH1_1	-3.03	-6.53	6.46E-11	YNL141W	Adenine deaminase (adenine aminohydrolase); converts adenine to hypoxanthine; involved in purine salvage; transcriptionally regulated by nutrient levels and growth phase; Aah1p degraded upon entry into quiescence via SCF and the proteasome
ADH4_1	-12.91	-10.93	8.53E-28	YGL256W	Alcohol dehydrogenase isoenzyme type IV; dimeric enzyme demonstrated to be zinc-dependent despite sequence similarity to iron-activated alcohol dehydrogenases; transcription is induced in response to zinc deficiency
AGP1_1	-4.72	-12.96	2.14E-38	YCL025C	Low-affinity amino acid permease with broad substrate range; involved in uptake of asparagine, glutamine, and other amino acids; expression regulated by SPS plasma membrane amino acid sensor system (Ssy1p-Ptr3p-Ssy5p); AGP1 has a paralog, GNP1, that arose from the whole genome duplication
AIM20_1	-2.10	-6.52	6.92E-11	YIL158W	Putative protein of unknown function; overexpression causes cell cycle delay or arrest; green fluorescent protein (GFP)-fusion protein localizes to vacuole; null mutant displays elevated frequency of mitochondrial genome loss; relocalizes from nucleus to cytoplasm upon DNA replication stress; AIM20 has a paralog, SKG1, that arose from the whole genome duplication
ANS1_1	-3.92	-3.59	3.35E-04	YHR126C	Putative GPI protein; transcription dependent upon Azf1p

ARE1_1	-2.01	-5.95	2.74E-09	YCR048W	Acyl-CoA:sterol acyltransferase; endoplasmic reticulum enzyme that contributes the major sterol esterification activity in the absence of oxygen; ARE1 has a paralog, ARE2, that arose from the whole genome duplication
ARG1_1	-3.10	-4.98	6.51E-07	YOL058W	Arginosuccinate synthetase; catalyzes the formation of L-argininosuccinate from citrulline and L-aspartate in the arginine biosynthesis pathway; potential Cdc28p substrate
ARG3_1	-3.29	-4.91	9.19E-07	YJL088W	Ornithine carbamoyltransferase; also known as carbamoylphosphate:L-ornithine carbamoyltransferase; catalyzes the biosynthesis of the arginine precursor citrulline
ATG15_1	-2.50	-6.38	1.77E-10	YCR068W	Lipase required for intravacuolar lysis of autophagic and Cvt bodies; targeted to intravacuolar vesicles during autophagy via the multivesicular body (MVB) pathway
BAT1_1	-2.40	-9.70	2.88E-22	YHR208W	Mitochondrial branched-chain amino acid (BCAA) aminotransferase; preferentially involved in BCAA biosynthesis; homolog of murine ECA39; highly expressed during logarithmic phase and repressed during stationary phase; BAT1 has a paralog, BAT2, that arose from the whole genome duplication
BIK1_1	-2.07	-6.18	6.33E-10	YCL029C	Microtubule-associated protein; component of the interface between microtubules and kinetochore, involved in sister chromatid separation; essential in polyploid cells but not in haploid or diploid cells; ortholog of mammalian CLIP-170
BIO3_1	-3.11	-3.89	9.99E-05	YNR058W	7,8-diamino-pelargonic acid aminotransferase (DAPA); catalyzes the second step in the biotin biosynthesis pathway; BIO3 is in a cluster of 3 genes (BIO3, BIO4, and BIO5) that mediate biotin synthesis; BIO3 and BIO4 were acquired by horizontal gene transfer (HGT) from bacteria

BIO4_1	-2.67	-3.55	3.85E-04	YNR057C	Dethiobiotin synthetase; catalyzes the third step in the biotin biosynthesis pathway; BIO4 is in a cluster of 3 genes (BIO3, BIO4, and BIO5) that mediate biotin synthesis; BIO3 and BIO4 were acquired by horizontal gene transfer (HGT) from bacteria; expression appears to be repressed at low iron levels
BIO5_1	-2.94	-4.04	5.35E-05	YNR056C	Putative transmembrane protein involved in the biotin biosynthesis; responsible for uptake of 7-keto 8-aminopelargonic acid; BIO5 is in a cluster of 3 genes (BIO3, BIO4, and BIO5) that mediate biotin synthesis
BNA4_1	-2.10	-3.36	7.90E-04	YBL098W	Kynurenine 3-mono oxygenase; required for the de novo biosynthesis of NAD from tryptophan via kynurenine; expression regulated by Hst1p; putative therapeutic target for Huntington disease
BNA5_1	-2.01	-5.18	2.17E-07	YLR231C	Kynureninase; required for the de novo biosynthesis of NAD from tryptophan via kynurenine; expression regulated by Hst1p
BPL1_1	-2.00	-4.81	1.48E-06	YDL141W	Biotin:apoprotein ligase; covalently modifies proteins with the addition of biotin, required for acetyl-CoA carboxylase (Acc1p) holoenzyme formation
BUD23_1	-2.87	-10.67	1.45E-26	YCR047C	Methyltransferase; methylates residue G1575 of 18S rRNA; required for rRNA processing and nuclear export of 40S ribosomal subunits independently of methylation activity; diploid mutant displays random budding pattern
CAR2_1	-2.08	-4.99	5.91E-07	YLR438W	L-ornithine transaminase (OTAse); catalyzes the second step of arginine degradation, expression is dually-regulated by allophanate induction and a specific arginine induction process; not nitrogen catabolite repression sensitive; protein abundance increases in response to DNA replication stress
CIS3_1	-2.33	-7.92	2.45E-15	YJL158C	Mannose-containing glycoprotein constituent of the cell wall; member of the PIR (proteins with internal repeats) family

CIT2_1	-3.08	-6.79	1.14E-11	YCR005C	Citrate synthase; catalyzes the condensation of acetyl coenzyme A and oxaloacetate to form citrate, peroxisomal isozyme involved in glyoxylate cycle; expression is controlled by Rtg1p and Rtg2p transcription factors; CIT2 has a paralog, CIT1, that arose from the whole genome duplication
COS12_1	-38.63	-3.94	8.27E-05	YGL263W	Protein of unknown function; member of the DUP380 subfamily of conserved, often subtelomerically-encoded proteins
CPR4_1	-2.25	-11.00	3.87E-28	YCR069W	Peptidyl-prolyl cis-trans isomerase (cyclophilin); catalyzes the cis-trans isomerization of peptide bonds N-terminal to proline residues; has a potential role in the secretory pathway; CPR4 has a paralog, CPR8, that arose from the whole genome duplication
CTR86_1	-2.10	-3.43	6.00E-04	YCR054C	Essential protein of unknown function; with orthologs in Ashbya gossypii and Candida albicans; similar to human ATXN10, mutations in which cause spinocerebellar ataxia type 10; codon usage corresponds to that observed for yeast genes expressed at low levels; relative distribution to the nucleus increases upon DNA replication stress
CWP2_1	-3.05	-13.38	7.78E-41	YKL096W-A	Covalently linked cell wall mannoprotein; major constituent of the cell wall; plays a role in stabilizing the cell wall; involved in low pH resistance; precursor is GPI-anchored
DBP2_1	-3.00	-7.58	3.34E-14	YNL112W	ATP-dependent RNA helicase of the DEAD-box protein family; has a strong preference for dsRNA; interacts with YRA1; required for the assembly of Yra1p, Nab2p and Mex67p onto mRNA and formation of nuclear mRNP; involved in mRNA decay and rRNA processing; may be involved in suppression of transcription from cryptic initiation sites

DCC1_1	-2.51	-5.32	1.05E-07	YCL016C	Subunit of a complex with Ctf8p and Ctf18p; shares some components with Replication Factor C; required for sister chromatid cohesion and telomere length maintenance
DOG1_1	-2.55	-8.71	3.06E-18	YHR044C	2-deoxyglucose-6-phosphate phosphatase; member of a family of low molecular weight phosphatases; confers 2-deoxyglucose resistance when overexpressed, in vivo substrate has not yet been identified; DOG1 has a paralog, DOG2, that arose from a single-locus duplication
ECM13_1	-2.01	-4.96	7.01E-07	YBL043W	Non-essential protein of unknown function; induced by treatment with 8-methoxypsoralen and UVA irradiation; ECM13 has a paralog, YJR115W, that arose from the whole genome duplication
ELO2_1	-2.24	-5.68	1.37E-08	YCR034W	Fatty acid elongase, involved in sphingolipid biosynthesis; acts on fatty acids of up to 24 carbons in length; mutations have regulatory effects on 1,3-beta-glucan synthase, vacuolar ATPase, and the secretory pathway; FEN1 has a paralog, ELO1, that arose from the whole genome duplication
ERS1_1	-2.87	-5.15	2.64E-07	YCR075C	Protein with similarity to human cystinosin; cystinosin is a H(+)-driven transporter involved in L-cystine export from lysosomes and implicated in the disease cystinosis; contains seven transmembrane domains
FEN2_1	-2.06	-4.07	4.77E-05	YCR028C	Plasma membrane H+-pantothenate symporter; confers sensitivity to the antifungal agent fenpropimorph; relocalizes from vacuole to cytoplasm upon DNA replication stress
FET4_1	-2.21	-6.49	8.62E-11	YMR319C	Low-affinity Fe(II) transporter of the plasma membrane

FKH1_1	-2.16	-3.91	9.14E-05	YIL131C	Forkhead family transcription factor; minor role in expression of G2/M phase genes; negatively regulates transcription elongation; positive role in chromatin silencing at HML, HMR; facilitates clustering and activation of early-firing replication origins; binds to recombination enhancer near HML, regulates donor preference during mating-type switching; relocalizes to cytosol in response to hypoxia; FKH1 has a paralog, FKH2, that arose from the whole genome duplication
FMP48_1	-3.89	-7.74	9.89E-15	YGR052W	Putative protein of unknown function; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies; induced by treatment with 8-methoxypsoralen and UVA irradiation
FUB1_1	-2.16	-6.85	7.41E-12	YCR076C	Putative protein of unknown function; interacts physically with multiple subunits of the 20S proteasome and genetically with genes encoding 20S core particle and 19S regulatory particle subunits; exhibits boundary activity which blocks the propagation of heterochromatic silencing; contains a PI31 proteasome regulator domain and sequence similarity with human PSMF1, a proteasome inhibitor; not an essential gene
FYV5_1	-2.04	-6.48	9.36E-11	YCL058C	Protein involved in regulation of the mating pathway; binds with Matalpha2p to promoters of haploid-specific genes; required for survival upon exposure to K1 killer toxin; involved in ion homeostasis
GAS3_1	-2.28	-5.56	2.77E-08	YMR215W	Putative 1,3-beta-glucanosyltransferase; has similarity go other GAS family members; low abundance, possibly inactive member of the GAS family of GPI-containing proteins; localizes to the cell wall; mRNA induced during sporulation

GBP2_1	-2.05	-8.44	3.06E-17	YCL011C	Poly(A+) RNA-binding protein; key surveillance factor for the selective export of spliced mRNAs from the nucleus to the cytoplasm; preference for intron-containing genes; similar to Npl3p; also binds single-stranded telomeric repeat sequence in vitro; relocalizes to the cytosol in response to hypoxia; GBP2 has a paralog, HRB1, that arose from the whole genome duplication
GFD2_1	-3.99	-5.87	4.49E-09	YCL036W	Protein of unknown function; identified as a high-copy suppressor of a dbp5 mutation; GFD2 has a paralog, YDR514C, that arose from the whole genome duplication
GPP2_1	-2.38	-9.37	7.38E-21	YER062C	DL-glycerol-3-phosphate phosphatase involved in glycerol biosynthesis; also known as glycerol-1-phosphatase; induced in response to hyperosmotic or oxidative stress, and during diauxic shift; GPP2 has a paralog, GPP1, that arose from the whole genome duplication
HCM1_1	-2.41	-3.59	3.36E-04	YCR065W	Forkhead transcription factor; drives S-phase specific expression of genes involved in chromosome segregation, spindle dynamics, and budding; suppressor of calmodulin mutants with specific SPB assembly defects; telomere maintenance role
HOT13_1	-2.31	-5.12	3.12E-07	YKL084W	Zinc-binding mitochondrial intermembrane space (IMS) protein; involved in a disulfide relay system for IMS import of cysteine-containing proteins; binds Mia40p and stimulates its Erv1p-dependent oxidation, probably by sequestering zinc
HSP12_1	-2.55	-3.46	5.46E-04	YFL014W	Plasma membrane protein involved in maintaining membrane organization; involved in maintaining organization during stress conditions; induced by heat shock, oxidative stress, osmostress, stationary phase, glucose depletion, oleate and alcohol; protein abundance increased in response to DNA replication stress and dietary restriction; regulated by the HOG and Ras-Pka pathways; required for dietary restriction-induced lifespan extension

HSP30_1	-2.04	-4.70	2.59E-06	YCR021C	Negative regulator of the H(+)-ATPase Pma1p; stress-responsive protein; hydrophobic plasma membrane localized; induced by heat shock, ethanol treatment, weak organic acid, glucose limitation, and entry into stationary phase
HTB2_1	-2.15	-9.91	3.74E-23	YBL002W	Histone H2B; core histone protein required for chromatin assembly and chromosome function; nearly identical to HTB1; Rad6p-Bre1p-Lge1p mediated ubiquitination regulates reassembly after DNA replication, transcriptional activation, meiotic DSB formation and H3 methylation
HTL1_1	-2.08	-5.82	5.72E-09	YCR020W-B	Component of the RSC chromatin remodeling complex; RSC functions in transcriptional regulation and elongation, chromosome stability, and establishing sister chromatid cohesion; involved in telomere maintenance
HXT11_1	-2.19	-4.31	1.66E-05	YOL156W	Putative hexose transporter that is nearly identical to Hxt9p; has similarity to major facilitator superfamily (MFS) transporters and is involved in pleiotropic drug resistance
IMD4_1	-2.11	-7.10	1.28E-12	YML056C	Inosine monophosphate dehydrogenase; catalyzes the rate-limiting step in the de novo synthesis of GTP; member of a four-gene family in S. cerevisiae, constitutively expressed; IMD4 has a paralog, IMD3, that arose from the whole genome duplication
ISU2_1	-2.14	-5.29	1.25E-07	YOR226C	Protein required for synthesis of iron-sulfur proteins; localized to the mitochondrial matrix; performs a scaffolding function in mitochondria during Fe/S cluster assembly; involved in Fe-S cluster assembly for both mitochondrial and cytosolic proteins; isu1 isu2 double mutant is inviable; protein abundance increases in response to DNA replication stress; evolutionarily conserved; ISU2 has a paralog, ISU1, that arose from the whole genome duplication

KAR2_1	-2.17	-7.64	2.13E-14	YJL034W	ATPase involved in protein import into the ER; also acts as a chaperone to mediate protein folding in the ER and may play a role in ER export of soluble proteins; regulates the unfolded protein response via interaction with Ire1p
LEU1_1	-3.26	-13.48	2.17E-41	YGL009C	Isopropylmalate isomerase; catalyzes the second step in the leucine biosynthesis pathway
LEU9_1	-2.20	-7.93	2.14E-15	YOR108W	Alpha-isopropylmalate synthase II (2-isopropylmalate synthase); catalyzes the first step in the leucine biosynthesis pathway; the minor isozyme, responsible for the residual alpha-IPMS activity detected in a leu4 null mutant; LEU9 has a paralog, LEU4, that arose from the whole genome duplication
LSB5_1	-2.27	-9.37	7.00E-21	YCL034W	Protein of unknown function; binds Las17p, which is a homolog of human Wiskott-Aldrich Syndrome protein involved in actin patch assembly and actin polymerization; may mediate disassembly of the Pan1 complex from the endocytic coat
MAE1_1	-2.81	-8.09	6.19E-16	YKL029C	Mitochondrial malic enzyme; catalyzes the oxidative decarboxylation of malate to pyruvate, which is a key intermediate in sugar metabolism and a precursor for synthesis of several amino acids
MGA1_1	-5.67	-6.92	4.57E-12	YGR249W	Protein similar to heat shock transcription factor; multicopy suppressor of pseudohyphal growth defects of ammonium permease mutants
MGR1_1	-2.39	-5.49	4.08E-08	YCL044C	Subunit of the mitochondrial (mt) i-AAA protease supercomplex; i-AAA degrades misfolded mitochondrial proteins; forms a subcomplex with Mgr3p that binds to substrates to facilitate proteolysis; required for growth of cells lacking mtDNA
MRS3_1	-2.01	-7.70	1.41E-14	YJL133W	Iron transporter, mediates Fe2+ transport across inner mito membrane; mitochondrial carrier family member; active under low-iron conditions; may transport other cations; MRS3 has a paralog, MRS4, that arose from the whole genome duplication

NCA3_1	-2.22	-14.83	9.32E-50	YJL116C	Protein involved in mitochondrion organization; functions with Nca2p to regulate mitochondrial expression of subunits 6 (Atp6p) and 8 (Atp8p) of the Fo-F1 ATP synthase; member of the SUN family; expression induced in cells treated with the mycotoxin patulin; NCA3 has a paralog, UTH1, that arose from the whole genome duplication
NPP1_1	-2.34	-6.67	2.65E-11	YCR026C	Nucleotide pyrophosphatase/phosphodiesterase; mediates extracellular nucleotide phosphate hydrolysis along with Npp2p and Pho5p; activity and expression enhanced during conditions of phosphate starvation; involved in spore wall assembly; NPP1 has a paralog, NPP2, that arose from the whole genome duplication, and an npp1 npp2 double mutant exhibits reduced dityrosine fluorescence relative to the single mutants
NRT1_1	-2.76	-7.34	2.14E-13	YOR071C	High-affinity nicotinamide riboside transporter; also transports thiamine with low affinity; major transporter for 5-aminoimidazole-4-carboxamide-1-beta-D-ribofuranoside (acadesine) uptake; shares sequence similarity with Thi7p and Thi72p; proposed to be involved in 5-fluorocytosine sensitivity
OAC1_1	-2.01	-13.39	6.54E-41	YKL120W	Mitochondrial inner membrane transporter; transports oxaloacetate, sulfate, thiosulfate, and isopropylmalate; member of the mitochondrial carrier family
OPT2_1	-2.68	-5.55	2.89E-08	YPR194C	Oligopeptide transporter; member of the OPT family, with potential orthologs in S. pombe and C. albicans; also plays a role in formation of mature vacuoles

PDI1_1	-3.07	-28.98	1.14E-184	YCL043C	Protein disulfide isomerase; multifunctional protein of ER lumen, essential for formation of disulfide bonds in secretory and cell-surface proteins, unscrambles non-native disulfide bonds; key regulator of Ero1p; forms complex with Mnl1p that has exomannosidase activity, processing unfolded protein-bound Man8GlcNAc2 oligosaccharides to Man7GlcNAc2, promoting degradation in unfolded protein response; PDI1 has a paralog, EUG1, that arose from the whole genome duplication
PET18_1	-3.02	-6.07	1.27E-09	YCR020C	Protein of unknown function; has weak similarity to proteins involved in thiamin metabolism; expression is induced in the absence of thiamin
PGK1_1	-2.95	-16.21	4.06E-59	YCR012W	3-phosphoglycerate kinase; catalyzes transfer of high-energy phosphoryl groups from the acyl phosphate of 1,3-bisphosphoglycerate to ADP to produce ATP; key enzyme in glycolysis and gluconeogenesis
PLB3_1	-2.93	-12.49	7.99E-36	YOL011W	Phospholipase B (lysophospholipase) involved in lipid metabolism; hydrolyzes phosphatidylinositol and phosphatidylserine and displays transacylase activity in vitro; PLB3 has a paralog, PLB1, that arose from the whole genome duplication
PMA1_1	-2.24	-12.40	2.66E-35	YGL008C	Plasma membrane H+-ATPase; pumps protons out of the cell; major regulator of cytoplasmic pH and plasma membrane potential; P2-type ATPase; Hsp30p plays a role in Pma1p regulation; interactions with Std1p appear to propagate [GAR+]
PMA2_1	-5.04	-7.55	4.19E-14	YPL036W	Plasma membrane H+-ATPase; isoform of Pma1p, involved in pumping protons out of the cell; regulator of cytoplasmic pH and plasma membrane potential
PMP1_1	-2.53	-21.16	2.18E-99	YCR024C-A	Regulatory subunit for the plasma membrane H(+)-ATPase Pma1p; small single-membrane span proteolipid; forms unique helix and positively charged cytoplasmic domain that is able to specifically segregate phosphatidylserines; PMP1 has a paralog, PMP2, that arose from the whole genome duplication

POL4_1	-2.16	-3.63	2.79E-04	YCR014C	DNA polymerase IV; undergoes pair-wise interactions with Dnl4p-Lif1p and Rad27p to mediate repair of DNA double-strand breaks by non-homologous end joining (NHEJ); homologous to mammalian DNA polymerase beta
PSA1_1	-2.44	-15.41	1.32E-53	YDL055C	GDP-mannose pyrophosphorylase (mannose-1-phosphate guanyltransferase); synthesizes GDP-mannose from GTP and mannose-1-phosphate in cell wall biosynthesis; required for normal cell wall structure
PWP2_1	-2.48	-6.68	2.35E-11	YCR057C	Conserved 90S pre-ribosomal component; essential for proper endonucleolytic cleavage of the 35 S rRNA precursor at A0, A1, and A2 sites; contains eight WD-repeats; PWP2 deletion leads to defects in cell cycle and bud morphogenesis
QDR2_1	-2.00	-9.25	2.25E-20	YIL121W	Plasma membrane transporter of the major facilitator superfamily; member of the 12-spanner drug:H(+) antiporter DHA1 family; exports copper; has broad substrate specificity and can transport many mono- and divalent cations; transports a variety of drugs and is required for resistance to quinidine, barban, cisplatin, and bleomycin; contributes to potassium homeostasis; expression is regulated by copper
RER1_1	-2.48	-9.27	1.94E-20	YCL001W	Protein involved in retention of membrane proteins; including Sec12p, in the ER; localized to Golgi; functions as a retrieval receptor in returning membrane proteins to the ER
RGS2_1	-2.94	-6.78	1.17E-11	YOR107W	Negative regulator of glucose-induced cAMP signaling; directly activates the GTPase activity of the heterotrimeric G protein alpha subunit Gpa2p
RHB1_1	-2.18	-4.47	7.78E-06	YCR027C	Putative Rheb-related GTPase; involved in regulating canavanine resistance and arginine uptake; member of the Ras superfamily of G-proteins
RIM1_1	-2.40	-12.04	2.32E-33	YCR028C-A	ssDNA-binding protein essential for mitochondrial genome maintenance; involved in mitochondrial DNA replication

RPS14A_1	-2.01	-13.04	6.81E-39	YCR031C	Protein component of the small (40S) ribosomal subunit; required for ribosome assembly and 20S pre-rRNA processing; mutations confer cryptopleurine resistance; homologous to mammalian ribosomal protein S14 and bacterial S11; RPS14A has a paralog, RPS14B, that arose from the whole genome duplication
RRM3_1	-2.08	-4.03	5.69E-05	YHR031C	DNA helicase involved in rDNA replication and Ty1 transposition; binds to and suppresses DNA damage at G4 motifs in vivo; relieves replication fork pauses at telomeric regions; structurally and functionally related to Pif1p
RRP43_1	-2.36	-8.98	2.71E-19	YCR035C	Exosome non-catalytic core component; involved in 3'-5' RNA processing and degradation in both the nucleus and the cytoplasm; has similarity to E. coli RNase PH and to human hRrp43p (OIP2, EXOSC8); protein abundance increases in response to DNA replication stress
RRT12_1	-2.28	-4.04	5.38E-05	YCR045C	Probable subtilisin-family protease; role in formation of the dityrosine layer of spore walls; localizes to the spore wall and also the nuclear envelope and ER region in mature spores
RSA4_1	-2.66	-6.32	2.58E-10	YCR072C	WD-repeat protein involved in ribosome biogenesis; may interact with ribosomes; required for maturation and efficient intra-nuclear transport or pre-60S ribosomal subunits, localizes to the nucleolus
RSC6_1	-2.01	-6.28	3.36E-10	YCR052W	Component of the RSC chromatin remodeling complex; essential for mitotic growth; RSC6 has a paralog, SNF12, that arose from the whole genome duplication
RSF2_1	-2.31	-5.28	1.31E-07	YJR127C	Zinc-finger protein; involved in transcriptional control of both nuclear and mitochondrial genes, many of which specify products required for glycerol-based growth, respiration, and other functions; RSF2 has a paralog, TDA9, that arose from the whole genome duplication; relocalizes from nucleus to cytoplasm upon DNA replication stress

RSN1_1	-2.16	-4.88	1.07E-06	YMR266W	Membrane protein of unknown function; overexpression suppresses NaCl sensitivity of sro7 mutant cells by restoring sodium pump (Ena1p) localization to the plasma membrane
RTC4_1	-3.73	-12.20	3.04E-34	YNL254C	Protein of unknown function; null mutation suppresses cdc13-1 temperature sensitivity; (GFP)-fusion protein localizes to both the cytoplasm and the nucleus
SAT4_1	-3.35	-9.18	4.17E-20	YCR008W	Ser/Thr protein kinase involved in salt tolerance; funtions in regulation of Trk1p-Trk2p potassium transporter; partially redundant with Hal5p; has similarity to Npr1p
SCS3_1	-2.05	-7.71	1.22E-14	YGL126W	Protein required for inositol prototrophy; required for normal ER membrane biosynthesis; ortholog of the FIT family of proteins involved in triglyceride droplet biosynthesis and homologous to human FIT2; disputed role in the synthesis of inositol phospholipids from inositol
SFG1_1	-2.32	-4.74	2.09E-06	YOR315W	Nuclear protein putative transcription factor; required for growth of superficial pseudohyphae (which do not invade the agar substrate) but not for invasive pseudohyphal growth; may act together with Phd1p; potential Cdc28p substrate
SGF29_1	-2.51	-8.70	3.22E-18	YCL010C	Component of the HAT/Core module of the SAGA, SLIK, and ADA complexes; HAT/Core module also contains Gcn5p, Ngg1p, and Ada2p; binds methylated histone H3K4; involved in transcriptional regulation through SAGA and TBP recruitment to target promoters and H3 acetylation
SIL1_1	-2.21	-8.84	9.60E-19	YOL031C	Nucleotide exchange factor for the ER lumenal Hsp70 chaperone Kar2p; required for protein translocation into the endoplasmic reticulum (ER); homolog of Yarrowia lipolytica SLS1; GrpE-like protein
SIP18_1	-2.38	-4.72	2.35E-06	YMR175W	Phospholipid-binding hydrophilin; essential to overcome desiccation-rehydration process; expression is induced by osmotic stress; SIP18 has a paralog, GRE1, that arose from the whole genome duplication

SOL2_1	-2.19	-7.74	1.02E-14	YCR073W-A	Protein with a possible role in tRNA export; shows similarity to 6-phosphogluconolactonase non-catalytic domains but does not exhibit this enzymatic activity; homologous to Sol3p and Sol4p; SOL2 has a paralog, SOL1, that arose from the whole genome duplication
SPB1_1	-2.24	-6.22	4.96E-10	YCL054W	AdoMet-dependent methyltransferase; involved in rRNA processing and 60S ribosomal subunit maturation; methylates G2922 in the tRNA docking site of the large subunit rRNA and in the absence of snR52, U2921; suppressor of PAB1 mutants
SPI1_1	-2.45	-5.53	3.27E-08	YER150W	GPI-anchored cell wall protein involved in weak acid resistance; basal expression requires Msn2p/Msn4p; expression is induced under conditions of stress and during the diauxic shift; SPI1 has a paralog, SED1, that arose from the whole genome duplication
SPS100_1	-2.24	-3.67	2.45E-04	YHR139C	Protein required for spore wall maturation; expressed during sporulation; may be a component of the spore wall; expression also induced in cells treated with the mycotoxin patulin; SPS100 has a paralog, YGP1, that arose from the whole genome duplication
SRD1_1	-3.09	-11.00	3.63E-28	YCR018C	Protein involved in the processing of pre-rRNA to mature rRNA; contains a C2/C2 zinc finger motif; srd1 mutation suppresses defects caused by the rrp1-1 mutation
SSK22_1	-2.39	-3.74	1.81E-04	YCR073C	MAP kinase kinase kinase of HOG1 mitogen-activated signaling pathway; functionally redundant with Ssk2p; interacts with and is activated by Ssk1p; phosphorylates Pbs2p; SSK22 has a paralog, SSK2, that arose from the whole genome duplication
STE50_1	-2.01	-5.89	3.81E-09	YCL032W	Adaptor protein for various signaling pathways; involved in mating response, invasive/filamentous growth, osmotolerance; acts as an adaptor that links G protein-associated Cdc42p-Ste20p complex to the effector Ste11p to modulate signal transduction

SUR2_1	-2.69	-14.02	1.16E-44	YDR297W	Sphinganine C4-hydroxylase; catalyses the conversion of sphinganine to phytosphingosine in sphingolipid biosyntheis
SUT2_1	-2.11	-4.68	2.86E-06	YPR009W	Putative transcription factor of the Zn2Cys6 family; regulates sterol uptake under anaerobic conditions along with SUT1; multicopy suppressor of mutations that cause low activity of the cAMP/protein kinase A pathway; positively regulates mating along with SUT1 by repressing the expression of genes (PRR2, NCE102 and RHO5) which function as mating inhibitors; SUT2 has a paralog, SUT1, that arose from the whole genome duplication
TAT1_1	-2.62	-12.06	1.71E-33	YBR069C	Amino acid transporter for valine, leucine, isoleucine, and tyrosine; low-affinity tryptophan and histidine transporter; overexpression confers FK506 and FTY720 resistance; protein abundance increases in response to DNA replication stress
THI7_1	-2.64	-4.04	5.24E-05	YLR237W	Plasma membrane transporter responsible for the uptake of thiamine; contributes to uptake of 5-aminoimidazole-4-carboxamide-1-beta-D-ribofuranoside (acadesine); member of the major facilitator superfamily of transporters; mutation of human ortholog causes thiamine-responsive megaloblastic anemia
TOS3_1	-2.05	-5.46	4.86E-08	YGL179C	Protein kinase; related to and functionally redundant with Elm1p and Sak1p for the phosphorylation and activation of Snf1p; functionally orthologous to LKB1, a mammalian kinase associated with Peutz-Jeghers cancer-susceptibility syndrome; TOS3 has a paralog, SAK1, that arose from the whole genome duplication
TRM11_1	-2.10	-5.41	6.43E-08	YOL124C	Catalytic subunit of adoMet-dependent tRNA methyltransferase complex; required for the methylation of the guanosine nucleotide at position 10 (m2G10) in tRNAs; contains a THUMP domain and a methyltransferase domain; another complex member is Trm112p

TUP1_1	-2.23	-7.89	2.97E-15	YCR084C	General repressor of transcription; forms complex with Cyc8p, involved in the establishment of repressive chromatin structure through interactions with histones H3 and H4, appears to enhance expression of some genes
ULI1_1	-4.37	-7.35	1.98E-13	YFR026C	Putative protein of unknown function; involved in and induced by the endoplasmic reticulum unfolded protein response (UPR)
URA1_1	-2.48	-22.26	8.91E-110	YKL216W	Dihydroorotate dehydrogenase; catalyzes the fourth enzymatic step in the de novo biosynthesis of pyrimidines, converting dihydroorotic acid into orotic acid
VAC17_1	-2.55	-3.85	1.17E-04	YCL063W	Phosphoprotein involved in vacuole inheritance; degraded in late M phase of the cell cycle; acts as a vacuole-specific receptor for myosin Myo2p
VEL1_1	-24.69	-4.43	9.48E-06	YGL258W	Protein of unknown function; highly induced in zinc-depleted conditions and has increased expression in NAP1 deletion mutants; VEL1 has a paralog, YOR387C, that arose from a single-locus duplication
VMA9_1	-2.22	-14.88	4.29E-50	YCL005W-A	Vacuolar H+ ATPase subunit e of the V-ATPase V0 subcomplex; essential for vacuolar acidification; interacts with the V-ATPase assembly factor Vma21p in the ER; involved in V0 biogenesis
WSC4_1	-2.07	-4.80	1.61E-06	YHL028W	Endoplasmic reticulum (ER) membrane protein; involved in the translocation of soluble secretory proteins and insertion of membrane proteins into the ER membrane; may also have a role in the stress response but has only partial functional overlap with WSC1-3
YBL100W-B_1	-2.51	-6.29	3.09E-10		
YBR200W-A_1	-3.71	-5.26	1.44E-07	YBR200W-A	Putative protein of unknown function; identified by fungal homology and RT-PCR
YCL012C_1	-2.40	-5.52	3.46E-08	YCL012C	Putative protein of unknown function; orthologs are present in S. bayanus, S. paradoxus and Ashbya gossypii; YCL012C is not an essential gene
YCL019W_1	-2.25	-5.06	4.19E-07		
YCL021W-A_1	-2.87	-4.00	6.29E-05	YCL021W-A	Putative protein of unknown function

YCL041C_1	-2.90	-6.81	9.48E-12		
YCL048W-A_1	-2.98	-7.15	8.85E-13		
YCR001W_1	-2.80	-3.47	5.26E-04		
YCR007C_1	-2.29	-4.71	2.42E-06	YCR007C	Putative integral membrane protein; member of DUP240 gene family; YCR007C is not an essential gene
YCR013C_1	-3.03	-15.28	1.09E-52		
YCR016W_1	-2.35	-7.83	4.99E-15	YCR016W	Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the nucleolus and nucleus; predicted to be involved in ribosome biogenesis
YCR023C_1	-2.17	-5.33	1.00E-07	YCR023C	Vacuolar membrane protein of unknown function; member of the multidrug resistance family; YCR023C is not an essential gene
YCR024C-B_1	-2.43	-23.58	5.86E-123	YCR024C-B	Putative protein of unknown function; identified by expression profiling and mass spectrometry
YCR061W_1	-2.02	-4.64	3.56E-06	YCR061W	Protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm in a punctate pattern; induced by treatment with 8-methoxypsoralen and UVA irradiation
YCR075W-A_1	-2.16	-5.01	5.43E-07	YCR075W-A	Putative protein of unknown function; identified by homology to Ashbya gossypii; YCR075W-A has a paralog, YNR034W-A, that arose from the whole genome duplication
YCR090C_1	-2.04	-5.86	4.59E-09	YCR090C	Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm and nucleus; YCR090C is not an essential gene
YDR034C-D_1	-2.05	-4.36	1.30E-05		
YDR210W-B_1	-3.38	-4.39	1.14E-05		
YDR261W-B_1	-2.44	-5.23	1.65E-07	YDR261W-B	Retrotransposon TYA Gag and TYB Pol genes; transcribed/translated as one unit; polyprotein is processed to make a nucleocapsid-like protein (Gag), reverse transcriptase (RT), protease (PR), and integrase (IN); similar to retroviral genes similar to retroviral genes

YDR365W-B_1	-4.19	-8.05	8.50E-16		
YER152C_1	-2.11	-6.84	8.19E-12	YER152C	Protein with 2-aminoadipate transaminase activity; shares amino acid similarity with the aminotransferases Aro8p and Aro9p; YER152C is not an essential gene
YER188W_1	-2.38	-4.66	3.23E-06		
YFL012W_1	-2.34	-3.35	8.02E-04	YFL012W	Putative protein of unknown function; transcribed during sporulation; null mutant exhibits increased resistance to rapamycin
YFR020W_1	-6.61	-18.33	4.72E-75		
YGK3_1	-2.26	-3.64	2.74E-04	YOL128C	Protein kinase related to mammalian GSK-3 glycogen synthase kinases; GSK-3 homologs (Mck1p, Rim11p, Mrk1p, Ygk3p) are involved in control of Msn2p-dependent transcription of stress responsive genes and in protein degradation; YGK3 has a paralog, MCK1, that arose from the whole genome duplication
YGL081W_1	-2.05	-3.57	3.53E-04	YGL081W	Putative protein of unknown function; non-essential gene; interacts genetically with CHS5, a gene involved in chitin biosynthesis
YGL262W_1	-22.99	-4.37	1.25E-05	YGL262W	Putative protein of unknown function; null mutant displays elevated sensitivity to expression of a mutant huntingtin fragment or of alpha-synuclein; YGL262W is not an essential gene
YGP1_1	-2.28	-7.62	2.61E-14	YNL160W	Cell wall-related secretory glycoprotein; induced by nutrient deprivation-associated growth arrest and upon entry into stationary phase; may be involved in adaptation prior to stationary phase entry; YGP1 has a paralog, SPS100, that arose from the whole genome duplication
YGR079W_1	-2.60	-6.86	6.86E-12	YGR079W	Putative protein of unknown function; YGR079W is not an essential gene
YIH1_1	-2.02	-7.22	5.16E-13	YCR059C	Negative regulator of eIF2 kinase Gcn2p; competes with Gcn2p for binding to Gcn1p; may contribute to regulation of translation in response to starvation via regulation of Gcn2p; binds to monomeric actin and to ribosomes and polyribosomes; ortholog of mammalian IMPACT

YIR042C_1	-6.67	-7.39	1.52E-13	YIR042C	Putative protein of unknown function; YIR042C is a non-essential gene
YJR115W_1	-7.41	-9.17	4.55E-20	YJR115W	Putative protein of unknown function; YJR115W has a paralog, ECM13, that arose from the whole genome duplication
YKL030W_1	-2.54	-6.29	3.08E-10		
YKL031W_1	-5.18	-4.39	1.12E-05		
YLR159C-A_1	-2.10	-4.38	1.18E-05		
YLR349W_1	-3.23	-3.73	1.89E-04		
YML122C_1	-5.44	-6.24	4.30E-10		
YMR244W_1	-4.09	-4.91	8.88E-07	YMR244W	Putative protein of unknown function
YMR265C_1	-2.03	-4.29	1.82E-05	YMR265C	Putative protein of unknown function
YMR320W_1	-2.45	-6.85	7.21E-12		
YNL234W_1	-2.33	-4.54	5.53E-06	YNL234W	Protein of unknown function with similarity to globins; has a functional heme-binding domain; mutant has aneuploidy tolerance; transcription induced by stress conditions; may be involved in glucose signaling or metabolism; regulated by Rgt1
YOL019W_1	-2.16	-7.97	1.63E-15	YOL019W	Protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the cell periphery and vacuole; YOL019W has a paralog, DCV1, that arose from the whole genome duplication
YOR225W_1	-2.00	-4.86	1.20E-06		
YOR387C_1	-6.91	-3.35	8.10E-04	YOR387C	Putative protein of unknown function; regulated by the metal-responsive Aft1p transcription factor; highly inducible in zinc-depleted conditions; localizes to the soluble fraction; YOR387C has a paralog, VEL1, that arose from a single-locus duplication
YPL014W_1	-3.80	-15.03	4.35E-51	YPL014W	Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm and to the nucleus
YPR158C-D_1	-2.01	-4.89	9.99E-07		

ZAP1_1	-2.36	-8.80	1.38E-18	YJL056C	Zinc-regulated transcription factor; binds to zinc-responsive promoters to induce transcription of certain genes in presence of zinc, represses other genes in low zinc; regulates its own transcription; contains seven zinc-finger domains
ZPS1_1	-14.26	-9.60	7.70E-22	YOL154W	Putative GPI-anchored protein; transcription is induced under low-zinc conditions, as mediated by the Zap1p transcription factor, and at alkaline pH
ZRT1_1	-3.12	-6.24	4.31E-10	YGL255W	High-affinity zinc transporter of the plasma membrane; responsible for the majority of zinc uptake; transcription is induced under low-zinc conditions by the Zap1p transcription factor

C. BY4741*adh1*Δ_#68-69-70 and BY4741*adh1*Δ#800-1454-903

Feature ID	Experiment - Fold Change (normalized values)	Baggerley's test: Host_Pathway vs Host_EmptyVector normalized values - Test statistic	Baggerley's test: Host_Pathway vs Host_EmptyVector normalized values - P-value	Annotations - Ensembl	Annotations - Gene title
ADH5_1	2.32	8.00	1.33E-15	YBR145W	Alcohol dehydrogenase isoenzyme V; involved in ethanol production; ADH5 has a paralog, ADH1, that arose from the whole genome duplication
ADH6_1	2.35	14.55	0	YMR318C	NADPH-dependent medium chain alcohol dehydrogenase; has broad substrate specificity; member of the cinnamyl family of alcohol dehydrogenases; may be involved in fusel alcohol synthesis or in aldehyde tolerance; protein abundance increases in response to DNA replication stress
ALD5_1	3.31	3.33	8.83E-04	YER073W	Mitochondrial aldehyde dehydrogenase; involved in regulation or biosynthesis of electron transport chain components and acetate formation; activated by K+; utilizes NADP+ as the preferred coenzyme; constitutively expressed
ARG1_1	2.11	9.61	0	YOL058W	Arginosuccinate synthetase; catalyzes the formation of L-argininosuccinate from citrulline and L-aspartate in the arginine biosynthesis pathway; potential Cdc28p substrate
ARG2_1	2.00	4.50	6.92E-06	YJL071W	Acetylglutamate synthase (glutamate N-acetyltransferase); mitochondrial enzyme that catalyzes the first step in the biosynthesis of the arginine precursor ornithine; forms a complex with Arg5,6p
ARG4_1	2.62	4.82	1.41E-06	YHR018C	Argininosuccinate lyase; catalyzes the final step in the arginine biosynthesis pathway

ARG5,6_1	2.06	3.83	1.26E-04	YER069W	Acetylglutamate kinase and N-acetyl-gamma-glutamyl-phosphate reductase; N-acetyl-L-glutamate kinase (NAGK) catalyzes the 2nd and N-acetyl-gamma-glutamyl-phosphate reductase (NAGSA), the 3rd step in arginine biosynthesis; synthesized as a precursor which is processed in the mitochondrion to yield mature NAGK and NAGSA; enzymes form a metabolon complex with Arg2p; NAGK C-terminal domain stabilizes the enzymes, slows catalysis and is involved in feed-back inhibition by arginine
ARO9_1	2.95	4.62	3.90E-06	YHR137W	Aromatic aminotransferase II; catalyzes the first step of tryptophan, phenylalanine, and tyrosine catabolism
BAP2_1	2.07	10.56	0	YBR068C	High-affinity leucine permease; functions as a branched-chain amino acid permease involved in uptake of leucine, isoleucine and valine; contains 12 predicted transmembrane domains; BAP2 has a paralog, BAP3, that arose from the whole genome duplication
BAT1_1	2.13	4.45	8.55E-06	YHR208W	Mitochondrial branched-chain amino acid (BCAA) aminotransferase; preferentially involved in BCAA biosynthesis; homolog of murine ECA39; highly expressed during logarithmic phase and repressed during stationary phase; BAT1 has a paralog, BAT2, that arose from the whole genome duplication
BNA3_1	2.13	4.74	2.10E-06	YJL060W	Kynurenine aminotransferase; catalyzes formation of kynurenic acid from kynurenine; potential Cdc28p substrate

BSC5_1	3.44	4.38	1.19E-05	YNR069C	Protein of unknown function; shows homology with N-terminal end of Bul1p; ORF exhibits genomic organization compatible with a translational readthrough-dependent mode of expression; readthrough expression includes YNR068C and the locus for this readthrough is termed BUL3; Bul3p is involved in ubiquitin-mediated sorting of plasma membrane proteins; readthrough and shortened forms of Bul3p interact with Rsp5p differently in vitro
BTN2_1	3.03	7.94	2.00E-15	YGR142W	v-SNARE binding protein; facilitates specific protein retrieval from a late endosome to the Golgi; modulates arginine uptake, possible role in mediating pH homeostasis between the vacuole and plasma membrane H(+)-ATPase; contributes to prion curing; BTN2 has a paralog, CUR1, that arose from the whole genome duplication
CPA2_1	2.18	8.49	0	YJR109C	Large subunit of carbamoyl phosphate synthetase; carbamoyl phosphate synthetase catalyzes a step in the synthesis of citrulline, an arginine precursor
CUR1_1	4.16	6.39	1.71E-10	YPR158W	Sorting factor, central regulator of spatial protein quality control; physically and functionally interacts with chaperones to promote sorting and deposition of misfolded proteins into cytosolic compartments; involved in destabilization of [URE3] prions; CUR1 has a paralog, BTN2, that arose from the whole genome duplication
ECL1_1	2.05	4.77	1.80E-06	YGR146C	Protein of unknown function; mitochondrial-dependent role in the extension of chronological lifespan; overexpression increases oxygen consumption and respiratory activity while deletion results in reduced oxygen consumption under conditions of caloric restriction; induced by iron homeostasis transcription factor Aft2p; multicopy suppressor of temperature sensitive hsf1 mutant; induced by treatment with 8-methoxypsoralen and UVA irradiation

ENB1_1	2.74	5.30	1.14E-07	YOL158C	Endosomal ferric enterobactin transporter; expressed under conditions of iron deprivation; member of the major facilitator superfamily; expression is regulated by Rcs1p and affected by chloroquine treatment
ERG4_1	2.32	3.43	6.07E-04	YGL012W	C-24(28) sterol reductase; catalyzes the final step in ergosterol biosynthesis; mutants are viable, but lack ergosterol
FMP23_1	3.41	6.09	1.15E-09	YBR047W	Putative protein of unknown function; proposed to be involved in iron or copper homeostasis; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies
GCY1_1	2.63	10.01	0	YOR120W	Glycerol dehydrogenase; involved in an alternative pathway for glycerol catabolism used under microaerobic conditions; also has mRNA binding activity; member of the aldo-keto reductase (AKR) family; protein abundance increases in response to DNA replication stress; GCY1 has a paralog, YPR1, that arose from the whole genome duplication
GGC1_1	2.24	5.52	3.47E-08	YDL198C	Mitochondrial GTP/GDP transporter; essential for mitochondrial genome maintenance; has a role in mitochondrial iron transport; member of the mitochondrial carrier family
GOR1_1	2.43	3.43	6.02E-04	YNL274C	Glyoxylate reductase; null mutation results in increased biomass after diauxic shift; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies; protein abundance increases in response to DNA replication stress
HIS4_1	2.54	9.81	0	YCL030C	Multifunctional enzyme containing phosphoribosyl-ATP pyrophosphatase; phosphoribosyl-AMP cyclohydrolase, and histidinol dehydrogenase activities; catalyzes the second, third, ninth and tenth steps in histidine biosynthesis

HOM3_1	2.10	5.18	2.24E-07	YER052C	Aspartate kinase (L-aspartate 4-P-transferase); cytoplasmic enzyme that catalyzes the first step in the common pathway for methionine and threonine biosynthesis; expression regulated by Gcn4p and the general control of amino acid synthesis
HSC82_1	2.01	13.26	0	YMR186W	Cytoplasmic chaperone of the Hsp90 family; plays a role in determining prion variants; redundant in function and nearly identical with Hsp82p, and together they are essential; expressed constitutively at 10-fold higher basal levels than HSP82 and induced 2-3 fold by heat shock; contains two acid-rich unstructured regions that promote the solubility of chaperone-substrate complexes; HSC82 has a paralog, HSP82, that arose from the whole genome duplication
HSP82_1	2.07	4.76	1.90E-06	YPL240C	Hsp90 chaperone; redundant in function with Hsc82p; required for pheromone signaling, negative regulation of Hsf1p; docks with Tom70p for mitochondrial preprotein delivery; promotes telomerase DNA binding, nucleotide addition; protein abundance increases in response to DNA replication stress; contains two acid-rich unstructured regions that promote solubility of chaperone-substrate complexes; HSP82 has a paralog, HSC82, that arose from the whole genome duplication
LYS1_1	2.67	6.29	3.18E-10	YIR034C	Saccharopine dehydrogenase (NAD+, L-lysine-forming); catalyzes the conversion of saccharopine to L-lysine, which is the final step in the lysine biosynthesis pathway; also has mRNA binding activity
MBF1_1	2.04	7.39	1.45E-13	YOR298C-A	Transcriptional coactivator; bridges the DNA-binding region of Gcn4p and TATA-binding protein Spt15p; suppressor of frameshift mutations; protein abundance increases in response to DNA replication stress
MET10_1	2.34	5.53	3.24E-08	YFR030W	Subunit alpha of assimilatory sulfite reductase; complex converts sulfite into sulfide

MET13_1	3.03	3.45	5.64E-04	YGL125W	Major isozyme of methylenetetrahydrofolate reductase; catalyzes the reduction of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate in the methionine biosynthesis pathway
MET14_1	2.40	3.84	1.21E-04	YKL001C	Adenylylsulfate kinase; required for sulfate assimilation and involved in methionine metabolism
MET16_1	2.61	4.65	3.31E-06	YPR167C	3'-phosphoadenylsulfate reductase; reduces 3'-phosphoadenylyl sulfate to adenosine-3',5'-bisphosphate and free sulfite using reduced thioredoxin as cosubstrate, involved in sulfate assimilation and methionine metabolism
MET1_1	2.63	4.97	6.80E-07	YKR069W	S-adenosyl-L-methionine uroporphyrinogen III transmethylase; involved in the biosynthesis of siroheme, a prosthetic group used by sulfite reductase; required for sulfate assimilation and methionine biosynthesis
MET22_1	2.06	4.87	1.13E-06	YOL064C	Bisphosphate-3'-nucleotidase; involved in salt tolerance and methionine biogenesis; dephosphorylates 3'-phosphoadenosine-5'-phosphate and 3'-phosphoadenosine-5'-phosphosulfate, intermediates of the sulfate assimilation pathway
MET32_1	2.01	4.02	5.93E-05	YDR253C	Zinc-finger DNA-binding transcription factor; involved in transcriptional regulation of the methionine biosynthetic genes; targets strong transcriptional activator Met4p to promoters of sulfur metabolic genes; feedforward loop exists in the regulation of genes controlled by Met4p and Met32p; lack of such a loop for MET31 may account for the differential actions of Met32p and Met31p; MET32 has a paralog, MET31, that arose from the whole genome duplication

MET3_1	2.43	3.66	2.50E-04	YJR010W	ATP sulfurylase; catalyzes the primary step of intracellular sulfate activation, essential for assimilatory reduction of sulfate to sulfide, involved in methionine metabolism
MET5_1	2.36	4.58	4.72E-06	YJR137C	Sulfite reductase beta subunit; involved in amino acid biosynthesis, transcription repressed by methionine
MET6_1	2.48	4.41	1.05E-05	YER091C	Cobalamin-independent methionine synthase; involved in methionine biosynthesis and regeneration; requires a minimum of two glutamates on the methyltetrahydrofolate substrate, similar to bacterial metE homologs
MPE1_1	2.55	5.82	5.83E-09	YKL059C	Essential conserved subunit of CPF cleavage and polyadenylation factor; plays a role in 3' end formation of mRNA via the specific cleavage and polyadenylation of pre-mRNA, contains a putative RNA-binding zinc knuckle motif; relocalizes to the cytosol in response to hypoxia
MSI1_1	4.00	9.89	0	YBR195C	Subunit of chromatin assembly factor I (CAF-1); chromatin assembly by CAF-1 affects multiple processes including silencing at telomeres, mating type loci, and rDNA; maintenance of kinetochore structure; deactivation of DNA damage checkpoint after DNA repair; chromatin dynamics during transcription; and repression of divergent noncoding transcription; Msi1p localizes to nucleus and cytoplasm and independently regulates the RAS/cAMP pathway via sequestration of Npr1p kinase
NDJ1_1	2.08	4.39	1.15E-05	YOL104C	Meiosis-specific telomere protein; required for bouquet formation, effective homolog pairing, ordered cross-over distribution, sister chromatid cohesion at meiotic telomeres, chromosomal segregation and telomere-led rapid prophase movement

NIT1_1	2.52	3.99	6.68E-05	YIL164C	Nitrilase; member of the nitrilase branch of the nitrilase superfamily; in closely related species and other S. cerevisiae strain backgrounds YIL164C and adjacent ORF, YIL165C, likely constitute a single ORF encoding a nitrilase gene
NRD1_1	2.47	6.73	1.67E-11	YNL251C	RNA-binding subunit of Nrd1 complex; complex interacts with exosome to mediate 3'-end formation of some mRNAs, snRNAs, snoRNAs, and CUTs; interacts with CTD of RNA pol II large subunit Rpo21p at phosphorylated Ser5 to direct transcription termination of non-polyadenylated transcripts; H3K4 trimethylation of transcribed regions by Set1p enhances recruitment of Nrd1p to those sites; role in regulation of mitochondrial abundance and cell size
NRK1_1	2.28	4.00	6.45E-05	YNL129W	Nicotinamide riboside kinase; catalyzes the phosphorylation of nicotinamide riboside and nicotinic acid riboside in salvage pathways for NAD+ biosynthesis
ODC2_1	2.84	5.18	2.22E-07	YOR222W	Mitochondrial inner membrane transporter; exports 2-oxoadipate and 2-oxoglutarate from the mitochondrial matrix to the cytosol for use in lysine and glutamate biosynthesis and in lysine catabolism; ODC2 has a paralog, ODC1, that arose from the whole genome duplication
PDC6_1	3.58	5.04	4.66E-07	YGR087C	Minor isoform of pyruvate decarboxylase; decarboxylates pyruvate to acetaldehyde, involved in amino acid catabolism; transcription is glucose- and ethanol-dependent, and is strongly induced during sulfur limitation
PHO11_1	3.42	4.99	5.89E-07	YAR071W /// YHR215W	One of three repressible acid phosphatases; glycoprotein that is transported to the cell surface by the secretory pathway; induced by phosphate starvation and coordinately regulated by PHO4 and PHO2; PHO11 has a paralog, PHO12, that arose from a segmental duplication
PHO12_1	5.05	4.63	3.67E-06		

PHO5_1	5.00	3.83	1.30E-04	YBR093C	Repressible acid phosphatase; 1 of 3 repressible acid phosphatases that also mediates extracellular nucleotide-derived phosphate hydrolysis; secretory pathway derived cell surface glycoprotein; induced by phosphate starvation and coordinately regulated by PHO4 and PHO2
SDT1_1	2.61	3.86	1.11E-04	YGL224C	Pyrimidine nucleotidase; responsible for production of nicotinamide riboside and nicotinic acid riboside; overexpression suppresses the 6-AU sensitivity of transcription elongation factor S-II, as well as resistance to other pyrimidine derivatives; SDT1 has a paralog, PHM8, that arose from the whole genome duplication
SEO1_1	2.70	5.03	4.98E-07	YAL067C	Putative permease; member of the allantoate transporter subfamily of the major facilitator superfamily; mutation confers resistance to ethionine sulfoxide
SPL2_1	2.96	5.01	5.44E-07	YHR136C	Protein with similarity to cyclin-dependent kinase inhibitors; downregulates low-affinity phosphate transport during phosphate limitation by targeting Pho87p to the vacuole; upstream region harbors putative hypoxia response element (HRE) cluster; overproduction suppresses a plc1 null mutation; promoter shows an increase in Snf2p occupancy after heat shock; GFP-fusion protein localizes to the cytoplasm
SSA2_1	2.37	9.45	0	YLL024C	ATP-binding protein; involved in protein folding and vacuolar import of proteins; member of heat shock protein 70 (HSP70) family; associated with the chaperonin-containing T-complex; present in the cytoplasm, vacuolar membrane and cell wall; 98% identical with paralog Ssa1p, but subtle differences between the two proteins provide functional specificity with respect to propagation of yeast [URE3] prions and vacuolar-mediated degradations of gluconeogenesis enzymes

SSU1_1	2.11	6.31	2.76E-10	YPL092W	Plasma membrane sulfite pump involved in sulfite metabolism; required for efficient sulfite efflux; major facilitator superfamily protein
STE2_1	2.18	4.12	3.76E-05	YFL026W	Receptor for alpha-factor pheromone; seven transmembrane-domain GPCR that interacts with both pheromone and a heterotrimeric G protein to initiate the signaling response that leads to mating between haploid a and alpha cells
STE3_1	2.06	4.39	1.11E-05	YKL178C	Receptor for a factor pheromone; couples to MAP kinase cascade to mediate pheromone response; transcribed in alpha cells and required for mating by alpha cells, ligand bound receptors endocytosed and recycled to the plasma membrane; GPCR
STI1_1	2.07	7.64	2.22E-14	YOR027W	Hsp90 cochaperone; interacts with the Ssa group of the cytosolic Hsp70 chaperones and activates Ssa1p ATPase activity; interacts with Hsp90 chaperones and inhibits their ATPase activity; homolog of mammalian Hop
STR2_1	2.01	5.91	3.52E-09	YJR130C	Cystathionine gamma-synthase, converts cysteine into cystathionine; STR2 has a paralog, YML082W, that arose from the whole genome duplication
SUL2_1	2.47	3.66	2.49E-04	YLR092W	High affinity sulfate permease; sulfate uptake is mediated by specific sulfate transporters Sul1p and Sul2p, which control the concentration of endogenous activated sulfate intermediates
TMT1_1	2.57	9.13	0	YER175C	Trans-aconitate methyltransferase; cytosolic enzyme that catalyzes the methyl esterification of 3-isopropylmalate, an intermediate of the leucine biosynthetic pathway, and trans-aconitate, which inhibits the citric acid cycle
YEL057C_1	3.39	4.67	2.94E-06	YEL057C	Protein of unknown function involved in telomere maintenance; target of UME6 regulation

YIL165C_1	2.53	6.09	1.15E-09	YIL165C	Putative protein of unknown function; mutant exhibits mitophagy defects; in closely related species and other S. cerevisiae strain backgrounds YIL165C and adjacent ORF, YIL164C, likely constitute a single ORF encoding a nitrilase gene
YLR152C_1	2.48	3.96	7.37E-05	YLR152C	Putative protein of unknown function; YLR152C is not an essential gene
YLR307C-A_1	2.44	5.25	1.52E-07	YLR307C-A	Putative protein of unknown function
YNR068C_1	3.16	5.35	8.95E-08	YNR068C	Putative protein of unknown function; exhibits homology to C-terminal end of Bul1p; expressed as a readthrough product of BSC5, the readthrough locus being termed BUL3; the BUL3 readthrough product is involved in ubiquitin-mediated sorting of plasma membrane proteins and interacts with WW domains of Rsp5p in vitro, but in a functionally different way than the non-readthrough form
YOR121C_1	2.39	8.06	6.66E-16		
ADH4_1	-8.44	-6.50	8.27E-11	YGL256W	Alcohol dehydrogenase isoenzyme type IV; dimeric enzyme demonstrated to be zinc-dependent despite sequence similarity to iron-activated alcohol dehydrogenases; transcription is induced in response to zinc deficiency
ANB1_1	-2.58	-7.07	1.53E-12	YJR047C	Translation elongation factor eIF-5A; previously thought to function in translation initiation; undergoes an essential hypusination modification; expressed under anaerobic conditions; ANB1 has a paralog, HYP2, that arose from the whole genome duplication
BIO3_1	-2.94	-3.54	4.00E-04	YNR058W	7,8-diamino-pelargonic acid aminotransferase (DAPA); catalyzes the second step in the biotin biosynthesis pathway; BIO3 is in a cluster of 3 genes (BIO3, BIO4, and BIO5) that mediate biotin synthesis; BIO3 and BIO4 were acquired by horizontal gene transfer (HGT) from bacteria

BIO4_1	-3.30	-4.17	3.10E-05	YNR057C	Dethiobiotin synthetase; catalyzes the third step in the biotin biosynthesis pathway; BIO4 is in a cluster of 3 genes (BIO3, BIO4, and BIO5) that mediate biotin synthesis; BIO3 and BIO4 were acquired by horizontal gene transfer (HGT) from bacteria; expression appears to be repressed at low iron levels
DBP2_1	-2.08	-4.67	2.99E-06	YNL112W	ATP-dependent RNA helicase of the DEAD-box protein family; has a strong preference for dsRNA; interacts with YRA1; required for the assembly of Yra1p, Nab2p and Mex67p onto mRNA and formation of nuclear mRNP; involved in mRNA decay and rRNA processing; may be involved in suppression of transcription from cryptic initiation sites
FMP48_1	-2.93	-5.50	3.74E-08	YGR052W	Putative protein of unknown function; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies; induced by treatment with 8-methoxypsoralen and UVA irradiation
MCD4_1	-2.45	-7.64	2.23E-14	YKL165C	Protein involved in GPI anchor synthesis; multimembrane-spanning protein that localizes to the endoplasmic reticulum; highly conserved among eukaryotes; GPI stands for glycosylphosphatidylinositol
MHF1_1	-2.21	-7.11	1.17E-12	YOL086W-A	Component of the heterotetrameric MHF histone-fold complex; in humans the MMF complex interacts with both DNA and Mph1p ortholog FANCM, a Fanconi anemia complementation group protein, to stabilize and remodel blocked replication forks and repair damaged DNA; mhf1 srs2 double mutants are MMS hypersensitive; ortholog of human centromere constitutive-associated network (CCAN) subunit CENP-S, also known as MHF1
RGS2_1	-2.42	-5.10	3.34E-07	YOR107W	Negative regulator of glucose-induced cAMP signaling; directly activates the GTPase activity of the heterotrimeric G protein alpha subunit Gpa2p

RRG8_1	-2.50	-3.97	7.22E-05	YPR116W	Putative protein of unknown function; required for mitochondrial genome maintenance; null mutation results in a decrease in plasma membrane electron transport
RTC4_1	-3.26	-5.99	2.14E-09	YNL254C	Protein of unknown function; null mutation suppresses cdc13-1 temperature sensitivity; (GFP)-fusion protein localizes to both the cytoplasm and the nucleus
SUR2_1	-2.16	-7.52	5.43E-14	YDR297W	Sphinganine C4-hydroxylase; catalyses the conversion of sphinganine to phytosphingosine in sphingolipid biosyntheis
THI7_1	-2.32	-5.21	1.90E-07	YLR237W	Plasma membrane transporter responsible for the uptake of thiamine; contributes to uptake of 5-aminoimidazole-4-carboxamide-1-beta-D-ribofuranoside (acadesine); member of the major facilitator superfamily of transporters; mutation of human ortholog causes thiamine-responsive megaloblastic anemia
ULI1_1	-4.42	-7.92	2.38E-15	YFR026C	Putative protein of unknown function; involved in and induced by the endoplasmic reticulum unfolded protein response (UPR)
VEL1_1	-15.10	-5.14	2.75E-07	YGL258W	Protein of unknown function; highly induced in zinc-depleted conditions and has increased expression in NAP1 deletion mutants; VEL1 has a paralog, YOR387C, that arose from a single-locus duplication
YBR200W-A_1	-4.56	-5.12	3.10E-07	YBR200W-A	Putative protein of unknown function; identified by fungal homology and RT-PCR
YLR154W-E_1	-2.09	-4.00	6.43E-05	YLR154W-E	
ZAP1_1	-2.11	-4.73	2.29E-06	YJL056C	Zinc-regulated transcription factor; binds to zinc-responsive promoters to induce transcription of certain genes in presence of zinc, represses other genes in low zinc; regulates its own transcription; contains seven zinc-finger domains
ZPS1_1	-11.30	-4.39	1.12E-05	YOL154W	Putative GPI-anchored protein; transcription is induced under low-zinc conditions, as mediated by the Zap1p transcription factor, and at alkaline pH

ZRT1_1	-2.67	-3.66	2.54E-04	YGL255W	High-affinity zinc transporter of the plasma membrane; responsible for the majority of zinc uptake; transcription is induced under low-zinc conditions by the Zap1p transcription factor