Title
Metabolic Engineering of Yeast to Maximize Precursor Formation and Polyketide Production

Author
McTaggart, Tami Lee

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Metabolic Engineering of Yeast to Maximize Precursor Formation and Polyketide Production

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Chemical and Biomolecular Engineering

by

Tami Lee McTaggart

Dissertation Committee:
Professor Nancy A. Da Silva, Chair
Professor Suzanne B. Sandmeyer
Assistant Professor Han Li

2020

DEDICATION

To

my parents, Tom and Valerie

and my partner, Paul

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

My sincere thanks and appreciation go first and foremost to my committee chair and mentor, Professor Nancy Da Silva. You always had faith in me and encouraged my ideas and growth as a scientist. Without your constant support, compassion and advice – especially in the last months of my PhD – graduate school would have been worlds more challenging and much less enjoyable. I would also like to thank my committee members Professor Han Li and Professor Suzanne Sandmeyer, who have given me advice and new ideas which broadened my understanding of the field. I also want to thank Professor Szu Wang and Professor Sheryl Tsai for sitting on my qualifying committee, which helped to shape the experiments and goals of the last years of my PhD. Thanks also to Professor Martha Mecartney for facilitating the GAANN fellowship within the department.

This journey would definitely not have been possible without the support of the Da Silva lab members. In particular, I want to thank Anh Pham, Dr. Pamela B. Besada Lombana, Dr. Jin Choi, Danielle Bever, Dr. Ruben Ferndadez-Moya, Dr. Richard Que and Hannah Yocum. The late night experiments and early morning growth curves were always better when shared. All of my undergraduate mentees also helped not only in the lab, but also in development of my own research skills and teaching abilities. Thanks to Shane Bassett, Shiaki Minami, Tej Shah, Eric Troyer and Jen Creso. It is such a joy to see you all succeed. I want to extend the appreciation to Steve Weinstock, who always was around to lend a wrench and fix ailing equipment.

I would also like to thank my cohort and mentors from the University of Washington – Mila Chistoserdova and Mary Lidstrom, who gave me the opportunity to pursue biological research and develop my skills as a scientist. I also thank Prof François Baneyx, who encouraged my decision to join Prof Da Silva at UCI.

My time in Irvine wouldn't have been fulfilling without the support of some of the friends I've met through UCI. My CBEMS cohort was essential for getting through the coursework and challenges of the first few years here. Rose Pier, Thao Nguyen, Seong-Min Kim – I'm grateful for our friendship that continued through the years. I also couldn't go without speaking to the friends I met through UCI Pokémon Go. Michael, I would be a different person if you never trusted me to lead. Ivan, Daniel, Andrew, Jennifer, Sarah, Kimberly, Rachel, Victor – the last minute travel, half asleep late night grinds, strategizing and trades were always worth it. Life was always an adventure with you guys by my side.

Thank you to my partner, Paul. You've been the most steadfast, caring and helpful ally throughout my journey in higher education and in life. You've never been afraid to push me to be the best version of myself. Lastly, I'm grateful for the support from my parents, who gave everything they could to give me the best opportunities in life. Thank you both for always advocating for me and holding me to the highest standards.

I thank Elsevier and John Wiley and Sons for permission to include copyrighted material in Chapters 1 and 2, respectively, as part of my thesis/dissertation. Financial support was

provided by the University of California, Irvine, NSF Grants CBET-1803677, EEC-0813570 DOE grant DE-SC0019093 and a GAANN fellowship provided by the ED.

VITA

Tami Lee McTaggart

2011-14	Undergraduate Research Assistant – Mary E. Lidstrom Lab
2011-2014	Chief Technical Officer, Quality Control Lead – UW Biodiesel Cooperative
2014	B.S. in Chemical Engineering, University of Washington
	B.A in Mathematics, University of Washington
2014-20	Graduate Research Assistant – Nancy Da Silva Lab
2015	M.S. in Chemical and Biochemical Engineering, University of California, Irvine
2016-17	Teaching Assistant – University of California, Irvine
2016-19	GAANN Fellow – US Department of Education
2020	Ph.D. in Chemical and Biomolecular Engineering, University of California, Irvine
2020	Research Scientist – Amazon

PRESENTATIONS

Synthesis of Polyketides from Low Cost Substrates by the Thermotolerant Yeast *Kluyveromyces marxianus,* American Chemical Society National Meeting (Orlando, FL, March 2019)

Introduction to CRISPR-Cas genome editing and regulation, Guest Lecture in CBEMS 218 Bioengineering with Recombinant Microorganisms (Irvine, CA, February 2019)

Synthesis of Polyketides from Low Cost Substrates by the Thermotolerant Yeast *Kluyveromyces marxianus*, International Conference on Biomolecular Engineering – AIChE (Newport Beach, CA, January 2019)

Combinatorial strategies for optimizing polyketide production in *Saccharomyces cerevisiae*, American Chemical Society National Meeting (San Francisco, CA, April 2017)

Engineering Saccharomyces cerevisiae for optimized production of 2-pyrones, Center for Biorenewable Chemicals Fall meeting (Ames, IA, October 2016)

Diversity and Phylogeny of Lake Washington Methylotrophs, University of Washington Undergraduate Research Symposium (Seattle, WA, May 2014)

PUBLICATIONS

Paul Ellenbogen, Tami L. McTaggart. “Unauthorized access to genetic genealogy services by saliva spoofing” Manuscript under review (2020).

Tami L. McTaggart, Danielle Bever, Shane Bassett and Nancy A. Da Silva. "Synthesis of polyketides from low cost substrates by the thermotolerant yeast *Kluyveromyces marxianus.” Biotechnology and Bioengineering* 116.7 (2019): 1721-1730.

Besada-Lombana, Pamela B., Tami L. McTaggart, and Nancy A. Da Silva. "Molecular tools for pathway engineering in Saccharomyces cerevisiae." *Current opinion in Biotechnology* 53 (2018): 39-49.

Oshkin, Igor Y., David AC Beck, Andrew E. Lamb, Veronika Tchesnokova, Gabrielle Benuska, Tami L. McTaggart, Marina G. Kalyuzhnaya, Svetlana N. Dedysh, Mary E. Lidstrom, and Ludmila Chistoserdova. "Methane-fed microbial microcosms show differential community dynamics and pinpoint taxa involved in communal response." *The ISME journal* 9, no. 5 (2015): 1119.

Kalyuzhnaya, Marina G., Andrew E. Lamb, Tami L. McTaggart, Igor Y. Oshkin, Nicole Shapiro, Tanja Woyke, and Ludmila Chistoserdova. "Draft genome sequences of gammaproteobacterial methanotrophs isolated from Lake Washington sediment." *Genome announcements* 3, no. 2 (2015): e00103-15.

McTaggart, Tami L., David AC Beck, Usanisa Setboonsarng, Nicole Shapiro, Tanja Woyke, Mary E. Lidstrom, Marina G. Kalyuzhnaya, and Ludmila Chistoserdova. "Genomics of methylotrophy in Gram-positive methylamine-utilizing bacteria." *Microorganisms* 3, no. 1 (2015): 94-112.

Beck, David AC, Tami L. McTaggart, Usanisa Setboonsarng, Alexey Vorobev, Lynne Goodwin, Nicole Shapiro, Tanja Woyke, Marina G. Kalyuzhnaya, Mary E. Lidstrom, and Ludmila Chistoserdova. "Multiphyletic origins of methylotrophy in Alphaproteobacteria, exemplified by comparative genomics of Lake Washington isolates." *Environmental microbiology* 17, no. 3 (2015): 547-554.

McTaggart, Tami L., Nicole Shapiro, Tanja Woyke, and Ludmila Chistoserdova. "Draft genomes of two strains of Flavobacterium isolated from Lake Washington sediment." *Genome announcements* 3, no. 1 (2015): e01597-14.

McTaggart, Tami L., Nicole Shapiro, Tanja Woyke, and Ludmila Chistoserdova. "Draft genome of Pseudomonas sp. strain 11/12A, isolated from Lake Washington sediment." *Genome announcements* 3, no. 1 (2015): e01587-14.

McTaggart, Tami L., Nicole Shapiro, Tanja Woyke, and Ludmila Chistoserdova. "Draft genome of Janthinobacterium sp. RA13 isolated from Lake Washington sediment." *Genome announcements* 3, no. 1 (2015): e01588-14.

McTaggart, Tami L., Gabrielle Benuska, Nicole Shapiro, Tanja Woyke, and Ludmila Chistoserdova. "Draft genome sequences of five new strains of Methylophilaceae isolated from Lake Washington sediment." *Genome announcements* 3, no. 1 (2015): e01511-14.

Beck, David AC, Tami L. McTaggart, Usanisa Setboonsarng, Alexey Vorobev, Marina G. Kalyuzhnaya, Natalia Ivanova, Lynne Goodwin, Tanja Woyke, Mary E. Lidstrom, and Ludmila Chistoserdova. "The expanded diversity of Methylophilaceae from Lake Washington through cultivation and genomic sequencing of novel ecotypes." *PLoS One* 9, no. 7 (2014): e102458.

LIST OF AUTHOR CONTRIBUTIONS

This summary highlights contributions for any co-authored materials in this thesis.

Chapter 2: Synthesis of polyketides from low cost substrates by the thermotolerant yeast Kluyveromyces marxianus

TLM designed the study. **NAD** supervised the project. **TLM, DB and SB** performed the experiments. **TLM, DB, SB and NAD** wrote the manuscript.

Appendix E: Unauthorized access to genetic genealogy services by saliva spoofing

PE conceived of the study. **PE** and **TLM** designed the study. **TLM** performed the experiments. **PE** analyzed the data. **PE** and **TLM** wrote the manuscript.

ABSTRACT OF THE DISSERTATION

Metabolic Engineering of Yeast to Maximize Precursor Formation and Polyketide Production

by

Tami Lee McTaggart

Doctor of Philosophy in Chemical and Biomolecular Engineering

University of California, Irvine, 2020

Professor Nancy A. Da Silva, Chair

Increased demand for petroleum resources coupled with political, environmental and supply-chain concerns has increased the investment and interest in bio-based chemical production. In particular, polyketide production in microorganisms facilitates the expansion of sustainable, local and scalable chemical production due to the versatility of polyketides as precursors in catalytic conversion to a range of high-value chemicals including pharmaceuticals, food additives and plastics. For these reasons, optimization of polyketide production in yeast aids the sustainability and affordability of various consumer products. This work engineers and evaluates the yeasts *Saccharomyces cerevisiae* and *Kluyveromyces marxianus* for the production of triacetic acid lactone (TAL), a platform chemical precursor to 1,3-pentadiene, sorbic acid and other commodity and high-value products. When considering the production of a specific product, one fundamental choice is host selection. In conjunction with traditional host *S. cerevisiae*, we explore development of the TAL production platform in a thermotolerant, rapidly

growing and industrially relevant yeast *K. marxianus.* Prior to metabolic engineering, we demonstrate production of 1.24 g/L TAL from xylose, a low-cost substrate. The yield of this process on xylose (0.0295 mol TAL/mol carbon) is comparable to the state-of-the art in other hosts from glucose and demonstrates a breakthrough for xylose utilization in yeast.

To build on this versatile *K. marxianus* platform, and to improve the *S. cerevisiae* one, CRISPR-Cas9 was utilized to introduce two heterologous enzymes that divert carbon away from CO_2 production, resulting in a 60% improvement in TAL yield. This pathway was then introduced into a genome scale model and flux balance analysis was performed using OptKnock, resulting in new gene knockout targets that are complementary to this new pathway. In parallel to both rationally designed and computational-based metabolic engineering strategies, we developed a novel growth selection method which is built from a genome-wide CRISPR-dCas9 library and a FapR/FapO malonyl-CoA sensor native to *Bacillus subtilis*. This method couples growth rate with intracellular malonyl-CoA levels such that over many generations, yeast strains with improved malonyl-CoA are enriched within a population and can be identified through next generation sequencing (NGS).

Due to significant developments in CRISPR-Cas9, metabolic modeling and sequencing affordability, metabolic engineering can advance via many different approaches including rational, computational and genome-scale methods. In this work, we incorporate each of these strategies to improve the understanding of both the *S. cerevisiae* and *K. marxianus* platforms for the production of polyketides and to develop a framework for enhanced production of other acetyl- and malonyl-CoA based products.

Chapter 1

Introduction

Motivation

Microorganisms have been used by humans to produce valuable products since the earliest known history – there is evidence of fermented beer and wine as early as ~7000 BC (McGovern et al. 2004) and cheese in 5200 BC (McClure et al. 2018). However, truly exploiting microbial metabolism required tools that were not available until fairly recently. It was not until 1973 that scientists first reported the in vitro construction of functional plasmids in bacteria (Cohen et al. 1973) and gave roots to the field. Metabolic engineering as a discipline was not defined until 1991 by Jay Bailey, where he states that "metabolic engineering is the improvement of cellular activities by manipulation of enzymatic, transport, and regulatory functions of the cell with the use of recombinant DNA technology" (Bailey 1991). To this day, the field is constantly shaped by new developments and innovations – notably in the last decade with molecular tools such as Gibson Assembly (Gibson et al. 2009), CRISPR-Cas9 (Jinek et al. 2012) as well as the rapidly decreasing cost of benchtop sequencing, by over 8 orders of magnitude since the early 1990s, and DNA synthesis, by 3-4 orders as of 2017 (Carlson, 2017). Metabolic engineering today encompasses many diverse strategies, tools and screening methods (Figure 1) (Besada-Lombana, McTaggart, and Da Silva 2018).

Development of these tools over the past 50 years has allowed development of a diverse range of organisms into "microbial cell factories" which can generate commodity as well as specialty chemicals. Concerns about depletion of petroleum-based feedstocks as well as objections to emissions associated with these products have driven the industry towards bio-

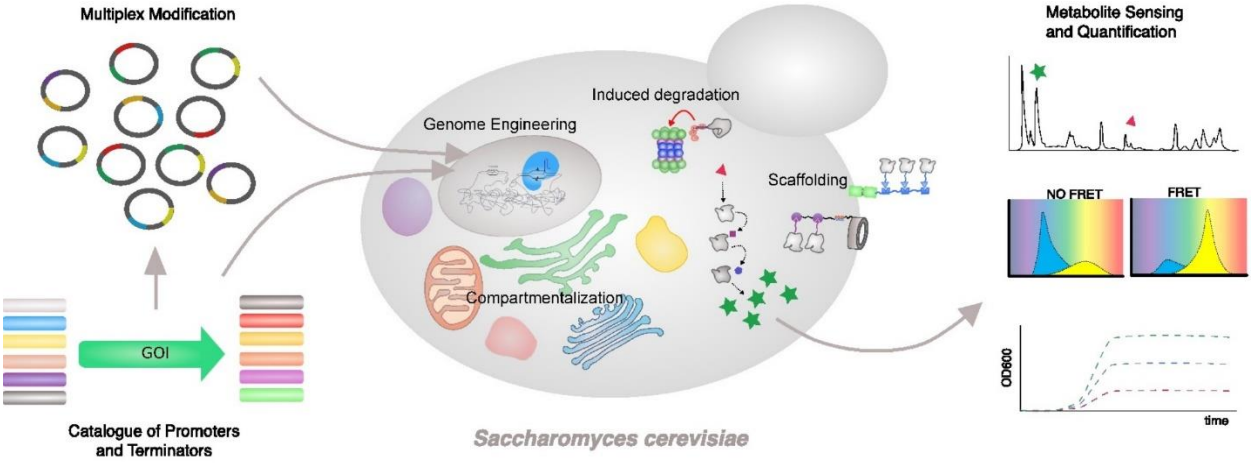


Figure 1 The diversity of molecular tools for metabolic engineering of *S. cerevisiae*

based alternatives. This biorefining industry, encompassing petroleum alternatives such as corn to ethanol, is expected to grow 8.9% per year through 2023 with a total global market of almost $717 billion; specifically, the biochemical production industry, which specializes in fine chemicals, is expected to grow at 2.1% annually (Daystar et al. 2018). While biochemicals are still a niche market, encompassing less than 1% of the chemicals market, recent supply chain issues have highlighted benefits to localized production of goods. Furthermore, metabolic engineering has applications broader than just the chemicals industry. There is huge potential in the pharmaceuticals market for the production of products such as insulin, vaccines, and recombinant protein therapies for the growing field of personalized medicine.

Saccharomyces cerevisiae (Baker's yeast) has an important role within the current metabolic engineering landscape. Robust and well-characterized, *S. cereivisiae* has been developed extensively for industrial applications in the production of natural products such as fragrances, food additives and fatty acids (Krivoruchko and Nielsen 2015; Fernandez-Moya and

Da Silva 2017). *Saccharomyces cerevisiae* is particularly suitable for expressing type I and type III polyketide synthases (Hashimoto, Nonaka, and Fujii 2014) since it can readily express a palate of plant and fungal enzymes unlike many bacterial hosts. For products to be fit for human consumption, a lengthy and expensive FDA approval process must satisfied. Therefore, host organisms such as *S. cerevisiae* which are generally recognized as safe (GRAS) and qualified presumption of safety (QPS) by the FDA (FDA 2015) and EU (Ricci et al. 2018), respectively, are highly desirable to reduce time to market. Moreover, *S. cerevisiae* has been studied as a model eukaryote for many years, with annotated genomic sequences (Engel et al. 2014), knockout libraries (Giaever and Nislow 2014) and various plasmid sets and engineering toolkits (Da Silva and Srikrishnan 2012; Besada-Lombana, McTaggart, and Da Silva 2018; M. E. Lee et al. 2015). These advantages enable safe and efficient polyketide production.

In addition to *S. cerevisiae,* other microbial hosts such as the yeast *Kluyveromyces marxianus* are also very promising for industrial chemical production. *K. marxianus* is particularly interesting due to its high thermotolerance (growth up to 52°C) (Banat, Nigam, and Marchant 1992), acid tolerance (down to pH 2.3) (Amrane and Prigent 1998), and rapid growth rate (T_d=50min) (Groeneveld, Stouthamer, and Westerhoff 2009). Like *S. cerevisiae*, *K. marxianus* also is GRAS and QPS (FDA 2015), and has been used widely in industrial dairy production of chymosin (Administration 2015) and in production of inulinase (Cruz-Guerrero et al. 1995) and β-galactosidase (H. X. Zhou et al. 2013). However, unlike *S. cerevisiae*, there is much less previous work and information on the genome, function and metabolism of this organism, and fewer

developed pathway engineering tools. Development of these tools can expedite the understanding of *K. marxianus* metabolism and thus improve the outlook for bio-based products.

Natural products such as polyketides can facilitate the expansion of sustainable chemical production due to their versatility as precursors in catalytic conversion to a range of high-value chemicals (Nikolau et al. 2008). Polyketide production in microorganisms upgrades affordable sugars to a variety of products including pharmaceuticals, food additives and plastics (Staunton and Weissman 2001).

The simple polyketide, triacetic acid lactone (TAL), is a constructed from the metabolic precursors acetyl-CoA and malonyl-CoA (extender molecule). TAL can be catalytically converted to many commodity chemicals such as sorbic acid, pogostone and acetyl acetone (Shanks and Keeling 2017), and previous work in our lab has engineered *S. cerevisiae* for production of 1.7 g/L in 5mL cultures (Cardenas and Da Silva 2014) and 10.5 g/L in fed-batch cultures (Vickery et al. 2018a). Additionally, since TAL is formed from simple metabolic building blocks, increased TAL production is an indicator of metabolic interventions that can translate to increases in more complex polyketide products. Production of TAL in *S. cerevisiae* is accomplished via introduction of a type III polyketide synthase (PKS) 2-pyrone synthase (2-PS) from the *Gebera hybrida* daisy (Eckermann et al. 1998). TAL is produced from one acetyl- and two malony-CoA molecules (Figure

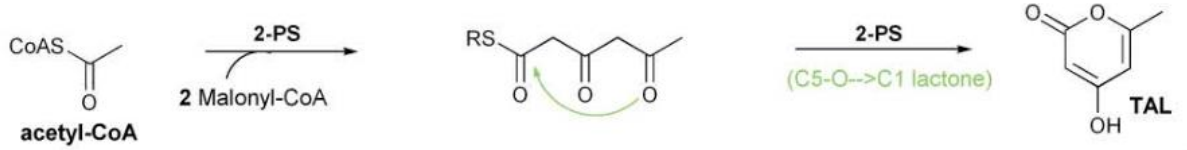


Figure 2 Cyclization of TAL from acetyl-CoA starter and malonyl-CoA extender molecules as catalyzed by 2PS

2) (Austin and Noel 2003). In order to increase TAL production, it is essential to maximize the concentration of these precursors. Due to competition for acetyl- and malonyl-CoA, it is important that these precursors are used quickly and efficiently. Therefore it is critical that the 2-PS enzyme itself is active and present in the cell at sufficient concentrations.

Unfortunately, residual components from the rich media employed for high level synthesis in *S. cerevisiae* are challenging to separate from TAL in an aqueous environment and reduce catalytic efficiency in downstream conversion (Schwartz et al. 2014). In Chapter 2, this work addresses some of the concerns with complex medium through the development of the TAL production platform in *K. marxianus*, which is able to utilize minimal medium and nonconventional feedstocks, in particular xylose, to produce TAL titers that rival those of *S. cerevisiae*. Unique and industrially desirable features of *K. marxianus* are highlighted in the development of the TAL production platform from a range of cost effective substrates and at temperatures as high as 42°C. Furthermore, our development of a high copy expression system in *K. marxianus* demonstrates the robustness and potential of this species for metabolic engineering.

Another prominent concern with bio-based chemical production is with the cost and sustainability of glucose as the primary substrate. Production of high-value chemicals such as pharmaceuticals are economically viable. However, there are well-known difficulties with lower cost products such as ethanol, due to the cost of sugar feedstocks and the poor efficiency in hydrolysis of lignocellulosic biomass (Sun and Cheng 2002). Chapter 3 aims to increase the theoretical yield of TAL from sugars through various metabolic interventions, including a new

pathway that avoids carbon dioxide formation, enabling higher conservation of carbon. Chapter 3 first explores the introduction of this pathway in *S. cerevisiae*, then extends the development into *K. marxianus.*

From previous work in the lab, many successful interventions have been identified for increasing acetyl-CoA pools and flux through central carbon metabolism; however, there is a challenge in combining these methods efficiently, strategically and reproducibly. Chapter 4 focuses on development of a new malonyl-CoA sensor method, which relies on CRISPR/dCas9 to generate randomized knockout libraries. These libraries can then be selectively grown and evaluated for improved malonyl-CoA production.

Strategies for improving microbial productivity are numerous, and a few are addressed within this thesis; however, combining them rapidly and strategically will be essential for the future growth and development of this industry and allow convergence on robust and economically viable strains.

Objectives

To discover novel metabolic engineering strategies for improvement of precursor formation and polyketide production in yeast, this thesis research had the following objectives:

Chapter 2: Synthesis of polyketides from low cost substrates by the thermotolerant yeast Kluyveromyces marxianus

Objective #1 – Characterize the *K. marxianus* yeast expression system as a function of temperature and carbon substrate

Objective #2 – Develop a polyketide expression system that is stable with high copy number

Objective #3 – Refine the expression system with promoter and enzyme engineering

Chapter 3: Improved polyketide synthesis via a heterologous carbon-saving pathway in the yeasts *Saccharomyces cerevisiae* and *Kluyveromyces marxianus*

Objective #1 – Design a heterologous carbon-saving pathway in both yeasts

Objective #2 – Refine CRISPR systems for the ease of genomic integration of heterologous pathway genes

Objective #3 – Implement and evaluate the carbon-saving pathway for the production of polyketides in both yeasts and on a range of substrates

Objective #4 – Implement carbon saving into a genome scale model

Chapter 4: Development of a novel CRISPRi growth screen for selection of malonyl-CoA overproducers

Objective #1 – Develop novel screening methodologies to select for malonyl-CoA overproducers from a large library of strain variants

Objective #2 – Build and validate the malonyl-CoA sensor and gRNA library methodology in both *S. cerevisiae* and *K. marxianus*

Objective #3 – Implement the malonyl-CoA sensor for the identification of novel gene downregulation candidates

References

Administration, U.S. Food and Drug. 2015. “Code of Federal Regulations Title 21.” Code of Federal Regulation Title 21. 2015. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=1271

Amrane, A., and Y. Prigent. 1998. “Effect of Culture Conditions of Kluyveromyces Marxianus on Its Autolysis, and Process Optimization.” Bioprocess Engineering 18 (5): 383–88. https://doi.org/10.1007/s004490050460.

Austin, Michael B., and Joseph P. Noel. 2003. “The Chalcone Synthase Superfamily of Type III Polyketide Synthases.” Natural Product Reports 20 (1): 79–110. https://doi.org/10.1039/b100917f.

Bailey, James E. 1991. “Toward a Science of Metabolic Engineering.” Science. https://doi.org/10.1126/science.2047876.

Banat, I. M., P. Nigam, and R. Marchant. 1992. “Isolation of Thermotolerant, Fermentative Yeasts Growing at 52-Degree C and Producing Ethanol at 45- Degree C and 50-Degree C.” World Journal of Microbiology & Biotechnology 8 (3): 259–63. https://doi.org/10.1007/BF01201874.

Besada-Lombana, P.B., T.L. McTaggart, and N.A. Da Silva. 2018. “Molecular Tools for Pathway Engineering in Saccharomyces Cerevisiae.” Current Opinion in Biotechnology 53. https://doi.org/10.1016/j.copbio.2017.12.002.

Cardenas, Javier, and Nancy A. Da Silva. 2014. “Metabolic Engineering of Saccharomyces Cerevisiae for the Production of Triacetic Acid Lactone.” Metabolic Engineering 25: 194–203. https://doi.org/10.1016/j.ymben.2014.07.008.

Carlson, Rob. n.d. “On DNA and Transistors — Synthesis.” Website. http://www.synthesis.cc/synthesis/2016/03/on_dna_and_transistors.

Cohen, S. N., A. C.Y. Chang, H. W. Boyer, and R. B. Helling. 1973. “Construction of Biologically Functional Bacterial Plasmids in Vitro.” Proceedings of the National Academy of Sciences of the United States of America. https://doi.org/10.1073/pnas.70.11.3240.

Cruz-Guerrero, Alma, Ines Garcia-Peña, Eduardo Barzana, Mariano Garcia-Garibay, and Lorena Gomez-Ruiz. 1995. “Kluyveromyces Marxianus CDBB-L-278: A Wild Inulinase Hyperproducing Strain.” Journal of Fermentation and Bioengineering 80 (2): 159–63. https://doi.org/10.1016/0922-338X(95)93212-3.

Daystar, Jesse, Robert Handfield, Jay Stuart Golden, Eric McConnell, and Brandon Morrison. 2018. “An Economic Impact Analysis of the U.S. Biobased Products Industry (2018).” file:///C:/Users/monic/Downloads/2018_Biobased_Products_Econ_Impact_Report.pdf.

Eckermann, Stefan, Gudrun Schröder, Jürgen Schmidt, Dieter Streck, Ru A. Edrada, Yrjö Helariutta, Paula Elomaa, et al. 1998. “New Pathway to Polyketides in Plants.” Nature 396 (6709): 387–90. https://doi.org/10.1038/24652.

Engel, Stacia R, Fred S Dietrich, Dianna G Fisk, Gail Binkley, Rama Balakrishnan, Maria C Costanzo, Selina S Dwight, et al. 2014. “The Reference Genome Sequence of

Saccharomyces Cerevisiae: Then and Now." G3 (Bethesda, Md.) 4 (3): 389–98. https://doi.org/10.1534/g3.113.008995.

FDA. 2015. "Microorganisms & Microbial-Derived Ingredients Used in Food (Partial List)." Food And Drug Administration. 2015. http://www.fda.gov/Food/IngredientsPackagingLabeling/GRAS/MicroorganismsMicrobialDerivedIngredients/default.htm.

Fernandez-Moya, Ruben, and Nancy A. Da Silva. 2017. "Engineering Saccharomyces Cerevisiae for High-Level Synthesis of Fatty Acids and Derived Products." FEMS Yeast Research. https://doi.org/10.1093/femsyr/fox071.

Giaever, Guri, and Corey Nislow. 2014. "The Yeast Deletion Collection: A Decade of Functional Genomics." Genetics 197 (2): 451–65. https://doi.org/10.1534/genetics.114.161620.

Gibson, Daniel G., Lei Young, Ray Yuan Chuang, J. Craig Venter, Clyde A. Hutchison, and Hamilton O. Smith. 2009. "Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases." Nature Methods. https://doi.org/10.1038/nmeth.1318.

Groeneveld, Philip, Adriaan H. Stouthamer, and Hans V. Westerhoff. 2009. "Super Life - How and Why 'cell Selection' Leads to the Fastest-Growing Eukaryote." FEBS Journal 276 (1): 254–70. https://doi.org/10.1111/j.1742-4658.2008.06778.x.

Hashimoto, Makoto, Takamasa Nonaka, and Isao Fujii. 2014. "Fungal Type III Polyketide Synthases." Nat. Prod. Rep. 31 (10): 1306–17. https://doi.org/10.1039/C4NP00096J.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. “A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity.” Science 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Krivoruchko, Anastasia, and Jens Nielsen. 2015. “Production of Natural Products through Metabolic Engineering of Saccharomyces Cerevisiae.” Current Opinion in Biotechnology. https://doi.org/10.1016/j.copbio.2014.12.004.

Lee, Michael E., William C. DeLoache, Bernardo Cervantes, and John E. Dueber. 2015. “A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly.” ACS Synthetic Biology 4 (9): 975–86. https://doi.org/10.1021/sb500366v.

McClure, Sarah B., Clayton Magill, Emil Podrug, Andrew M.T. Moore, Thomas K. Harper, Brendan J. Culleton, Douglas J. Kennett, and Katherine H. Freeman. 2018. “Fatty Acid Specific Δ13C Values Reveal Earliest Mediterranean Cheese Production 7,200 Years Ago.” PLoS ONE. https://doi.org/10.1371/journal.pone.0202807.

McGovern, Patrick E., Juzhong Zhang, Jigen Tang, Zhiqing Zhang, Gretchen R. Hall, Robert A. Moreau, Alberto Nuñez, et al. 2004. “Fermented Beverages of Pre- and Proto-Historic China.” Proceedings of the National Academy of Sciences of the United States of America. https://doi.org/10.1073/pnas.0407921102.

Nikolau, Basil J., M. Ann D.N. Perera, Libuse Brachova, and Brent Shanks. 2008. “Platform Biochemicals for a Biorenewable Chemical Industry.” Plant Journal. https://doi.org/10.1111/j.1365-313X.2008.03484.x.

Ricci, Antonia, Ana Allende, Declan Bolton, Marianne Chemaly, Robert Davies, Rosina Girones, Konstantinos Koutsoumanis, et al. 2018. “Update of the List of QPS-recommended Biological Agents Intentionally Added to Food or Feed as Notified to EFSA 7: Suitability of Taxonomic Units Notified to EFSA until September 2017.” EFSA Journal 16 (1). https://doi.org/10.2903/j.efsa.2018.5131.

Schwartz, Thomas J., Robert L. Johnson, Javier Cardenas, Adam Okerlund, Nancy A. Da Silva, Klaus Schmidt-Rohr, and James A. Dumesic. 2014. “Engineering Catalyst Microenvironments for Metal-Catalyzed Hydrogenation of Biologically Derived Platform Chemicals.” Angewandte Chemie - International Edition 53 (47): 12718–22. https://doi.org/10.1002/anie.201407615.

Shanks, Brent H., and Peter L. Keeling. 2017. “Bioprivileged Molecules: Creating Value from Biomass.” Green Chem. 19 (14): 3177–85. https://doi.org/10.1039/C7GC00296C.

Silva, Nancy A. Da, and Sneha Srikrishnan. 2012. “Introduction and Expression of Genes for Metabolic Engineering Applications in Saccharomyces Cerevisiae.” FEMS Yeast Research 12 (2): 197–214. https://doi.org/10.1111/j.1567-1364.2011.00769.x.

Staunton, James, and Kira J. Weissman. 2001. “Polyketide Biosynthesis: A Millennium Review.” Natural Product Reports 18 (4): 380–416. https://doi.org/10.1039/a909079g.

Sun, Ye, and Jiayang Cheng. 2002. “Hydrolysis of Lignocellulosic Materials for Ethanol Production : A Review.” Bioresource Technology 83 (1): 1–11. https://doi.org/10.1016/S0960-8524(01)00212-7.

Vickery, Christopher R., Javier Cardenas, Marianne. Bowman, Michael D. Burkart, Nancy A. Da Silva, and Joseph P. Noel. 2018. “A Coupled in Vitro/in Vivo Approach for Engineering a Heterologous Type III PKS to Enhance Polyketide Biosynthesis in Saccharomyces Cerevisiae.” Biotechnology and Bioengineering, no. December 2017: 1–9. https://doi.org/10.1002/bit.26564.

Zhou, Hai Xiang, Jin Li Xu, Zhe Chi, Guang Lei Liu, and Zhen Ming Chi. 2013. “β-Galactosidase over-Production by a Mig1 Mutant of Kluyveromyces Marxianus KM for Efficient Hydrolysis of Lactose.” Biochemical Engineering Journal 76. https://doi.org/10.1016/j.bej.2013.04.010.

Chapter 2

Synthesis of polyketides from low cost substrates by the thermotolerant yeast *Kluyveromyces marxianus*

Abstract

Kluyveromyces marxianus is a promising nonconventional yeast for biobased chemical production due to its rapid growth rate, high TCA cycle flux, and tolerance to low pH and high temperature. Unlike *S. cerevisiae, K. marxianus* grows on low-cost substrates to cell densities that equal or surpass densities in glucose, which can be beneficial for utilization of lignocellulosic biomass (xylose), biofuel production waste (glycerol), and whey (lactose). We have evaluated *K. marxianus* for the synthesis of polyketides, using triacetic acid lactone (TAL) as the product. The 2-pyrone synthase (2-PS) was expressed on a CEN/ARS plasmid in three different strains, and the effects of temperature, carbon source, and cultivation strategy on TAL levels were determined. The highest titer was obtained in defined 1% xylose medium at 37^{o}C, with substantial titers at 41^{o}C and 43^{o}C. The introduction of a high-stability 2-PS mutant and a promoter substitution increased titer 4-fold. 2-PS expression from a multi-copy pKD1-based plasmid improved TAL titers a further 5-fold. Combining the best plasmid, promoter, and strain resulted in a TAL titer of 1.24 g/L and a yield of 0.030 mol TAL/mol C for this otherwise unengineered strain in 3mL tube culture. This is an excellent titer and yield (on xylose) prior to metabolic engineering or fed-batch culture relative to other hosts (on glucose), and demonstrates the promise of this rapidly growing and thermotolerant yeast species for polyketide production.

Introduction

The yeast *Kluyveromyces marxianus* has significant industrial potential due to its desirable growth characteristics and metabolism. *K. marxianus* is thermotolerant to 52°C (Banat, Nigam, and Marchant 1992), resistant to heat shock and temperature fluctuation (Matsumoto, Arai, and Nishimoto 2018), and acid tolerant to pH 2.3 (Amrane and Prigent 1998), all important attributes to both prevent contamination and reduce cooling costs in industrial fermentation. This yeast also boasts a very rapid doubling time of 45 mins in glucose medium (Groeneveld, Stouthamer, and Westerhoff 2009) and is considered the fastest growing eukaryote. It has the ability to grow on a wide range of carbon sources including xylose, glycerol, lactose, and many others (M. M. Lane and Morrissey 2010). *K. marxianus* is commonly found in fermented dairy products, and has applications for production of fragrance and flavor molecules (Morrissey et al. 2015) enabled by GRAS (FDA 2015) and QPS (Ricci et al. 2018) status in the United States and European Union, respectively. This yeast has also been observed to readily secrete enzymes, an attribute that has been utilized for simultaneous saccharification and fermentation of cellulose for ethanol production (Ballesteros et al. 2004; Tomás-Pejó et al. 2009). These traits make *K. marxianus* particularly suited to industrial production of enzymes such as pectinase (Espinoza et al. 1992), inulinase (Kushi, Monti, and Contiero 2000; Galindo-Leva et al. 2016) and β-galactosidase (Rech et al. 1999; H. X. Zhou et al. 2013), PCV2 virus-like particles (Duan et al. 2018), as well as other chemicals such as ethyl acetate (Löser et al. 2015; Löbs et al. 2017), xylitol (B. Zhang et al. 2016; J. S. Kim et al. 2015) and 2-phenylethanol (Martínez et al. 2018; F. Gao and Daugulis 2009).

Despite significant industrial advantages, *K. marxianus* has not been widely applied relative to other well-developed yeast hosts such as *Saccharomyces cerevisiae* due to a limited toolkit available for metabolic engineering. While low-copy plasmids have been developed (Iborra and Ball 1994; M. M. Ball et al. 1999) from genomic centromere and autonomously replicating (CEN/ARS) sequences, there is no known native 2µ -like plasmid for multi-copy plasmid development. In the closely related *Kluyveromyces lactis*, the multi-copy 2µ-like pKD1 plasmid (Falcone et al. 1986) has been studied extensively for expression of a range of heterologous enzymes (Spohner et al. 2016; Hsieh and Da Silva 1998; Cai et al. 2005). However, pKD1-based vectors have been applied in only limited studies in *K. marxianus* and these vectors have been observed to be relatively unstable in both CBS 712 (21.4%) and CBS 6556 (35.1%). As in *K. lactis*, linearization of the pKD1 sequence at SphI did improve stability in *K. marxianus* strain BKM Y-719 to 72% (Bartkeviciute, Siekstele, and Sasnauskas 2000). An alternate multi-copy pDblet plasmid from *Schizosaccharomyces pombe* has also been used in CBS6556 with reported stability of 75% (De Souza and De Morais 2000). Rapid development of CRISPR/Cas9 systems for *K. marxianus* (Nambu-Nishida et al. 2017; Löbs et al. 2017; M. H. Lee et al. 2018; Juergens et al. 2018) has recently expanded the ability to modify this yeast as a cell factory.

Metabolic studies of *K. marxianus* have demonstrated unique traits that are beneficial for waste substrate utilization including a range of sugar degradation enzymes and multiple significant and independent sugar transport systems (Postma and Van den Broek 1990; De Bruijne et al. 1988). Metabolism of single sugars as well as mixed sugar feedstocks results in minimal by-product formation relative to other yeasts, including very little ethanol (Crabtree-

negative) (Fonseca, de Carvalho, and Gombert 2013; Fonseca et al. 2007), which is desirable for product separation and yield. *K. marxianus* also has higher TCA-cycle and pentose phosphate pathway flux in glucose relative to *S. crerevisiae* (Blank, Lehmbeck, and Sauer 2005). Although growth rate is highest in glucose, with significant catabolite repression of other sugar degradation pathways (Rodrussamee et al. 2011; Lertwattanasakul et al. 2011), the ability of *K. marxianus* to uptake xylose is notably high (Stambuk et al. 2003) and has been recently enhanced (Sharma et al. 2016), and a weak xylose transporter had been identified (Knoshaug et al. 2015). Increased biomass accumulation in xylose media relative to glucose is thought to be related to phosphoketolase activity (Evans and Ratledge 1984). Recent metabolic models of *K. marxianus* also show xylose to be a promising substrate, especially after adjusting enzyme cofactor preferences and activity in the xylose reductase step (Pentjuss et al. 2017). This organism's unique metabolism and ability to assimilate a wide range of substrates makes it a promising host for the synthesis of acetyl-CoA based products such as polyketides.

Polyketides are a diverse class of natural products of great importance due to their bioactive properties and structural diversity (Pfeifer and Khosla 2001). Important applications include their use as therapeutics and more recently as precursors for conversion to biorenewable chemicals (Shanks and Keeling 2017; Kraus, Wanninayake, and Bottoms 2016). Yeast are particularly promising for the expression of Type I (iterative) and Type III fungal and plant polyketide synthases (Hashimoto, Nonaka, and Fujii 2014), and for synthesis of polyketides built from acetyl-CoA and malonyl-CoA. Examples include dihydromonacolin L (the precursor to lovastatin) using the *Aspergillus terreus* lovastatin nonaketide synthase (LovB) (Ma et al. 2009),

6-methylsalicylic acid using the *Penicillium patulum* 6-methylsalicylic acid synthase (6-MSAS) (Kealey et al. 1998), and triacetic acid lactone (TAL) using the *Gerbera hybrida* 2-pyrone synthase (2-PS) (Cardenas and Da Silva 2014; Xie et al. 2006). There has been significant attention on the synthesis of TAL as it is can be converted into a wide range of high-value and commodity products (Chia et al. 2012). TAL is also a simple polyketide that requires expression of only one synthase enzyme and is easily assayed; it can thus be used as an effective and rapid indicator of strains with the high acetyl-CoA and malonyl-CoA pools needed for polyketide production. Extensive strain, pathway, and synthase engineering have been performed to increase TAL synthesis in *S. cerevisiae* (Cardenas and Da Silva 2016; L. P. Saunders et al. 2015; Cardenas and Da Silva 2014). More recently, high TAL titers have been achieved via engineering of the oleaginous yeast *Yarrowia lipolitica* (Markham et al. 2018; Yu et al. 2018), taking advantage of the native high flux through acetyl-CoA. In both yeast species, TAL was synthesized from glucose-based media.

K. marxianus holds significant potential for polyketide synthesis from a variety of carbon sources and under industrially favorable conditions due to its fast growth rate, thermotolerance, and acid tolerance. To demonstrate the promise of this yeast species, we evaluated both growth and TAL production using eleven different carbon sources in three *K. marxianus* strains, and over a temperature range of 30^{o}C to 43^{o}C. We also improved the plasmid-based expression system and demonstrated the ability of this yeast to produce significant TAL titers in very minimal medium. The results clearly demonstrate the promise of *K. marxianus* for polyketide production on substrates such as xylose, glycerol, and lactose and over a large temperature range.

Materials and methods

Strains and plasmids

Escherichia coli strain DH5α (Invitrogen, Carlsbad, CA) was used for plasmid maintenance and amplification. The *K. marxianus* strains employed were CBS6556Δ*ura3* (Löbs et al. 2017), CBS 712 (ATCC 200963; ATCC®, Manassas, VA), and KM1Δ*ura3* (Pecota, Rajgarhia, and Da Silva 2007). For CBS 712, the *URA3* locus was disrupted as previously described (Pecota, Rajgarhia, and Da Silva 2007) resulting in CBS712Δ*ura3*. The Δ*ura3* strains were transformed using a standard *S. cerevisiae* lithium acetate method (Gietz et al. 1992) with the heat shock step performed at 45°C for 45 minutes. The Frozen-EZ Yeast Transformation II Kit (Zymo Research, Irvine, CA) was also used as it was found to be an efficient and rapid method for *K. marxianus* transformation.

Table 1 Primers used for cloning of *K. marxianus* plasmids and *URA3* knockout

Primer Sequence	Primer Purpose
TTCAGAATTCGAGCTCCTTTCATTTCTGAT	CEN/ARS$_{KM}$ origin f
TTCAGAATTCATCGATTGAAGTTTTGTCCA	CEN/ARS$_{KM}$ origin r
CGAGGTACCGTAACTTCTGATCCGAGTACA	URA3$_{KM}$ knockout donor f
CTGAAATTAGGTGCCTGTCACGG	URA3$_{KM}$ knockout donor r
ATCGGACGTCGGAGCACGATATCTTGGTCATTGC	PGK1$_{KM}$ f
ATCGACTAGTCATTTTTGTATCTTTATATAGGTAG	PGK1$_{KM}$ r
cgtcgtatgcttttcattagtgatgcatgcGAATTCGGTAATCATGGTCATAGCTGTTTC	pKD1(SphI) Gibson f
cgtgccgattcgcacgctgcaacggcatgcGAATTCGAGCTCGGTACCCGGGATAACTTC	pKD1(SphI) Gibson r
CAGGCATATGATTTAGCGGC	ADH2p NheI swap f
ATTACATGAGCTAGCTattacgatatagttaatag	ADH2p NheI swap r
gtaatAGCTAGCTCATGTAATTAGTTATGTCACGC	CYC1t NheI replacement f
ctgcaggcatgcaagcttGG	CYC1t NheI replacement r
taactatatcgtaatAGCTAGCatgggatcttactc	2-PS insert for pKD-A2PS f
TAACTAATTACATGAGTTTAAACtcagtttccattggc	2-PS insert for pKD-A2PS r

To construct the *K. marxianus* CEN/ARS plasmid, a 1.2kb fragment containing a centromere and autonomously replicating sequence was PCR-amplified from CBS 712 genomic DNA and flanked with EcoRI restriction sites. The plasmid pXP842-2PS (Cardenas and Da Silva 2014) which contains the *g2ps1* gene encoding 2-PS from *G. hybrida* fused to a HIS-tag (2PSHT) was digested with EcoRI to remove the *S. cerevisiae* 2μ origin and ligated to the *Km*CEN/ARS to generate pCA-A2PS. The native *K. marxianus PGK1* promoter was PCR amplified from the genomic DNA of CBS 712 and used to replace the *ScADH2p* in pCA-A2PS after digestion with AatII and SpeI, resulting in pCA-P2PS. These plasmids were digested at SpeI and XhoI restriction sites to replace the 2PS-HT with the 2-PSHT[C35S] mutant sequence (Vickery et al. 2018b). To develop the high-copy plasmid, pKD-P2PS, the *Kluyveromces lactis* pKD1 sequence was isolated from pSphI (Panuwatsuk and Da Silva 2002) by SphI digestion and Gibson assembled with a PCR product of pCA-P2PS that removed the *Km*CEN/ARS and added 30 base pairs of homology to either end of pKD1. The multi-copy plasmid with the *ScADH2* promoter (pKD-A2PS) was constructed from pCA-A using the same Gibson primers used to construct pKD-P2PS but without the 2-PS gene, resulting in pKD-A(SpeXho). We replaced the SpeI and XhoI sites with NheI using Gibson assembly with overhangs in the *ADH2* promoter and *CYC1* terminator generating pKD-A. The 2-PS gene was PCR amplified and Gibson assembled with pKD-A to form pKD-A2PS.

Plasmid recovery was performed using the GeneJet™ Plasmid Miniprep Kit (Thermo Scientific, Waltham, MA), and DNA sequence analysis confirmed the correct sequence of all PCR-amplified inserts (GeneWiz, South Plainfield, NJ; Eton Biosciences, San Diego, CA). Q5® Hot Start High-Fidelity DNA Polymerase, T4 DNA ligase, and deoxynucleotides were purchased from

New England Biolabs (Ipswich, MA). Oligonucleotide primers were purchased from IDT DNA (San Diego, CA). All primer sequences are provided in **Table 1**. The plasmids constructed are shown in **Table 2**.

Table 2 List of plasmids for *K. marxianus* expression

Plasmid	Description
pCA-P2PS	CEN/ARS_{KM}, *PGK1pKM-2PSHT-CYCt,* $URA3_{SC}$
pCA-P[C35S]	CEN/ARS_{KM}, *PGK1pKM-2PSHT[C35S]-CYCt,* $URA3_{SC}$
pCA-A	CEN/ARS_{KM}, *ADH2pSC-CYCt,* $URA3_{SC}$
pCA-A2PS	CEN/ARS_{KM}, *ADH2pSC-2PSHT-CYCt,* $URA3_{SC}$
pCA-A[C35S]	CEN/ARS_{KM}, *ADH2pSC-2PSHT[C35S]-CYCt,* $URA3_{SC}$
pKD-P2PS	PKD1KL, *PGK1pKM-2PSHT-CYCt,* $URA3_{SC}$
pKD-A	PKD1KL, *ADH2pSC-CYCt,* $URA3_{SC}$
pKD-A2PS	PKD1KL, *ADH2pSC-2PSHT-CYCt,* $URA3_{SC}$

Media and cultivation

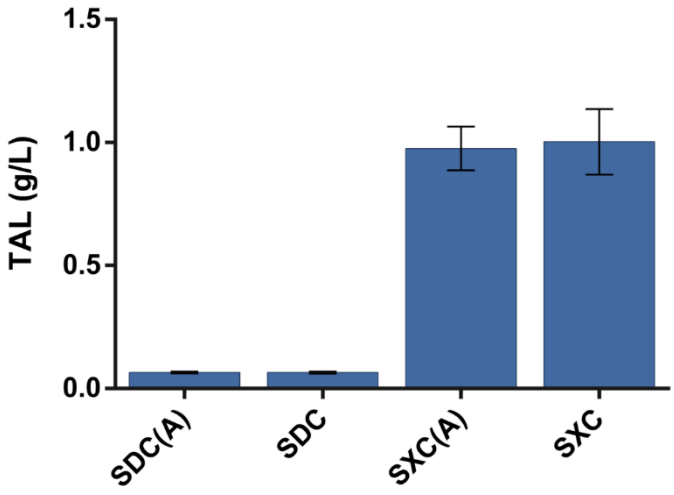


Figure 3 Strain CBS712Δ*ura3* +pKD-P2PS was grown in 1% SC medium with or without supplemented adenine (100mg/L) and TAL production was measured after 48hr cultivation at 37°C. Values are reported as mean ± standard deviation (n=3 biological replicates).

Luria-Bertani (LB) media with 150 mg/L ampicillin was used for *E. coli* culture. *K. marxianus* was cultivated in complex YP medium (1% Bacto yeast extract, 2% Bacto peptone), selective SC medium (0.67% yeast nitrogen base, 0.5% ammonium sulfate, 100mg/L adenine-sulfate), and minimal selective S medium (0.67% yeast nitrogen base, 0.5% ammonium sulfate, 0.5% casamino acids, 100mg/L adenine-sulfate), each supplemented with a variety of carbon sources. In all cases, an equimolar amount of carbon was used resulting in supplements of 1.00 g dextrose, 0.95 g cellobiose, 1.00 g xylose, 0.95 g lactose, 1.00 g galactose, 0.95 g sucrose, 1.01 g sorbitol, 1.01 g mannitol, 1.02 g glycerol, 0.98 g acetate, or 0.98 g succinate per 100 mL of media. Media containing xylose are designated YPX, SXC, and SX. The adenine supplemented in the media had no effect on growth or TAL production (**Figure 3**).

K. marxianus was first inoculated from frozen glycerol stocks or plates into 3 or 5 mL medium in 15x125mm borosilicate glass culture tubes and cultivated at 37°C and 250 rpm overnight in an air incubator shaker (New Brunswick Scientific Co. Excella E25, Edison, NJ) or a gyratory water bath shaker (New Brunswick Scientific Co. Model G67D, Edison NJ). Culture

tubes were maintained at ~45° angle for the duration of incubation to enhance gas exchange. The overnight tube culture was used to inoculate subsequent tubes or flasks to an initial optical density (OD_{600}) of 0.1 (Shimadzu UV-2450 UV-Vis Spectrophotometer, Columbia, MD) in various carbon sources and cultivated at temperatures ranging from 30 to 43°C. A 500 µL sample was taken periodically to determine cell density and to measure TAL concentration in the medium. Maximum specific growth rate (μ_{max}) was determined by measuring the OD_{600} hourly during exponential growth (generally between 4- and 15-hours post-inoculation); μ_{max} values were calculated using a minimum of 4 points during exponential phase (**Figure 4**).

Plasmid stability analysis

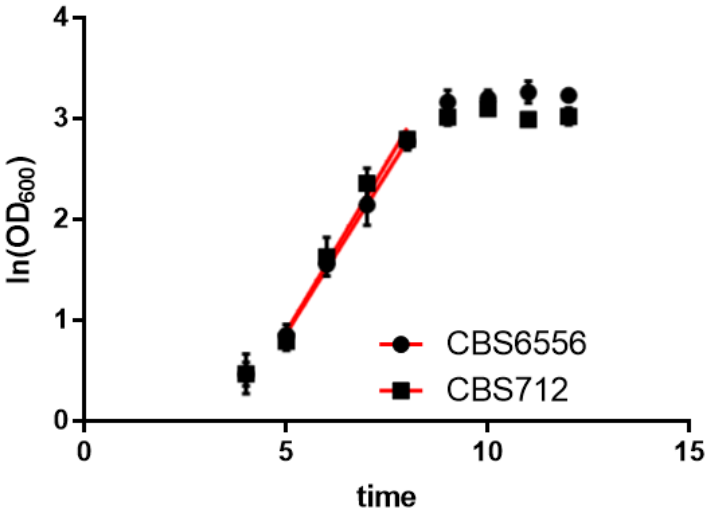


Figure 4 Sample growth curve for strains CBS712 and CBS6556 grown in YPD at 43°C in 5mL test tubes. The values of μ_{MAX} (**Table 2**) were determined using a minimum of 4 points during exponential growth, shown in red. Values shown are mean ± standard deviation (n=3 biological replicates).

Yeast strains harboring pKD1-based plasmids were cultivated on SXC plates, resuspended in water, and then inoculated to OD_{600} 0.05 in 3mL of SXC or YPX and grown for 48 hours at 37°C. Samples were taken in early exponential phase (OD_{600}=4), late exponential phase (OD_{600}=10), and stationary phase (48 h), diluted, and plated onto YPX medium. After 24 hours of growth at 37°C, 100 single colonies were streaked onto selective SXC plates for each of the three independent cultures (or ~300 colonies per time, per medium). The percent of

plasmid-containing cells was determined as the number of colonies that grew on the selective plates divided by the total number of colonies transferred.

HPLC assay

To measure TAL concentration, samples were centrifuged at 3500 rpm (2400 x g) for 5 min (Beckman Coulter Allegra X-22R Centrifuge, Brea, CA), followed by collection of the supernatants and storage at 4°C.

The concentration of TAL was measured by HPLC using a Shimadzu HPLC system: LC-10AT pumps (Shimadzu), UV–vis detector (SPD-10A VP, Shimadzu), Zorbax SB-C18 reversed-phase column (2.1×150 mm, Agilent Technologies). Acetonitrile buffered in 1% acetic acid was used as the mobile phase, while HPLC grade water buffered in 1% acetic acid was used as the aqueous phase. A gradient program using a 95–85% Pump B gradient (H_2O with 1% acetic acid) provided an elution time of approximately 12 min (flow rate 0.25 mL/min, column temperature 25 °C).

Sugars and fermentation products were also analyzed from supernatant via HPLC. Substances were detected using an RID detector (RD20-A, Shimadzu) eluting with 100% HPLC grade water in an Aminex HPX-42A column. Elution times were observed to be 26.6 min for xylose, 23 min lactose, 26 min glucose, 28.7 min for xylitol, 33.4 min ethanol, 28 min glycerol (flow rate 0.4 mL/min, column temperature 85°C). Organic acids were quantified from supernatant via HPLC using an Aminex HPX-87H column and eluting with 5mM sulfuric acid in HPLC grade water. Peaks were measured both with an RID (RD20-A, Shimadzu) and a UV-Vis (SPD10-A VP, Shimadzu) detector. Elution times observed were 13.85 min succinic acid; 15.2 min

lactic acid; 18.1 min acetic acid. TAL is included in the broad peak at 23 min (flow rate 0.5 mL/min, column temperature 65°C).

Results and discussion

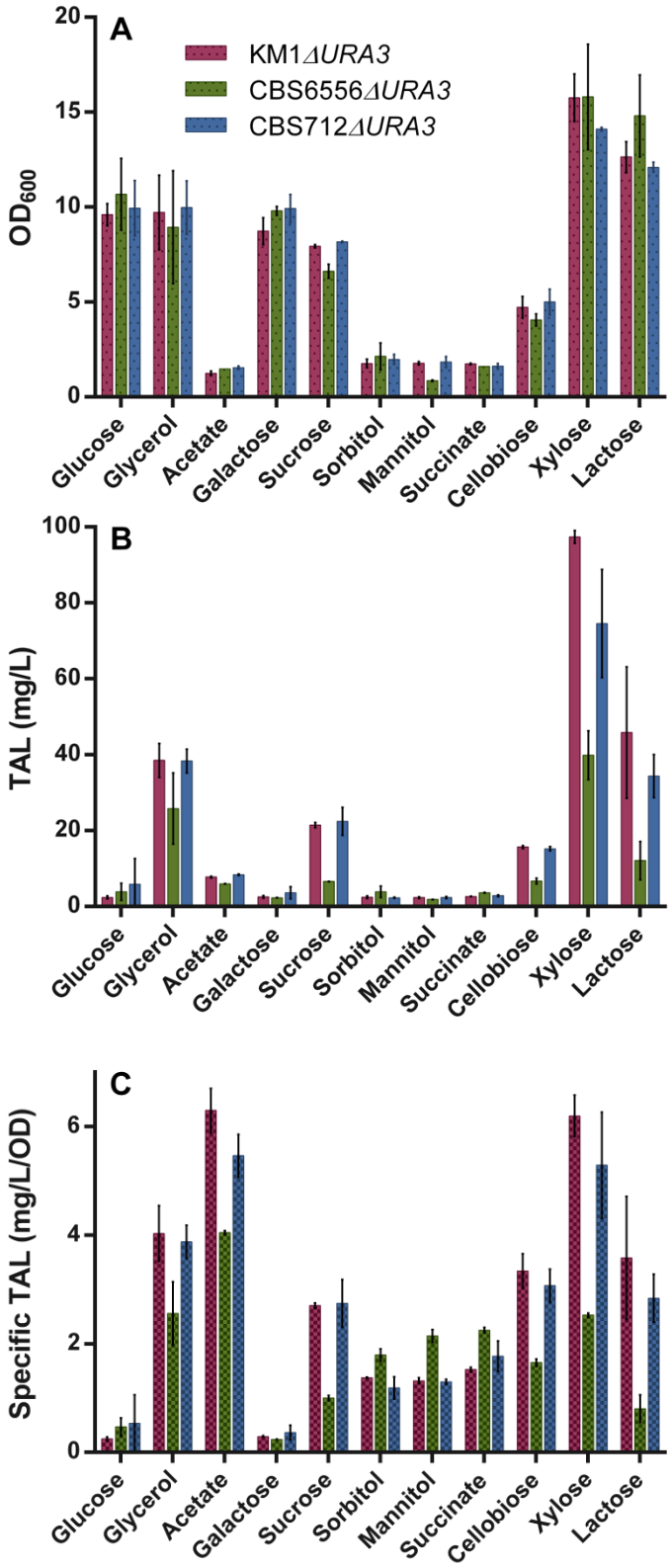
Carbon substrate effects on growth and TAL production

K. marxianus is known to grow on a wide array of feedstocks. We thus first evaluated the effect of the carbon substrate on both growth and TAL synthesis. Strains CBS6556Δ*ura3*, CBS712Δ*ura3,* and KM1Δ*ura3* were transformed with plasmid pCA-P2PS carrying a *K. marxianus* CEN/ARS sequence and the *K. marxianus PGK1* promoter for expression of the 2-PS gene. This *Km*CEN/ARS plasmid has very high stability in selective medium (ca. 90% after 48 h) (**Table 3**).

Table 3 Plasmid stability of pCA-P2PS in *K. marxianus* strains CBS6556Δ*ura3* and CBS712Δ*ura3* in selective SXC medium compared to rich YPX medium at stationary (48h) phase. Average values ± standard deviation (n=300 biological replicates).

	CBS6556Δ*ura3*		**CBS712Δ*ura3***	
	SXC	YPX	SXC	YPX
48h	0.96 ± 0.01	0.93 ± 0.05	0.89 ± 0.08	0.87 ± 0.06

The three strains were cultivated at 37ºC in selective SC medium containing eleven independent carbon sources (at equimolar carbon) for 48 h. Growth was observed on all substrates evaluated, with the highest final OD on xylose, and excellent growth on glucose, glycerol, galactose, sucrose, and lactose (**Figure 5A**). All three *K. marxianus* strains had the same



growth profile on the various carbon sources. The excellent growth on xylose, glycerol, and lactose is noteworthy. These three substrates can be derived from waste biomass, biodiesel, or waste whey, respectively, and are often poorly metabolized by other yeasts such as *S. cerevisiae*, even afer significant metabolic engineering (Strucko et al. 2018; H. Zhou et al. 2012; Kwak and Jin 2017).

We also evaluated the effect of the eleven media on the synthesis of our model polyketide TAL. At 48 h, the extracellular TAL was quantified via HPLC. Cultivation in glycerol and xylose

Figure 5 Growth and TAL synthesis for three *K. marxianus* strains carrying the low-copy pCA-P2PS plasmid. Strains were cultivated for 48h at 37°C in 5mL SC medium with a range of carbon sources (at equimolar carbon, equivalent to 1% glucose). **A)** Growth (optical density at 600nm). **B)** TAL titer. **C)** Specific titer. Bars represent mean values ± standard deviation (*n*=3 biological replicates).

resulted in notably higher TAL titers (**Figure 5B**), with somewhat lower titers in sucrose, cellobiose, and lactose. The same general trends were observed for all three strains; however, significantly higher TAL levels were reached with strains CBS712Δ*ura3* and KM1Δ*ura3* relative to strain CBS6556Δ*ura3*. Interestingly, glucose proved to be a poor carbon source for TAL synthesis, while xylose significantly outperformed all other substrates. Over 10-fold higher titers were observed in xylose relative to glucose. In addition, TAL synthesis in lactose was significantly higher than that in glucose or galactose. Specific production on acetate was comparable to xylose but growth and cell density were very low, indicating that increasing acetate feeding may be a viable strategy for TAL production without significant growth. Growth rate was significantly slower with xylose (0.25 h^{-1}) and lactose (0.52 h^{-1}) relative to glucose (0.66 h^{-1}), and less ethanol, glycerol, and organic acid byproducts were formed in selective xylose cultures of CBS712Δ*ura3* (**Figure 6**). Increased byproduct formation in glucose relative to xylose has been previously observed, as well as reduced transcription of amino acid synthesis genes (Schabort et al. 2016) which may contribute to slower growth rates.

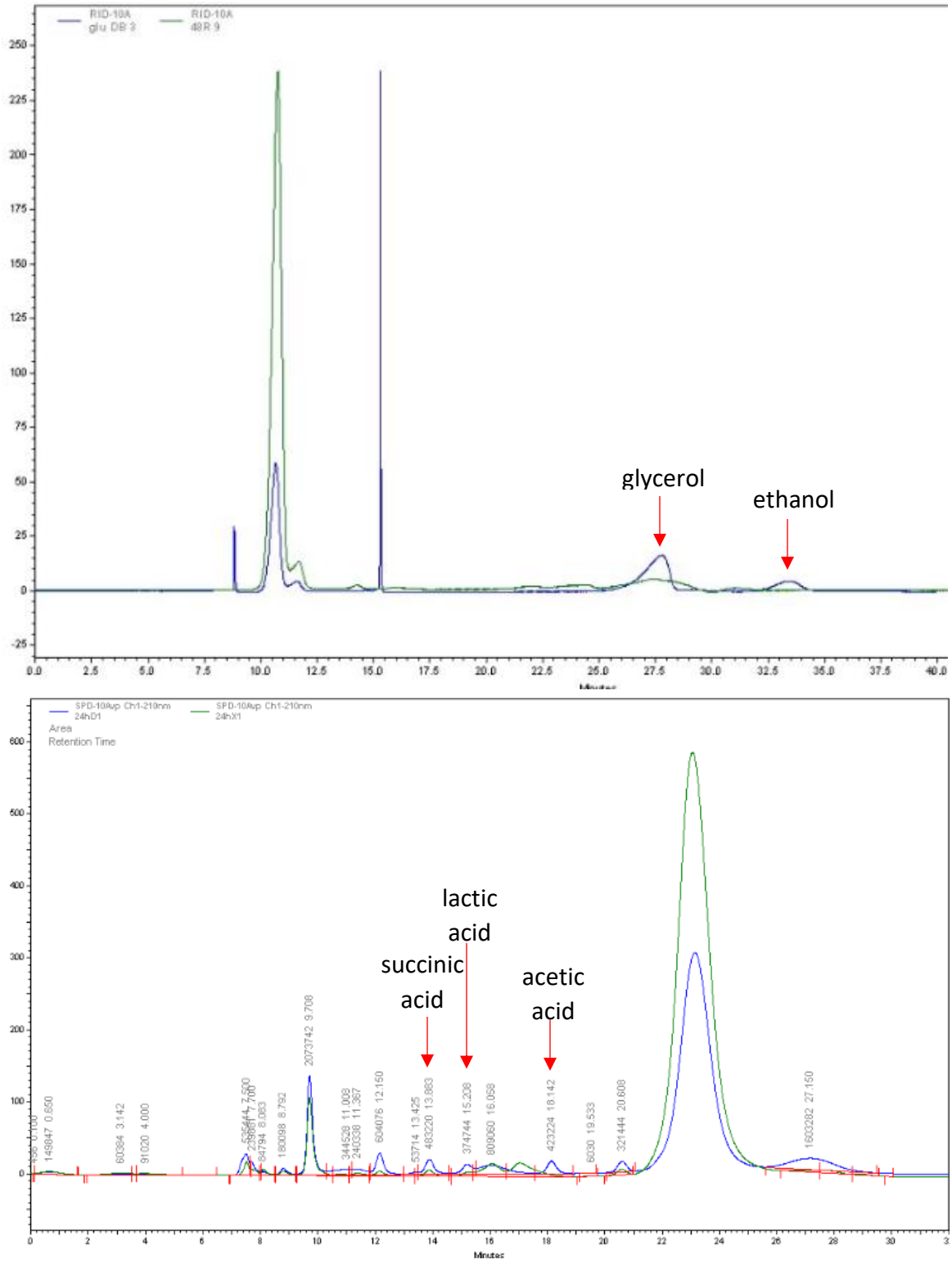


Figure 6 A) Chromatogram of the supernatant after 48h cultivation of CBS712Δ*ura3* +pKD-P2PS in SXC(A) (green) and SDC(A) (blue) medium using an HPX-42A column and RID detector. Peaks at 10.5 – 11.5 are medium components. **B)** Chromatogram of the supernatant after 24h cultivation of CBS712Δ*ura3* +pKD-P2PS in SXC(A) (green) and SDC(A) (blue) using an HPX-87H column and UV-Vis detector. The broad peak at 23 min includes TAL as well as other products.

In *K. marxianus*, high-affinity sugar transport is known to be symport and introduce one H+ into the cytosol for every sugar molecule (Van Leeuwen et al. 1991); however, xylose transporters may be lower affinity than those for glucose, resulting in relatively slower uptake and growth (Stambuk et al. 2003). The minimal byproduct formation observed in xylose may contribute to the higher levels of TAL produced relative to glucose; however, these byproducts do not account for the significant difference in titers. Transcription analyses for *K. marxianus* report greater than 100-fold higher transcription of fatty acid and lipid catabolism genes in xylose relative to glucose as well as increased conversion of ethanol via the TCA cycle (Lertwattanasakul et al. 2015; Schabort et al. 2016); these may contribute to higher acetyl-CoA for conversion to TAL. While the full metabolism of *K. marxianus* in glucose and xylose is still not fully understood, important factors such as lipid degradation and accumulation, byproduct formation, cofactor balances, sugar transport and growth rate as well as putative phosphoketolase activity (Evans and Ratledge 1984) may contribute to the improved growth and titers from xylose.

TAL was also found to be more detrimental to the growth of *K. marxianus* in xylose (**Table 4**) than to *S. cerevisiae* in glucose (Cardenas and Da Silva 2014), influencing final cell density at titers of >1g/L. A reduction in growth rate was also observed, and was similar that for *Y. lipolytica* in glucose (Markham et al. 2018). For both *Yarrowia* and *E. coli*, TAL toxicity did not prevent the late-phase accumulation of high TAL titers (Tang et al. 2013; Markham et al. 2018).

TAL toxicity to *K. marxianus*

To evaluate the effect of TAL concentration on the growth *K. marxianus,* strain CBS6556 was cultivated for 48 h at 37°C in 3mL of 1% SXC(A) media supplemented with uracil (100 mg/L)

and solid TAL at a concentration of 0, 1, 3, or 5 g/L (with pH adjusted to 5 by addition of HCl or NaOH). Optical density (600 nm) measurements were used to determine maximum specific growth rate and the cell growth at 48h (**Table 4**). The growth rate of CBS6556 dropped with increasing TAL concentration, indicating TAL toxicity.

A similar, less exaggerated trend was previously observed in *Saccharomyces cerevisiae* grown in selective media containing TAL (Cardenas and Da Silva 2014). It was hypothesized that increasing concentrations of the protonated form of TAL at lower pH values may have detrimental effects on the membrane of the yeast cells, resulting in reduced growth rate. Despite *K. marxianus* exhibiting less resistance to TAL than *S. cerevisiae*, high concentrations of TAL are not expected until late in cultivation when the optical density of the culture is already high. In *Yarrowia lipolytica*, similar toxicity did not prevent the accumulation of high TAL titers.

Table 4 TAL toxicity for strain CBS6556Δ*ura3* at increasing concentrations of TAL in 1% SXC(A) medium in 3mL, 37°C. Values are reported ± standard deviation (n=3 biological replicates).

Concentration of TAL in 1% SXC(A) (g/L)	Maximum Specific Growth Rate (h^{-1})	Optical Density (600 nm) at 48 h
0	0.42 ± 0.02	17.2 ± 0.7
1	0.22 ± 0.02	16.9 ± 1.0
3	0.12 ± 0.01	12.1 ± 0.4
5	0.082 ± 0.002	7.1 ± 0.3

Table 5 Maximum specific growth rates and final optical density (OD_{600}) for *K. marxianus* strains CBS712*Δura3* and CBS6556*Δura3* cultivated in YP media over a range of temperatures. μ_{MAX} is determined from a minimum of 4 points during exponential growth, where optical density is measured hourly.

		CBS6556Δ*ura3*				**CBS712Δ*ura3***			
		30°C	37 °C	41 °C	43 °C	30 °C	37 °C	41 °C	43 °C
μ_{MAX}	GLUCOSE	0.78 ± 0.07	0.73 ± 0.03	0.65 ± 0.02	0.56 ± 0.04	0.52 ± 0.09	0.67 ± 0.10	0.66 ± 0.02	0.58 ± 0.02
	GLYCEROL	0.56 ± 0.16	0.68 ± 0.08	0.47 ± 0.03	0.26 ± 0.01	0.55 ± 0.06	0.39 ± 0.06	0.40 ± 0.02	0.32 ± 0.01
	XYLOSE	0.39 ± 0.05	0.40 ± 0.03	0.40 ± 0.03	0.12 ± 0.01	0.31 ± 0.08	0.28 ± 0.08	0.28 ± 0.05	0.28 ± 0.05
OD_{600}	GLUCOSE	37.6 ± 5.5	26.1 ± 1.5	20.0 ± 2.3	15.3 ± 1.0	39.8 ± 3.4	22.9 ± 1.5	14.6 ± 0.8	14.7 ± 1.0
	GLYCEROL	23.6 ± 0.5	30.8 ± 1.2	26.1 ± 0.8	18.4 ± 0.2	17.8 ± 0.8	11.6 ± 0.7	9.7 ± 0.25	4.7 ± 0.3
	XYLOSE	38.6 ± 2.4	33.2 ± 3.3	19.6 ± 1.7	15.5 ± 0.3	43.7 ± 3.3	44.6 ± 3.2	34.9 ± 1.6	26.4 ± 0.6

Values are listed as a mean ± standard deviation (n=3 biological replicates)

Temperature effects on growth and TAL production

The thermotolerance of *K. marxianus* is beneficial in an industrial setting to reduce cooling costs and contamination. We thus evaluated the effect of cultivation temperature on growth rate, final biomass, and TAL titer. The growth of sequenced strains CBS6556Δ*ura3* and CBS712Δ*ura3* at 30, 37, 41, and 43°C was first evaluated in rich YP medium containing glucose, glycerol, or xylose. Growth rate in glucose medium remained high despite temperature increases; however there was a marked decrease in final cell density with increasing temperatures (**Table 5**). In CBS6556Δ*ura3*, there is a large reduction in growth rate in xylose (by 70%) as temperature increases. Despite slower growth overall, temperature had little to no detriment to growth rate for CBS712Δ*ura3* when grown on xylose and glucose. In all cases, final OD_{600} was reduced

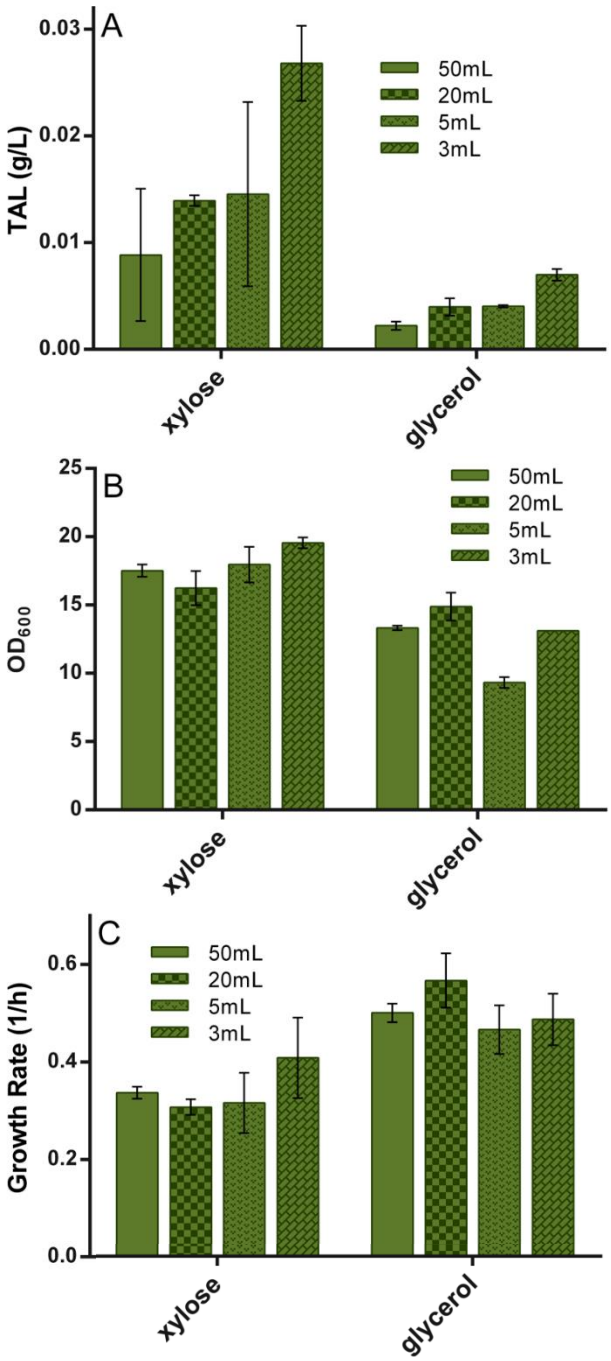


Figure 7 Effect of culture volume and flask/tube on growth and TAL production. Strain CBS6556Δ*ura3*+pCA-P2PS was cultivated in 1% xylose or 1% glycerol SC medium at 37oC for 48h in either baffled flasks (50mL and 30mL) or test tubes (5mL and 3mL). **A)** TAL production. **B)** Final cell density. **C)** Growth rate. Values are reported as mean ± standard deviation (n=3 biological replicates).

as temperature increased. Interestingly, cell density in xylose matched or exceeded that of glucose and glycerol. The combination of CBS712Δ*ura3* and xylose resulted in the lowest decrease in final density at 43°C as well as the highest overall cell density observed at this temperature.

For further studies, we used *K. marxianus* CBS712Δ*ura3* in xylose medium because the CBS712 sequence is available, and the strain demonstrates high growth and TAL production on this carbon source (**Figure 5B, Table 5**). Strain CBS712Δ*ura3* transformed with plasmid pCA-P2PS was cultivated in 3 mL tube cultures in selective, defined SXC medium at 30, 37, 41, and 43°C. The 2-PS gene is known to be stable at these temperatures with a melting temperature of 60.7°C (Vickery et al. 2018a). We found that using 3mL medium in tubes gave comparable growth and TAL production as baffled flasks (**Figure 7**). Both TAL titer and specific TAL production were similar at 37, 41, and 43°C, and 2-3-fold higher than at 30°C (**Figure 8**). It is worth noting that, unlike in rich medium, final OD does not decrease with increasing temperature in selective SXC (**Table 6**). Consistent production of TAL over a range of temperatures offers significant flexibility for industrial processes.

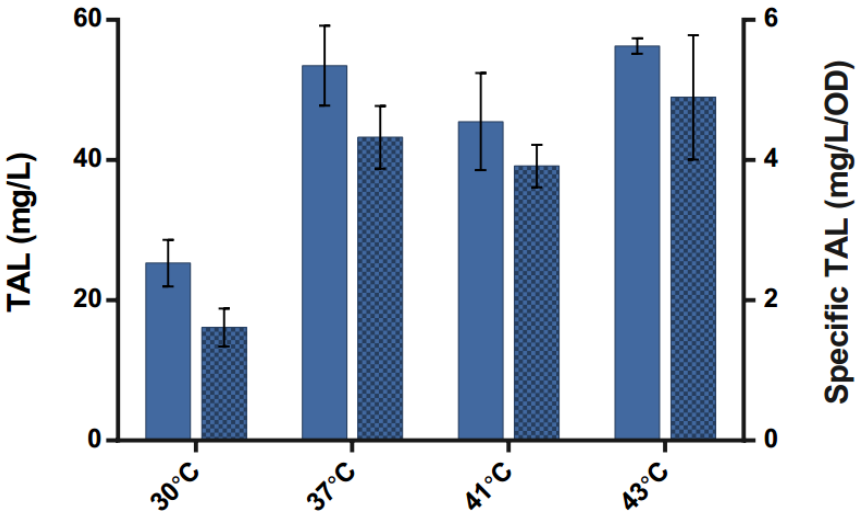


Figure 8 TAL production of strain CBS712Δ*ura3*+pCA-P2PS after 48h of culture in SXC medium over a range of temperatures. TAL titer: solid bars. Specific TAL: hatched bars. Bars represent mean ± standard deviation (n=3 biologial replicates).

Table 6 Final optical density (OD600) of CBS712Δ*ura3*+pCA-P2PS cultivated in SXC media over a range of temperatures.

	CBS712Δ*ura3*			
	30 °C	37 °C	41 °C	43 °C
OD_{600}	16.2 ± 0.33	10.5 ± 0.35	9.76 ± 0.84	14.9 ± 0.49

Values are listed as a mean ± standard deviation (n=3 biological replicates)

Improvements in expression system for TAL synthesis

To further increase TAL expression in *K. marxianus* CBS712Δ*ura3*, we evaluated the impact of a variant *g2ps1* gene and promoter strength. In our previous work on TAL synthesis in *S. cerevisiae*, we developed 2-PS mutants with higher *in vitro* and *in vivo* stability, enabling increased enzyme concentration and TAL synthesis (Vickery et al. 2018a). We replaced 2-PS with one of these mutants, 2PS[C35S], resulting in pCA-P[C35S], and measured TAL levels in SC medium supplemented with glucose, glycerol or xylose. As observed previously, TAL production was negligible in the glucose medium. However, in both glycerol and xylose, titers increased by 80% and 36%, respectively, for strains harboring 2-PS[C35S] relative to the wildtype 2-PS enzyme (**Figure 9**). This demonstrates the potential for further application of successful strategies developed for *S. cerevisiae* in *K. marxianus*.

Our initial studies used the *K. marxianus PGK1* promoter (P_{KmPGK1}) to control expression of the 2-PS gene on the *Km*CEN/ARS vector, resulting in yields of ~0.01 mol TAL/mol carbon – in the same range as initial work in *S. cerevisiae* (**Table 7**). We replaced this promoter with the *S. cerevisiae ADH2* promoter (P_{ScADH2}) that we have used successfully for TAL expression in *S. cerevisiae*. Strain CBS712Δ*ura3* was transformed with pCA-P2PS or pCA-A2PS and cultivated for

48 h in SXC medium. A 2.2-fold improvement in TAL levels was observed when the alternate promoter P_{ScADH2} was employed (**Figure 9**). We then substituted the 2-PS variant (2-PS[C35S]) under this new promoter and repeated the experiment. Use of 2-PS[C35S] increased TAL titer by 57% relative to the wildtype 2-PS, and the combination of the more stable 2-PS with the stronger P_{ScADH2} increased TAL levels TAL titer and yield by 3.3-fold relative to our initial system with the wildtype 2-PS and P_{KmPGK1}.

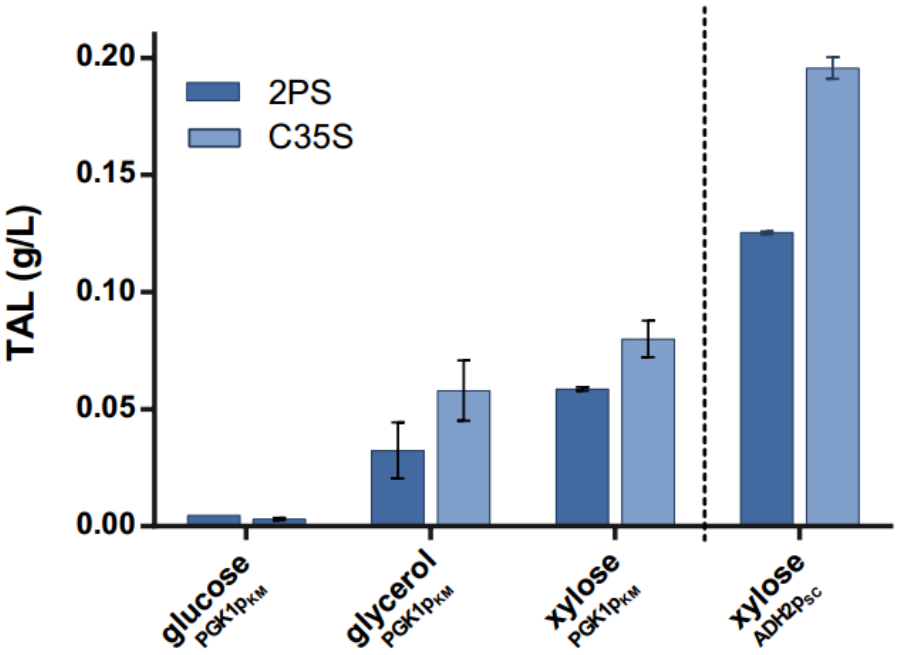


Figure 9 Comparison of 2-PS enzymes and promoter strength on TAL titer. Strain CBS712Δ*ura3* transformed with pCA-P2PS or pCA-P[C35S] was cultivated at 37°C for 48h in SC medium with three different carbon sources. CBS712Δ*ura3* was also transformed with pCA-A2PS or pCA-A[C35S] (in both vectors, the *ADH2* promoter replaces the *PGK1* promoter) and cultivated at 37°C for 48h in SXC medium. Bars represent mean ± standard deviation (n=3 biological replicates).

Table 7 TAL titers and yields for three yeast species.

Strain	Plasmid or Integrants	Medium	Sampling time (h)	Titer (g/L)	Yield (mol TAL/ mol sugar)	Yield (mol TAL/ mol C)
S. cerevisiae						
BY4741[1]	2u-PGK1p-2PS	YPD (1% glucose)	48	0.058	0.008	0.00138
BY4741[1]	2u-ADH2p-2PS	YPD (1% glucose)	48	0.225	0.032	0.00536
BYeng[1,4]	2u-ADH2p-2PS pJCT-PDHm	YPD (1% glucose)	48	1.6	0.229	0.03810
K. marxianus						
CBS712 (this study)	pCA-P2PS	SXC (1% xylose)	48	0.06	0.007	0.00143
CBS712 (this study)	pCA-A2PS[C35S]	SXC (1% xylose)	48	0.2	0.024	0.00476
CBS712 (this study)	pKD-P2PS	SXC (1% xylose)	48	0.89	0.106	0.02120
CBS712 (this study)	pKD-A2PS	SXC (1% xylose)	48	1.07	0.127	0.02550
KM1 (this study)	pKD-A2PS	SXC (1% xylose)	48	1.24	0.148	0.02950
Yarrowia lipolytica						
yJY2039[2,5]	3 Integrations	Minimal (2% glucose)	96	0.6	0.043	0.00714
YT0[3,6]	2 Integrations	CSM (2% glucose)	96	1.1	0.079	0.01310
YT[3,6]	4 Integrations	CSM (2% glucose)	96	1.51	0.108	0.01800
YT-PDH[3,7]	4 Integrations	CSM (2% glucose)	96	2.8	0.200	0.03330

[1]Cardenas and Da Silva, 2016
[2]Yu et al., 2018
[3]Markham et al., 2018
[4]BYeng: BY4741*Δtrp1Δpor2Δmpc2Δpda1Δyat2Δzwf1*
[5]yJY2039: ATTC 20460ura3leu2 and 3 integrations of 2-PS
[6]PO1f with 2 (YT0) or 4 (YT) integrations of 2-PS
[7]YT-PDH: YT- *ACS1, ALD5, PDC2, ACC1*

Interestingly, both P_{KmPGK1} and P_{ScADH2} resulted in similar TAL synthesis profiles, with P_{ScADH2} producing consistently higher titers throughout the batch culture (**Figure 10**).

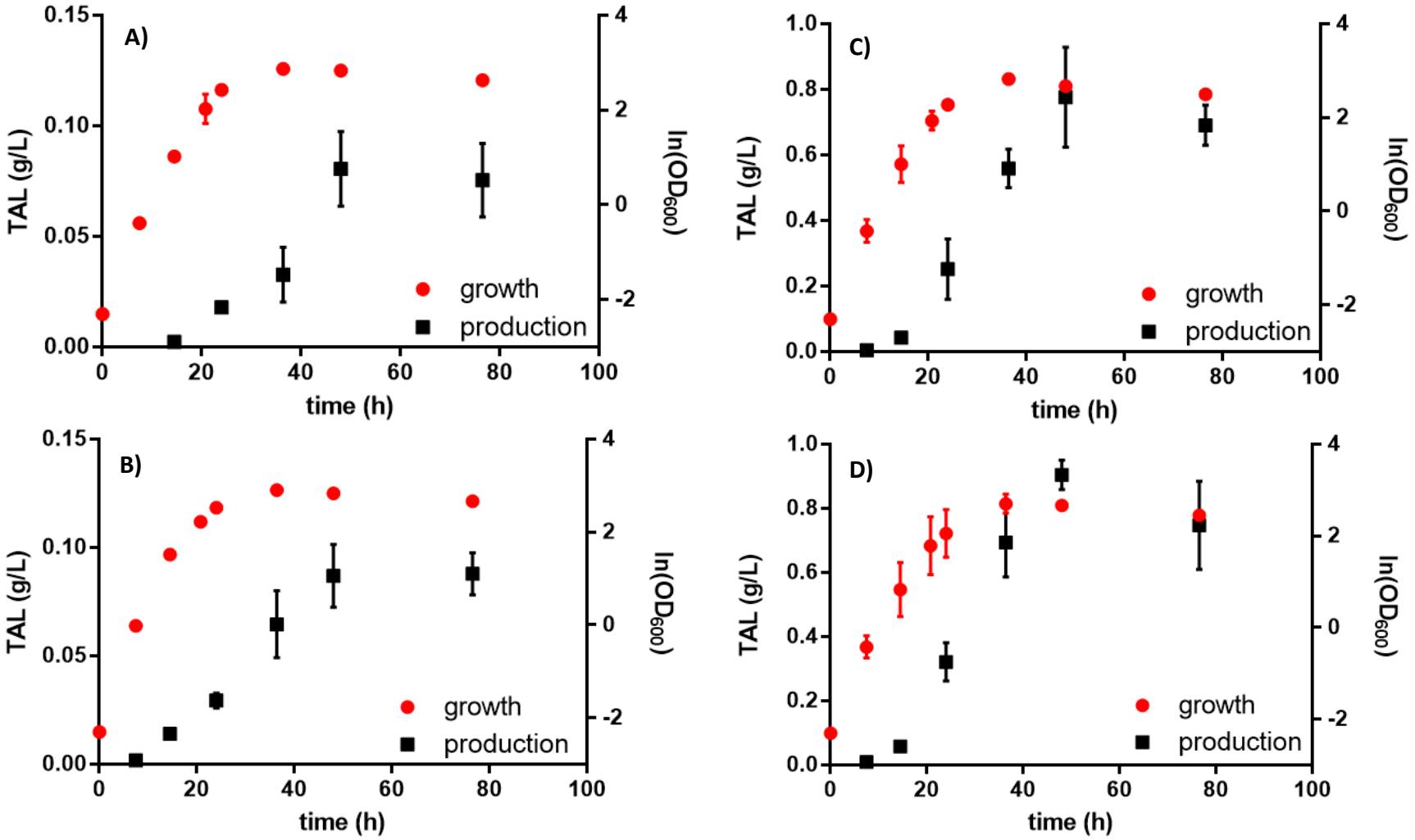


Figure 10 Growth and TAL production in 1% SXC(A) medium at 37 °C as a function of time for strain CBS712Δ*ura3* carrying plasmid **A)** pCA-P2PS **B)** pCA-A2PS **C)** pKD-P2PS or **D)** pKD-A2PS. Values are reported as mean ± standard deviation (n≥2 biological replicates).

In our studies, TAL was detected as early as 5h with this promoter; however, TAL does not approach 1 g/L until stationary phase. This indicates that despite the toxicity effects of TAL (**Table 4**), the TAL produced should have minor effects on final cell density and titer. The similar TAL synthesis profile observed for P_{KmPGK1} and P_{ScADH2} suggests that regulation may be similar for these promoters in this yeast. The native *ADH2* promoter has not been characterized in *K. marxianus.* Of the six constitutive promoters (*PGK*, *ADH1*, *TDH2* promoters from *K. marxianus* or *S. cerevisiae*) previously compared in *K. marxianus*, P_{KmPGK1} was the strongest during growth on all three carbon sources (xylose, glucose, and glycerol) tested, and the promoter that maintained the highest expression levels at elevated temperatures (Yang et al. 2015). It will be interesting to test the native *K. marxianus ADH2* promoter to determine the native regulation. In *K. marxianus*, the native promoters have generally been found to be stronger than the *S. cerevisiae* versions (for the three promoters compared) (Yang et al. 2015). Replacing P_{ScADH2} with the native P_{KmADH2} may further improve polyketide levels.

Increases in gene copy number to improve TAL titer

All of our initial studies were performed using a *K. marxianus* CEN/ARS vector maintained at high stability but low copy number. Increasing copy number in *K. marxianus* (via a higher copy plasmid or gene integration) is necessary for higher TAL titers. Although no multi-copy plasmids are known for *K. marxianus*, pKD1-based plasmids from *K. lactis* have been successfully used in *K. marxianus* (De Souza and De Morais 2000; Duan et al. 2018). For our studies, we replaced the *Km*CEN/ARS sequence on pCA-P2PS with the full pKD1 sequence (linearized at the native SphI

site to improve stability (Hsieh and Da Silva 1998; Bartkeviciute, Siekstele, and Sasnauskas 2000)), resulting in plasmid pKD-P2PS.

Strains CBS6556Δ*ura3* and CBS712Δ*ura3* were transformed with this new plasmid and cultivated in SXC medium for 48 h. Use of this higher copy plasmid (pKD-P2PS) increased TAL titers and yield by greater than 14.8-fold (**Figure 11, Table 7**) relative to the *Km*CEN/ARS based vector (pCA-P2PS). TAL titers of up to 0.89 g/L were observed from xylose in 3 mL tube cultures using an otherwise unengineered strain. This yield is also an improvement of 15-fold relative to initial studies of *S. cerevisiae* (with a similar expression system, growing on glucose) prior to metabolic engineering (**Table 7**). Production was highest at 37°C (**Figure 12**); in contrast, titers were similar at 37°C, 41°C, and 43°C with the *Km*CEN/ARS plasmid (**Figure 8**).

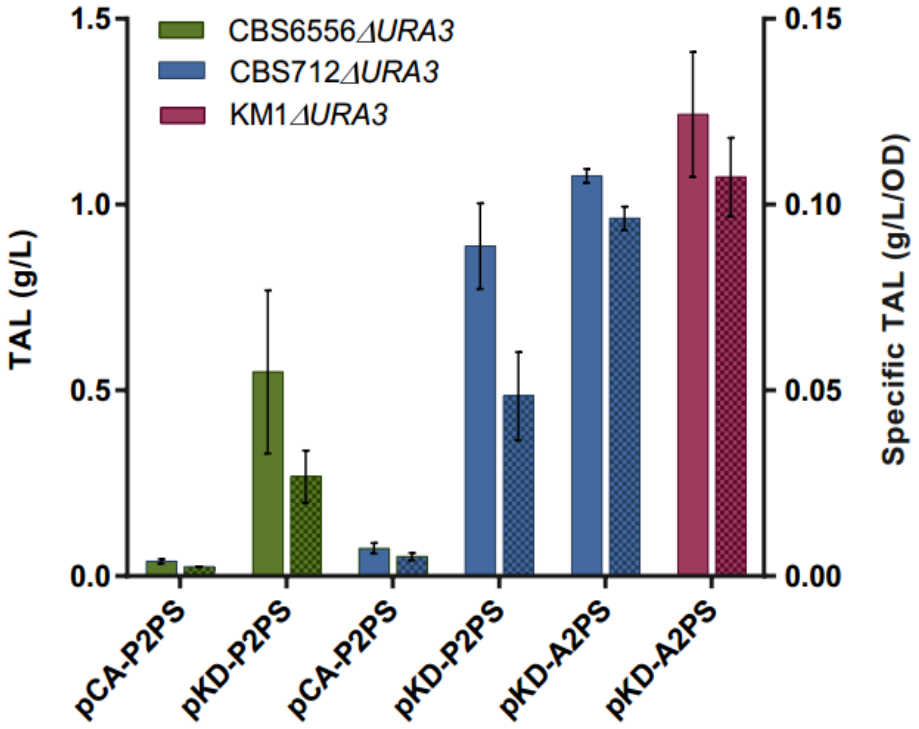


Figure 11 Comparison of expression with low and high copy plasmids. Strains CBS6656Δ*ura3*, CBS712Δ*ura3*, and KMΔ*ura3* were transformed with pCA-P2PS, pKD-P2PS, or pKD-A2PS and cultivated at 37°C for 48h in 3 mL SXC medium. TAL titer: solid bars. Specific TAL: hatched bars. Bars represent mean ± standard deviation (n≥3 biological replicates).

The use of P_{ScADH2} in the CEN/ARS plasmid increased TAL levels; therefore, we replaced P_{KmPGK1} with P_{ScADH2} in the multi-copy pKD plasmid (pKD-A2PS) and repeated our experiment. In CBS712Δ*ura3*, we observed a 20% improvement in TAL titer, and a 2-fold improvement in specific titer when the *ADH2* promoter was used. Furthermore, in our most robust industrial strain KM1Δ*ura3*, we reached titers of 1.24 g/L with a yield of 0.030 mol TAL/mol C from xylose in 3 mL culture. This is an excellent titer and yield relative to other yeast hosts; prior to metabolic engineering, the yield for *K. marxianus* on xylose approached those for the best engineered *S. cerevisiae* and *Y. lipolytica* (**Table 7**) strains on glucose.

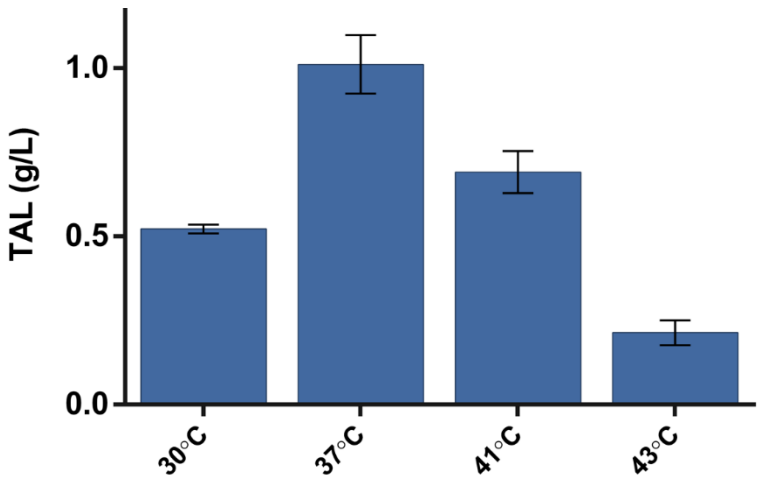


Figure 12 TAL titers of CBS712Δ*ura3* with pKD-P2PS after 48h cultivation in 3mL at 30, 37, 41 and 43°C in SXC. Values are reported as mean ± standard deviation (n=3 biological replicates).

Given the ability of *K. marxianus* to grow on very minimal medium, we compared growth and TAL production for CBS712Δ*ura3*+pKD-P2PS in selective SXC and SX media at 37°C. In the very minimal SX medium, TAL titer was 30% lower and specific titer (g/L/OD) only 25% lower than in SXC (that contains casamino acids) (**Figure 13**). We measured the plasmid stability at various times during cultivation in SXC medium; for both strains, plasmid stability was greater than 80% after 48 hours (**Table 8**). Surprisingly, stability was also high in complex medium (55% in

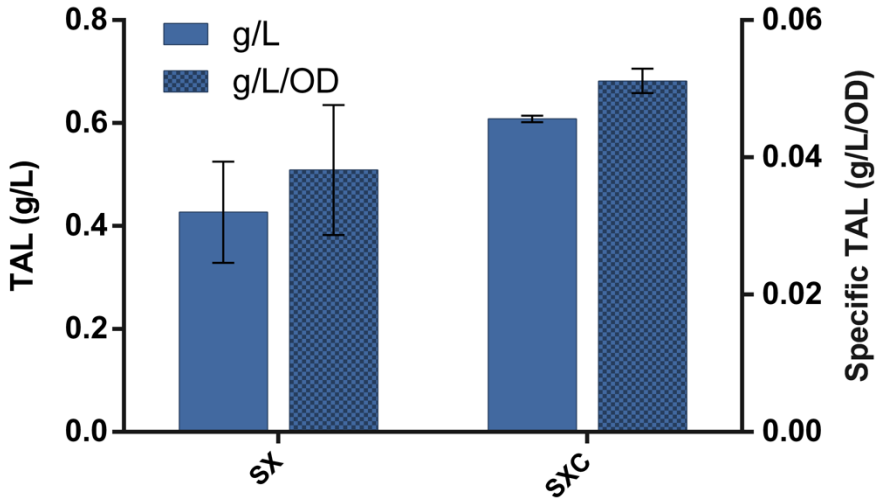


Figure 13 TAL titers after 48h of cultivation of strain CBS712Δ*ura3* +pKD-P2PS in xylose S minimal medium compared to xylose SC defined medium. Values are reported as mean ± standard deviation (n=3 biological replicates).

CBS712*Δura3* and 87% in CBS6556*Δura3*). The substantially higher TAL titers and the high plasmid stability demonstrates the promise of pKD1-based plasmids for polyketide production. In addition, the ability of *K. marxianus* to produce these titers in minimal xylose medium is advantageous for industrial production and downstream separation.

Table 8 Plasmid stability of pKD-P2PS in *K. marxianus* strains CBS6556Δ*ura3* and CBS712Δ*ura3* in selective SXC medium compared to rich YPX medium at early exponential (OD=4), late exponential (OD=10) and stationary (48h) phases. Average values ± standard deviation (n=300 colonies).

	CBS6556		**CBS712**	
	SXC	**YPX**	**SXC**	**YPX**
OD=4	1 ± 0	0.97 ± 0.019	0.88 ± 0.042	0.57 ± 0.048
OD=10	0.94 ± 0.059	0.96 ± 0.024	0.85 ± 0.061	0.64 ± 0.026
48h	0.90 ± 0.015	0.87 ± 0.046	0.81 ± 0.057	0.55 ± 0.042

Conclusions

This study is the first to evaluate polyketide production in *K. marxianus*, achieving excellent titers of 1.24 g/L TAL in test tube culture prior to metabolic engineering or bioreactor cultivation. Successful transfer of strategies effective in *S. cerevisiae* such as enzyme and strain engineering can be further explored to optimize *K. marxianus* as an industrial workhorse. Sustained production at elevated temperatures of 41 and 43°C, in true minimal medium, and on a variety of substrates, in particular xylose, demonstrates the promise of this rapidly growing, thermotolerant yeast species for sustainable and low-cost production of acetyl-CoA based polyketides.

Acknowledgements

This research was supported by the National Science Foundation: Grant No. CBET-1803677 and EEC-0813570 (through the Engineering Research Center CBiRC (Center for Biorenewable Chemicals)) and the Department of Energy/Office of Science: Grant No. DE-SC0019093. The authors thank Prof. Ian Wheeldon (University of California, Riverside) for strain CBS6556*Δura3*, Cesar Villavicencio for constructing pCA-P2PS, Tej Shah for performing introductory growth studies, and Ruben Fernandez-Moya for his helpful input and advice.

References

Amrane, A., and Y. Prigent. 1998. “Effect of Culture Conditions of Kluyveromyces Marxianus on Its Autolysis, and Process Optimization.” Bioprocess Engineering 18 (5): 383–88. https://doi.org/10.1007/s004490050460.

Ball, M M, A Raynal, M Guerineau, and F Iborra. 1999. “Construction of Efficient Centromeric, Multicopy and Expression Vectors for the Yeast Kluyveromyces Marxianus Using Homologous Elements and the Promoter of a Purine-Cytosine-like Permease.” Journal of Molecular Microbiology and Biotechnology.

Ballesteros, M., J. M. Oliva, M. J. Negro, P. Manzanares, and I. Ballesteros. 2004. “Ethanol from Lignocellulosic Materials by a Simultaneous Saccharification and Fermentation Process (SFS) with Kluyveromyces Marxianus CECT 10875.” Process Biochemistry. https://doi.org/10.1016/j.procbio.2003.09.011.

Banat, I. M., P. Nigam, and R. Marchant. 1992. “Isolation of Thermotolerant, Fermentative Yeasts Growing at 52-Degree C and Producing Ethanol at 45- Degree C and 50-Degree C.” World Journal of Microbiology & Biotechnology 8 (3): 259–63. https://doi.org/10.1007/BF01201874.

Bartkeviciute, Danguole, Rimantas Siekstele, and Kęstutis Sasnauskas. 2000. “Heterologous Expression of the Kluyveromyces Marxianus Endopolygalacturonase Gene (EPG1) Using

Versatile Autonomously Replicating Vector for a Wide Range of Host." In Enzyme and Microbial Technology. https://doi.org/10.1016/S0141-0229(00)00155-1.

Blank, Lars M., Frank Lehmbeck, and Uwe Sauer. 2005. "Metabolic-Flux and Network Analysis in Fourteen Hemiascomycetous Yeasts." In FEMS Yeast Research. https://doi.org/10.1016/j.femsyr.2004.09.008.

Bruijne, Adriaan W. De, Johanna Schuddemat, Peter J.A. Van den Broek, and Johnny Van Steveninck. 1988. "Regulation of Sugar Transport Systems of Kluyveromyces Marxianus: The Role of Carbohydrates and Their Catabolism." BBA - Biomembranes. https://doi.org/10.1016/0005-2736(88)90104-6.

Cai, X. P., J. Zhang, H. Y. Yuan, Z. A. Fang, and Y. Y. Li. 2005. "Secretory Expression of Heterologous Protein in Kluyveromyces Cicerisporus." Applied Microbiology and Biotechnology. https://doi.org/10.1007/s00253-004-1834-9.

Cardenas, Javier, and Nancy A. Da Silva. 2014. "Metabolic Engineering of Saccharomyces Cerevisiae for the Production of Triacetic Acid Lactone." Metabolic Engineering 25: 194–203. https://doi.org/10.1016/j.ymben.2014.07.008.

Cardenas, Javier, and Nancy A. Da Silva. 2016. "Engineering Cofactor and Transport Mechanisms in Saccharomyces Cerevisiae for Enhanced Acetyl-CoA and Polyketide Biosynthesis." Metabolic Engineering 36: 80–89. https://doi.org/10.1016/j.ymben.2016.02.009.

Chia, Mei, Thomas J. Schwartz, Brent H. Shanks, and James A. Dumesic. 2012. “Triacetic Acid Lactone as a Potential Biorenewable Platform Chemical.” Green Chemistry 14 (7): 1850. https://doi.org/10.1039/c2gc35343a.

Duan, Jinkun, Deqiang Yang, Lei Chen, Yao Yu, Jungang Zhou, and Hong Lu. 2018. “Efficient Production of Porcine Circovirus Virus-like Particles Using the Nonconventional Yeast Kluyveromyces Marxianus” 2016.

Espinoza, Patricia, Eduardo Bárzana, Mariano García-Garibay, and Lorena Ǵmez-Ruiz. 1992. “Evaluation of Kluyveromyces Marxianus for the Production of Lactase Simultaneously to Pectinase or Inulinase.” Biotechnology Letters. https://doi.org/10.1007/BF01021058.

Evans, Christopher Thomas, and Colin Ratledge. 1984. “Induction of Xylulose-5-Phosphate Phosphoketolase in a Variety of Yeasts Grown on d-Xylose: The Key to Efficient Xylose Metabolism.” Archives of Microbiology 139 (1): 48–52. https://doi.org/10.1007/BF00692711.

Falcone, C., M. Saliola, X. J. Chen, L. Frontali, and H. Fukuhara. 1986. “Analysis of a 1.6-Mm Circular Plasmid from the Yeast Kluyveromyces Drosophilarum: Structure and Molecular Dimorphism.” Plasmid. https://doi.org/10.1016/0147-619X(86)90044-2.

FDA. 2015. “Microorganisms & Microbial-Derived Ingredients Used in Food (Partial List).” Food And Drug Administration. 2015. http://www.fda.gov/Food/IngredientsPackagingLabeling/GRAS/MicroorganismsMicrobialDerivedIngredients/default.htm.

Fonseca, Gustavo Graciano, Nuno Miguel Barbosa de Carvalho, and Andreas Karoly Gombert. 2013. “Growth of the Yeast Kluyveromyces Marxianus CBS 6556 on Different Sugar Combinations as Sole Carbon and Energy Source.” Applied Microbiology and Biotechnology. https://doi.org/10.1007/s00253-013-4748-6.

Fonseca, Gustavo Graciano, Andreas Karoly Gombert, Elmar Heinzle, and Christoph Wittmann. 2007. “Physiology of the Yeast Kluyveromyces Marxianus during Batch and Chemostat Cultures with Glucose as the Sole Carbon Source.” FEMS Yeast Research 7 (3): 422–35. https://doi.org/10.1111/j.1567-1364.2006.00192.x.

Galindo-Leva, Luz Ángela, Stephen R. Hughes, Juan Carlos López-Núñez, Joshua M. Jarodsky, Adam Erickson, Mitchell R. Lindquist, Elby J. Cox, et al. 2016. “Growth, Ethanol Production, and Inulinase Activity on Various Inulin Substrates by Mutant Kluyveromyces Marxianus Strains NRRL Y-50798 and NRRL Y-50799.” Journal of Industrial Microbiology and Biotechnology. https://doi.org/10.1007/s10295-016-1771-5.

Gao, Fang, and Andrew J. Daugulis. 2009. “Bioproduction of the Aroma Compound 2-Phenylethanol in a Solid-Liquid Two-Phase Partitioning Bioreactor System by Kluyveromyces Marxianus.” Biotechnology and Bioengineering. https://doi.org/10.1002/bit.22387.

Gietz, Daniel, Andrew St Jean, Robin A. Woods, and Robert H. Schiestl. 1992. “Improved Method for High Efficiency Transformation of Intact Yeast Cells.” Nucleic Acids Research 20 (6): 1425.

Groeneveld, Philip, Adriaan H. Stouthamer, and Hans V. Westerhoff. 2009. “Super Life - How and Why ‘cell Selection’ Leads to the Fastest-Growing Eukaryote.” FEBS Journal 276 (1): 254–70. https://doi.org/10.1111/j.1742-4658.2008.06778.x.

Hashimoto, Makoto, Takamasa Nonaka, and Isao Fujii. 2014. “Fungal Type III Polyketide Synthases.” Nat. Prod. Rep. 31 (10): 1306–17. https://doi.org/10.1039/C4NP00096J.

Hsieh, H. P., and N. A. Da Silva. 1998. “Partial-PKD1 Plasmids Provide Enhanced Structural Stability for Heterologous Protein Production in Kluyveromyces Lactis.” Applied Microbiology and Biotechnology 49 (4): 411–16. https://doi.org/10.1007/s002530051191.

Iborra, François, and Maria M. Ball. 1994. “Kluyveromyces Marxianus Small DNA Fragments Contain Both Autonomous Replicative and Centromeric Elements That Also Function in Kluyveromyces Lactis.” Yeast. https://doi.org/10.1002/yea.320101211.

Juergens, Hannes, Javier A Varela, Arthur R Gorter de Vries, Thomas Perli, Veronica J M Gast, Nikola Y Gyurchev, Arun S Rajkumar, et al. 2018. “Genome Editing in Kluyveromyces and Ogataea Yeasts Using a Broad-Host-Range Cas9/GRNA Co-Expression Plasmid.” FEMS Yeast Research, no. February: 1–16. https://doi.org/10.1093/femsyr/foy012.

Kealey, James T, Lu Liu, Daniel V Santi, Mary C Betlach, and Philip J Barr. 1998. “Production of a Polyketide Natural Product in Nonpolyketide-Producing Prokaryotic and Eukaryotic Hosts.” Proceedings of the National Academy of Sciences of the United States of America. https://doi.org/10.1073/pnas.95.2.505.

Kim, Jin Seong, Jae Bum Park, Seung Won Jang, and Suk Jin Ha. 2015. “Enhanced Xylitol Production by Mutant Kluyveromyces Marxianus 36907-FMEL1 Due to Improved Xylose Reductase Activity.” Applied Biochemistry and Biotechnology 176 (7): 1975–84. https://doi.org/10.1007/s12010-015-1694-z.

Knoshaug, Eric P., Virve Vidgren, Frederico Magalhães, Eric E. Jarvis, Mary Ann Franden, Min Zhang, and Arjun Singh. 2015. “Novel Transporters from Kluyveromyces Marxianus and Pichia Guilliermondii Expressed in Saccharomyces Cerevisiae Enable Growth on L-Arabinose and d-Xylose.” Yeast. https://doi.org/10.1002/yea.3084.

Kraus, George A., Umayangani K. Wanninayake, and Jashaun Bottoms. 2016. “Triacetic Acid Lactone as a Common Intermediate for the Synthesis of 4-Hydroxy-2-Pyridones and 4-Amino-2-Pyrones.” Tetrahedron Letters. https://doi.org/10.1016/j.tetlet.2016.02.043.

Kushi, R. T., R. Monti, and J. Contiero. 2000. “Production, Purification and Characterization of an Extracellular Inulinase from Kluyveromyces Marxianus Var. Bulgaricus.” Journal of Industrial Microbiology and Biotechnology. https://doi.org/10.1038/sj.jim.7000032.

Kwak, Suryang, and Yong Su Jin. 2017. “Production of Fuels and Chemicals from Xylose by Engineered Saccharomyces Cerevisiae: A Review and Perspective.” Microbial Cell Factories. https://doi.org/10.1186/s12934-017-0694-9.

Lane, Melanie M., and John P. Morrissey. 2010. “Kluyveromyces Marxianus: A Yeast Emerging from Its Sister’s Shadow.” Fungal Biology Reviews. https://doi.org/10.1016/j.fbr.2010.01.001.

Lee, Ming Hsuan, Jinn Jy Lin, Yu Ju Lin, Jui Jen Chang, Huei Mien Ke, Wen Lang Fan, Tzi Yuan Wang, and Wen Hsiung Li. 2018. “Genome-Wide Prediction of CRISPR/Cas9 Targets in Kluyveromyces Marxianus and Its Application to Obtain a Stable Haploid Strain.” Scientific Reports 8 (1): 1–10. https://doi.org/10.1038/s41598-018-25366-z.

Leeuwen, C. C M Van, E. Postma, P. J A Van den Broek, and J. Van Steveninck. 1991. “Proton-Motive Force-Driven D-Galactose Transport in Plasma Membrane Vesicles from the Yeast Kluyveromyces Marxianus.” Journal of Biological Chemistry.

Lertwattanasakul, Noppon, Tomoyuki Kosaka, Akira Hosoyama, Yutaka Suzuki, Nadchanok Rodrussamee, Minenosuke Matsutani, Masayuki Murata, et al. 2015. “Genetic Basis of the Highly Efficient Yeast Kluyveromyces Marxianus: Complete Genome Sequence and Transcriptome Analyses.” Biotechnology for Biofuels. https://doi.org/10.1186/s13068-015-0227-x.

Lertwattanasakul, Noppon, Nadchanok Rodrussamee, Suprayogi, Savitree Limtong, Pornthap Thanonkeo, Tomoyuki Kosaka, and Mamoru Yamada. 2011. “Utilization Capability of Sucrose, Raffinose and Inulin and Its Less-Sensitiveness to Glucose Repression in Thermotolerant Yeast Kluyveromyces Marxianus DMKU 3-1042.” AMB Express. https://doi.org/10.1186/2191-0855-1-20.

Löbs, Ann Kathrin, Ronja Engel, Cory Schwartz, Andrew Flores, and Ian Wheeldon. 2017. “CRISPR-Cas9-Enabled Genetic Disruptions for Understanding Ethanol and Ethyl Acetate

Biosynthesis in Kluyveromyces Marxianus." Biotechnology for Biofuels 10 (1). https://doi.org/10.1186/s13068-017-0854-5.

Löser, Christian, Thanet Urit, Erik Gruner, and Thomas Bley. 2015. "Efficient Growth of Kluyveromyces Marxianus Biomass Used as a Biocatalyst in the Sustainable Production of Ethyl Acetate." Energy, Sustainability and Society 5 (1): 1–15. https://doi.org/10.1186/s13705-014-0028-2.

Ma, Suzanne M., Jesse W.H. Li, Jin W. Choi, Hui Zhou, K. K.Michael Lee, Vijayalakshmi A. Moorthie, Xinkai Xie, et al. 2009. "Complete Reconstitution of a Highly Reducing Iterative Polyketide Synthase." Science. https://doi.org/10.1126/science.1175602.

Markham, Kelly A., Claire M. Palmer, Malgorzata Chwatko, James M. Wagner, Clare Murray, Sofia Vazquez, Arvind Swaminathan, Ishani Chakravarty, Nathaniel A. Lynd, and Hal S. Alper. 2018. "Rewiring Yarrowia Lipolytica toward Triacetic Acid Lactone for Materials Generation." Proceedings of the National Academy of Sciences, 201721203. https://doi.org/10.1073/pnas.1721203115.

Martínez, Oscar, Antoni Sánchez, Xavier Font, and Raquel Barrena. 2018. "Bioproduction of 2-Phenylethanol and 2-Phenethyl Acetate by Kluyveromyces Marxianus through the Solid-State Fermentation of Sugarcane Bagasse." Applied Microbiology and Biotechnology, 1–14. https://doi.org/10.1007/s00253-018-8964-y.

Matsumoto, Izumi, Takahiro Arai, and Y U I Nishimoto. 2018. “Thermotolerant Yeast Kluyveromyces Marxianus Reveals More Tolerance to Heat Shock than the Brewery Yeast Saccharomyces Cerevisiae” 23 (3).

Morrissey, John P., Maria M.W. Etschmann, Jens Schrader, and Gustavo M. de Billerbeck. 2015. “Cell Factory Applications of the Yeast Kluyveromyces Marxianus for the Biotechnological Production of Natural Flavour and Fragrance Molecules.” Yeast. https://doi.org/10.1002/yea.3054.

Nambu-Nishida, Yumiko, Keiji Nishida, Tomohisa Hasunuma, and Akihiko Kondo. 2017. “Development of a Comprehensive Set of Tools for Genome Engineering in a Cold- A Nd Thermo-Tolerant Kluyveromyces Marxianus Yeast Strain.” Scientific Reports 7 (1). https://doi.org/10.1038/s41598-017-08356-5.

Panuwatsuk, W., and N. Da Silva. 2002. “Evaluation of PKD1-Based Plasmid Systems for Heterologous Protein Production in Kluyveromyces Lactis.” Applied Microbiology and Biotechnology 58 (2): 195–201. https://doi.org/10.1007/s002530100815.

Pecota, Douglas C., Vineet Rajgarhia, and Nancy A. Da Silva. 2007. “Sequential Gene Integration for the Engineering of Kluyveromyces Marxianus.” Journal of Biotechnology. https://doi.org/10.1016/j.jbiotec.2006.07.031.

Pentjuss, A., E. Stalidzans, J. Liepins, A. Kokina, J. Martynova, P. Zikmanis, I. Mozga, et al. 2017. “Model-Based Biotechnological Potential Analysis of Kluyveromyces Marxianus Central

Metabolism." Journal of Industrial Microbiology and Biotechnology. https://doi.org/10.1007/s10295-017-1946-8.

Pfeifer, B., and C. Khosla. 2001. "Biosynthesis of Polyketides in Heterologous Hosts." Microbiology and Molecular Biology Reviews. https://doi.org/10.1128/MMBR.65.1.106-118.2001.

Postma, E., and P. J.A. Van den Broek. 1990. "Continuous-Culture Study of the Regulation of Glucose and Fructose Transport in Kluyveromyces Marxianus CBS 6556." Journal of Bacteriology. https://doi.org/10.1128/jb.172.6.2871-2876.1990.

Rech, R., C. F. Cassini, A. Secchi, and M. A Z Ayub. 1999. "Utilization of Protein-Hydrolyzed Cheese Whey for Production of ??-Galactosidase by Kluyveromyces Marxianus." Journal of Industrial Microbiology and Biotechnology. https://doi.org/10.1038/sj.jim.2900692.

Ricci, Antonia, Ana Allende, Declan Bolton, Marianne Chemaly, Robert Davies, Rosina Girones, Konstantinos Koutsoumanis, et al. 2018. "Update of the List of QPS-recommended Biological Agents Intentionally Added to Food or Feed as Notified to EFSA 7: Suitability of Taxonomic Units Notified to EFSA until September 2017." EFSA Journal 16 (1). https://doi.org/10.2903/j.efsa.2018.5131.

Rodrussamee, Nadchanok, Noppon Lertwattanasakul, Katsushi Hirata, Suprayogi, Savitree Limtong, Tomoyuki Kosaka, and Mamoru Yamada. 2011. "Growth and Ethanol Fermentation Ability on Hexose and Pentose Sugars and Glucose Effect under Various

Conditions in Thermotolerant Yeast Kluyveromyces Marxianus." Applied Microbiology and Biotechnology. https://doi.org/10.1007/s00253-011-3218-2.

Saunders, Lauren P., Michael J. Bowman, Jeffrey A. Mertens, Nancy A. Da Silva, and Ronald E. Hector. 2015. "Triacetic Acid Lactone Production in Industrial Saccharomyces Yeast Strains." Journal of Industrial Microbiology and Biotechnology. https://doi.org/10.1007/s10295-015-1596-7.

Schabort, Du Toit W.P., Precious Letebele, Laurinda Steyn, Stephanus G. Kilian, and James C. Du Preez. 2016. "Differential RNA-Seq, Multi-Network Analysis and Metabolic Regulation Analysis of Kluyveromyces Marxianus Reveals a Compartmentalised Response to Xylose." PLoS ONE. https://doi.org/10.1371/journal.pone.0156242.

Shanks, Brent H., and Peter L. Keeling. 2017. "Bioprivileged Molecules: Creating Value from Biomass." Green Chem. 19 (14): 3177–85. https://doi.org/10.1039/C7GC00296C.

Sharma, Nilesh Kumar, Shuvashish Behera, Richa Arora, and Sachin Kumar. 2016. "Enhancement in Xylose Utilization Using Kluyveromyces Marxianus NIRE-K1 through Evolutionary Adaptation Approach." Bioprocess and Biosystems Engineering 39 (5): 835–43. https://doi.org/10.1007/s00449-016-1563-3.

Souza, C. G. De, and M. A. De Morais. 2000. "The Use of the Replicating PDblet Plasmid as a Cloning Vector with Enhanced Stability in Kluyveromyces Marxianus." Biotechnology Letters. https://doi.org/10.1023/A:1005624913974.

Spohner, Sebastian C., Vivienne Schaum, Hendrich Quitmann, and Peter Czermak. 2016. "Kluyveromyces Lactis: An Emerging Tool in Biotechnology." Journal of Biotechnology. https://doi.org/10.1016/j.jbiotec.2016.02.023.

Stambuk, Boris U., Mary Ann Franden, Arjun Singh, and Min Zhang. 2003. "D-Xylose Transport by Candida Succiphila and Kluyveromyces Marxianus." In Applied Biochemistry and Biotechnology - Part A Enzyme Engineering and Biotechnology. https://doi.org/10.1385/ABAB:106:1-3:255.

Strucko, Tomas, Katharina Zirngibl, Filipa Pereira, Eleni Kafkia, Elsayed T. Mohamed, Mandy Rettel, Frank Stein, et al. 2018. "Laboratory Evolution Reveals Regulatory and Metabolic Trade-Offs of Glycerol Utilization in Saccharomyces Cerevisiae." Metabolic Engineering. https://doi.org/10.1016/j.ymben.2018.03.006.

Tang, Shuang Yan, Shuai Qian, Olubolaji Akinterinwa, Christopher S. Frei, Joseph A. Gredell, and Patrick C. Cirino. 2013. "Screening for Enhanced Triacetic Acid Lactone Production by Recombinant Escherichia Coli Expressing a Designed Triacetic Acid Lactone Reporter." Journal of the American Chemical Society 135 (27): 10099–103. https://doi.org/10.1021/ja402654z.

Tomás-Pejó, E., J. M. Oliva, A. González, I. Ballesteros, and M. Ballesteros. 2009. "Bioethanol Production from Wheat Straw by the Thermotolerant Yeast Kluyveromyces Marxianus CECT 10875 in a Simultaneous Saccharification and Fermentation Fed-Batch Process." Fuel. https://doi.org/10.1016/j.fuel.2009.01.014.

Vickery, Christopher R., Javier Cardenas, Marianne. Bowman, Michael D. Burkart, Nancy A. Da Silva, and Joseph P. Noel. 2018a. “A Coupled in Vitro/in Vivo Approach for Engineering a Heterologous Type III PKS to Enhance Polyketide Biosynthesis in Saccharomyces Cerevisiae.” Biotechnology and Bioengineering, no. December 2017: 1–9. https://doi.org/10.1002/bit.26564.

Vickery, Christopher R., Javier Cardenas, Marianne E. Bowman, Michael D. Burkart, Nancy A. Da Silva, and Joseph P. Noel. 2018b. “A Coupled in Vitro/in Vivo Approach for Engineering a Heterologous Type III PKS to Enhance Polyketide Biosynthesis in Saccharomyces Cerevisiae.” Biotechnology and Bioengineering 115 (6): 1394–1402. https://doi.org/10.1002/bit.26564.

Xie, Dongming, Zengyi Shao, Jihane Achkar, Wenjuan Zha, John W. Frost, and Huimin Zhao. 2006. “Microbial Synthesis of Triacetic Acid Lactone.” Biotechnology and Bioengineering. https://doi.org/10.1002/bit.20759.

Yang, Chun, Shenglin Hu, Songli Zhu, Dongmei Wang, Xiaolian Gao, and Jiong Hong. 2015. “Characterizing Yeast Promoters Used in Kluyveromyces Marxianus.” World Journal of Microbiology and Biotechnology 31 (10): 1641–46. https://doi.org/10.1007/s11274-015-1899-x.

Yu, James, Jenny Landberg, Farbod Shavarebi, Virginia Bilanchone, Adam Okerlund, Umayangani Wanninayake, Le Zhao, George Kraus, and Suzanne Sandmeyer. 2018.

"Bioengineering Triacetic Acid Lactone Production in Yarrowia Lipolytica for Pogostone Synthesis." Biotechnology and Bioengineering. https://doi.org/10.1002/bit.26733.

Zhang, Biao, Jia Zhang, Dongmei Wang, Ruixiang Han, Rui Ding, Xiaolian Gao, Lianhong Sun, and Jiong Hong. 2016. "Simultaneous Fermentation of Glucose and Xylose at Elevated Temperatures Co-Produces Ethanol and Xylitol through Overexpression of a Xylose-Specific Transporter in Engineered Kluyveromyces Marxianus." Bioresource Technology. https://doi.org/10.1016/j.biortech.2016.05.068.

Zhou, Hai Xiang, Jin Li Xu, Zhe Chi, Guang Lei Liu, and Zhen Ming Chi. 2013. "β-Galactosidase over-Production by a Mig1 Mutant of Kluyveromyces Marxianus KM for Efficient Hydrolysis of Lactose." Biochemical Engineering Journal 76. https://doi.org/10.1016/j.bej.2013.04.010.

Zhou, Hang, Jing sheng Cheng, Benjamin L. Wang, Gerald R. Fink, and Gregory Stephanopoulos. 2012. "Xylose Isomerase Overexpression along with Engineering of the Pentose Phosphate Pathway and Evolutionary Engineering Enable Rapid Xylose Utilization and Ethanol Production by Saccharomyces Cerevisiae." Metabolic Engineering. https://doi.org/10.1016/j.ymben.2012.07.011.

Chapter 3

Improved polyketide synthesis via a heterologous carbon-saving pathway in the yeasts *Saccharomyces cerevisiae* and *Kluyveromyces marxianus*

Abstract

A heterologous carbon-saving pathway was introduced to increase the synthesis of polyketides in *Saccharomyces cerevisiae* and *Kluyveromyces marxianus.* This pathway was constructed through the addition of just two heterologous genes, a phosphoketolase (xpk) and a phosphotransacetylase (pta), which reroute carbon away from CO_2 and toward acetyl-CoA. Multiple enzyme sources were screened and studies using plasmids guided our genomic integration strategies. Integration of the two enzymes improved the production of triacetic acid lactone (TAL) by up to 1.6-fold in *K. marxianus* with xylose as the carbon source. Improvements on this system were supported by computational modeling, and a genome scale model was designed with the addition of these heterologous pathways. This model is used for metabolic flux balance analysis using the COBRA toolbox to predict optimal gene disruptions in *K. marxianus*. This system demonstrates a new schema for approaching production yields by reducing carbon losses in the form of CO_2.

Introduction

Sustainable bioprocessing is one of many ways to reduce our human impact on the environment to make the goods and products that we need. Using microbes found in nature, we can engineer new pathways and develop production systems that can replace traditional manufacturing. With extensive development already in place for *Escherichia coli*, *Saccharomyces cerevisiae* and many other hosts, there is a need to develop tools to explore and exploit the unique properties of non-model organisms.

Kluveromyces marxianus is a yeast with many advantages as an industrial microorganism, including a broad substrate range, thermotolerance up to 52C (Banat, Nigam, and Marchant 1992) and rapid growth rate (twice that of *S. cerevisiae*) (Groeneveld, Stouthamer, and Westerhoff 2009). While not widely used industrially, the promise of this host can now be realized due to new tools that can ease engineering including CRISPR-Cas9 systems (Löbs et al. 2017; Rajkumar et al. 2019) and genome-scale models (Marcišauskas, Ji, and Nielsen 2019). Although the full metabolism behind this organism are still poorly understood, recent work has focused on understanding the unique substrate utilization and temperature-sensitive abilities of *K. marxianus* using RNAseq (J. Gao et al. 2015; Schabort et al. 2016). The unique and broad substrate range has been demonstrated to be, in part, due to unique sugar transport systems (Hua et al. 2019; Varela et al. 2019) and improved flux through fatty acid catabolism (Schabort et al. 2016).

Previously, we've shown that this yeast has significant potential in the production of acetyl-CoA based products through simple enzyme and expression system engineering, producing the polyketide triacetic acid lactone (TAL) at yields that are comparable to some of the best engineered *S. cerevisiae* and *Yarrowia lipolytica* strains (Chapter 2). Metabolic pathway engineering in this promising yeast can further improve these yields.

Our production platform focuses on optimization of acetyl-CoA based products. The simple polyketide, triacetic acid lactone, is formed from one acetyl-CoA and two malonyl-CoA molecules by a heterologous 2-pyrone synthase enzyme (2-PS) and serves as an indicator for improvements in acetyl-CoA precursor formation. TAL is also a valuable platform chemical that is a precursor to sorbic acid, pogostone and many other fine and commodity chemicals (Chia et al. 2012). Development of this platform simultaneously in *S. cerevisiae* and *K. marxianus* furthers our understanding of yeast metabolism for optimization of synthesis of a wide range of polyketide as well as other acetyl-CoA based products.

One fundamental strategy for maximizing production yield of a variety of products is through increasing theoretical carbon yield. Mechanisms to increase theoretical yield in yeast minimize production of waste components and reduce overproduction of pathway intermediates, with two examples being cofactor specificity swapping (King and Feist 2014) and carbon-saving (C-saving) (Sonderegger, Schümperli, and Schu 2004). The method of C-saving reduces CO_2 formation via introduction of heterologous pathways, and has been applied in *E. coli* (Bogorad, Lin, and Liao 2013) and extended to *S. cerevisiae* (Meadows et al. 2016). The benefit

of C-saving is that it is proven effective via addition of as few as two heterologous enzymes and does not require extensive enzyme engineering, unlike swapping enzymes' cofactor requirements. These new pathways are not without issues, however, and in *S. cerevisiae* have been shown to increase acetate accumulation and reduce cellular fitness without additional engineering (Bergman et al. 2016).

The simplest method for C-saving works to reduce CO_2 formation through the introduction of two genes that enable the diversion of fructose-6-phosphate away from glycolysis and instead through non-oxidative pentose phosphate towards the production of acetyl-CoA. Two heterologous enzymes, phosphoketolase (Xpk) and phosphotransacetylase (Pta), promote the conversion of fructose 6-phosphate into acetyl-phosphate and acetyl-phosphate into acetyl-CoA, respectively. In native acetyl-CoA production, CO_2 is formed during acetaldehyde generation, losing 1/3 of the total carbon from glucose. The heterologous Xpk/Pta pathway provides an alternative pathway to acetyl-CoA which is able to convert 100% of the carbon from glucose into carbon for the desired product. Although this conversion is based on direct synthesis and neglects the carbon needs of other cellular processes, incorporating this schematic provides an alternative to the native pathway that reduces CO_2 generation and can be deemed 100% carbon efficient.

The potential of alternative substrates coupled with C-saving is even more promising when applied to a yeast such as *Kluyveromyces marxianus*, which already yields high TAL titers in both glycerol and xylose. In Meadows et al., addition of C-saving enzymes to a *S. cerevisiae*

genome scale model decreased oxygen needs per mol of farnesene by 48%, a phenomena that was verified in the bioreactor (Meadows et al. 2016). As a Crabtree-negative and fast growing yeast with high oxygen demands, *K. marxianus* would further benefit from this reduced oxygen utilization observed in other organisms with this pathway implemented in glucose.

One systematic method for understanding and predicting pathway engineering targets is *in silico* software tools. Metabolic models can be generated as a list of stoichiometric reactions of every known pathway of an organism. Genome scale models (GEMs) encompass all the known reactions within an organism based on stoichiometry and have been well-developed in *S. cereivisiae* and is continually updated by the community (Lu et al. 2019). GEMs for non-model organisms such as *K. marxianus* are beginning to be developed (Marcišauskas, Ji, and Nielsen 2019). Coupling a GEM with flux balance analysis software, a steady state model of an organism can be generated that can optimize for production, growth rate or many other factors given constraints such as substrate or oxygen availability. One such software package, the OptKnock package within the COBRA toolbox (Heirendt et al. 2019), can be used to predict and optimize various traits by removing one or more genes within a model. This tool can be useful for predicting how subsequent gene additions or removals can influence the organism as a whole as well as improve our understanding of combinatorial gene interventions.

To engineer these pathways in *K. marxianus*, we first had to build our strains that were more amenable to pathway engineering. One key limitation to efficient strain engineering in *K. marxianus* is native nonhomologous end joining. Nonhomologous end joining (NHEJ) is the mechanism used by many organisms to repair DNA double strand breaks. The NHEJ pathway (**Figure 14**) relies on a series of enzymes – Ku70/Ku80, Lig4/Lif1, and Nej1, which aid in binding, holding together and then ligating DNA that has a double strand break (Dudášová, Dudáš, and Chovanec 2004). Without NHEJ, an organism is unable to repair a double strand break and therefore unable to successfully replicate its genome. Removal of NHEJ is essential for improving CRISPR-Cas9 efficiency, since it enables selection of strains that

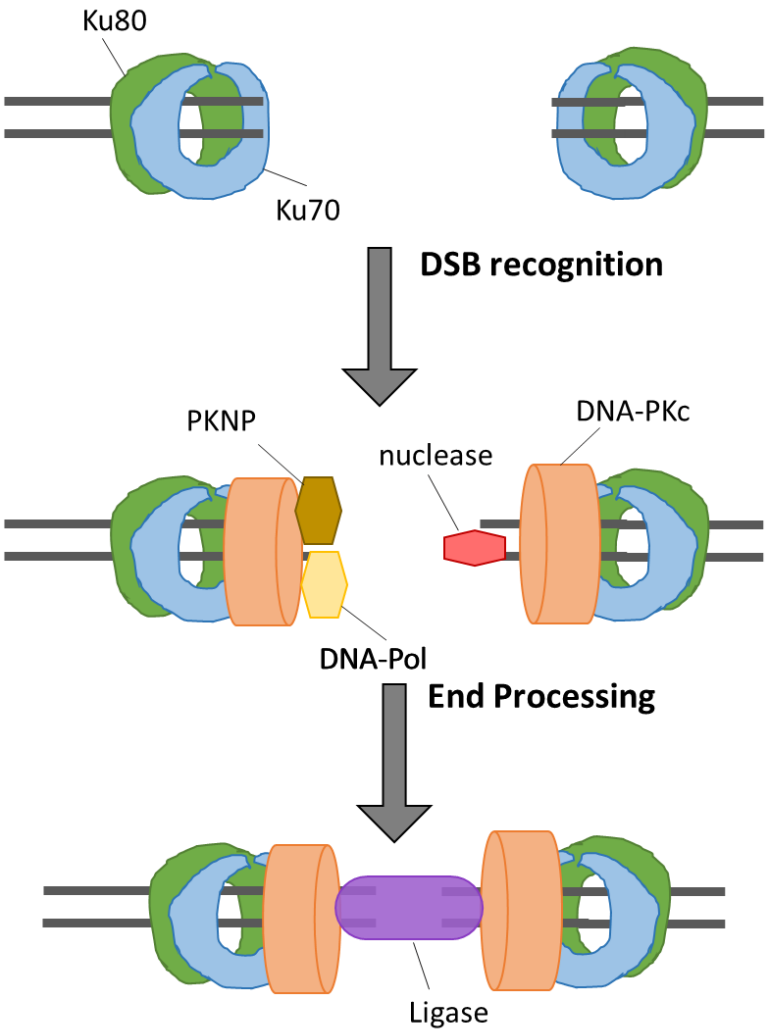


Figure 14 Simplified schematic of enzymes involved in NHEJ in yeast. In particular, the Ku70 and Ku80 enzymes are essential for double strand break (DSB) recognition and are the first steps in the NHEJ pathway.

successfully repair the cut generated by Cas9 without the need for integration and then removal of antibiotic or auxotrophic markers.

Removing genes within this NHEJ pathway, in particular the enzymes involved in recognition of DNA ends (*KU70/KU80*) enables much more efficient selection of successfully repaired CRISPR-modified strains. Although Ku70 deficient strains have demonstrated lower fitness and increased genome instability over time (Juergens et al. 2018), the expected increase in CRISPR/Cas9 efficiency should simplify the restoration of *KU70* functionality after strain optimization and screening. An alternative to *KU70* disruption would be *NEJ1*, which is a protein important for DNA ligase recruitment, or *DNL4*, the ligase itself, which both have been applied for increased CRIPSR efficiency and reduced NHEJ in *K. marxianus* (Nambu-Nishida et al. 2017).

Improving the efficiency of the CRISPR-Cas9 expression system in *K. marxainus* can also improve our ability to engineer the yeast pathway. For example, incorporation of a tRNA and hepatitis delta virus (HDV) ribozyme based system, which relies on RNA polymerase II (Pol II), has improved CRISPR efficiency by increasing gRNA stability when compared to frequently used *SNR52* which utilizes RNA polymerase III (Pol III) (Ryan et al. 2014; Juergens et al. 2018).

In this study, we aimed to increase polyketide titer and yield in *S. cerevisiae* (during growth on glucose) and *K. marxianus* (during growth on xylose) by reducing carbon loss in the form of CO_2. This was achieved through heterologous enzyme screening in *S. cerevisiae*, and then subsequent genomic integration via a new library of CRISPR gRNA targeting plasmids. These validated enzymes were then evaluated in *K. marxianus*, both on plasmids and after integration

into the genome. *K. marxianus* integration was achieved through removal of nonhomologous end joining and subsequent CRISPR-Cas9 platform development. Using a genome scale model, we added these heterologous pathway genes in order to identify new CRISPR targets that can further optimize this pathway through removal of competing genes. Reducing CO_2 production will lower cost, increase yield, improve fermentation through lowered oxygen demand, and reduce the overall waste produced per unit of product.

Materials and methods

Strains

Escherichia coli strain DH5α (Invitrogen, Carlsbad, CA) was used for plasmid maintenance and amplification. *S. cerevisiae* BY4741Δ*trp1* (Open Biosystems, Huntsville, AL) was used as the background strain. The *K. marxianus* strains employed were CBS6556Δ*ura3* (Löbs et al., 2017), CBS 712 (ATCC 200963; ATCC®, Manassas, VA), and KM1Δ*ura3* (Pecota et al., 2007). For CBS 712, the *URA3* locus was disrupted as previously described (Pecota et al., 2007) resulting in CBS712Δ*ura3* (**Table 9**). In all three *K. marxianus* strains lacking *URA3*, the *KU70* locus was removed using homologous recombination and a URA-blaster cassette. The native *KU70* gene was PCR amplified from gDNA of CBS6556 and inserted into the plasmid pBluescript SKII at the BamHI and EcoRI restriction sites. Next, the URA-blaster cassette from pδdUb-ADH2 (W. Lee and DaSilva 2006) was amplified with homology internal to *KU70,* such that the gene is disrupted, and the two fragments were assembled using Gibson 2x HiFi master mix (NEB, Ipswich, MA) to form the plasmid KU70-URAblaster. To form the donor for homologous recombination, the KU70-

URAblaster plasmid was digested with NaeI and SacI, gel extracted and then used for transformation with *K. marxianus* Δ*ura3* strains, then plated on selective medium without uracil. Once the integration was verified, the URA-blaster cassette was removed via cultivation in 5-FOA, leaving a *hisG* site inside the disrupted *KU70* gene. The disruption of *KU70* was PCR and sequence verified via Sanger sequencing (Genewiz, South Plainfield, NJ) (**Table 10**).

Table 9 *K.* marxianus and *S. cerevisiae* strains used in this study

Strain name	Description	Source
DH5α	*E. coli*, fhuA2 a(argF-lacZ)U169 phoA glnV44 a80a(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	Invitrogen
BY4741	*S. cerevisiae,* MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	Open Biosystems
BYt	*S. cerevisiae,* BY4741 trp1::KanM	Open Biosystems
BYt-2PS int	*S. cerevisiae,* BY4741 trp1::KanM, leu2::ADH2p-2PS	this study
CBS6556*Δura3*	*K. marxianus* Δ*ura3*	Löbs et al., 2017
KM1*Δura3*	*K. marxianus* Δ*ura3*	Pecota et al., 2007
CBS712	*K. marxianus*	ATCC
CBS712*Δura3*	*K. marxianus* Δ*ura3*	this study
CBS6556*Δura3Δku70*	CBS6556 Δ*ura3* ku70::hisG	this study
CBS712*Δura3Δku70*	CBS712 Δ*ura3* ku70::hisG	this study
KM1*Δura3Δku70*	KM1 Δ*ura3* ku70::hisG	this study

Table 10 Primers and plasmids used to remove *KU70* from *K. marxianus*

Primer Name	Sequence	Purpose
KM_KU70_F	tgagcttcaataccctgattgactggaaggGGAAAGCTGCGGCCAAACTC	*KU70* backbone amplification
KM_KU70_R	tcatgcgttcatgcaccactggaagatccCCACTTCTCCAGTTTCCGCATC	
KM_KU70_native_F	TTC AGG ATC CAT GTA GAA AGC CGT TCG AGA TTA AGA	*KU70* amplification from gDNA
KM_KU70_native_R	TTC AGA ATT CGG AAT TGC CAT GTA AAA ACA TAC GG	
Plasmid Name	**Description**	**Source**
pδdUB-ADH2	*URA3*-blaster (flanked with *hisG*) integrating vector, AmpR	Lee and DaSilva, 2006
pBluescript SKII	*E. coli* cloning vector, AmpR	Addgene
pB-SKII-KU70	*E. coli cloning vector*, AmpR with $KU70_{KM}$	This study
KU70-URAblaster	pB-SKII_KU70 with URAblaster casette internal to *KU70*	This study

Media and cultivation

Luria-Bertani (LB) media with 150 mg/L ampicillin was used for *E. coli* culture. *S. cerevisiae* and *K. marxianus* were cultivated in complex YP medium (1% Bacto yeast extract, 2% Bacto peptone), selective SC medium (0.67% yeast nitrogen base, 0.5% ammonium sulfate, 0.5% casamino acids), or minimal selective S medium (0.67% yeast nitrogen base, 0.5% ammonium sulfate, 100mg/L adenine-sulfate), each supplemented with either 10 g/L of glucose (*S. cerevisiae*) or xylose (*K. marxianus*). Media containing glucose are designated YPD, SDC and SD, whereas xylose are designated YPX, SXC, and SX. Supplementation of 100mg/L adenine-sulfate, uracil or tryptophan are denoted by parenthesis – for example, SDC(A) includes adenine, SDC(A,T) include adenine and tryptophan.

S. cerevisiae and *K. marxianus* were inoculated from glycerol stocks or plates into either 5 mL (*S. cerevisiae*) or 3 mL (*K. marxianus*) of medium in 15 x 125mm borosilicate glass culture tubes and cultivated at 30°C (*S. cerevisiae*) or 37°C (*K. marxianus*) and 250 rpm overnight in an air incubator shaker (New Brunswick Scientific Co. Excella E25, Edison, NJ) or a gyratory water bath shaker (New Brunswick Scientific Co. Model G67D, Edison, NJ or Amerex GYROMAX 929, Concord, CA). Culture tubes were maintained at ~45° angle for the duration of incubation to enhance gas exchange.

For TAL expression, an overnight tube culture was used to inoculate fresh medium to an initial optical density (OD_{600}) of 0.3 for *S. cerevisiae* or 0.1 for *K. marxianus* (Shimadzu UV-2450 UV-Vis Spectrophotometer, Columbia, MD) and cultivated at 30°C (*S. cerevisiae*) or 37°C for (*K. marxianus*). *K. marxianus* strains underwent a second inoculation the next day to an OD_{600} of 0.1

at which point the timed cultivation began. Expression in *S. cerevisiae* used glucose as a substrate at 1% concentration (1% SDC(A) supplemented with tryptophan, if necessary) and *K. marxianus* used 1% xylose (1% SXC(A)). After 48 hours of cultivation, the culture was removed from the incubator and prepared for TAL quantification.

xpk and pta gene isolation and plasmid construction

The xpk and pta genes were isolated from four microorganisms: *Methanosarcina thermophilia* (ATCC 43570), *Bifidobacterium animalis subsp. lactis* (ATCC 700541), *Bacillus subtilis* (Bacilus subtillis wild type isolate (Branda et al. 2001) was a gift from the lab of Prof. Allon Hochbaum, UCI), and *Oenococcus oeni* (ATCC BAA-331D-5, genomic DNA aliquot).

M. thermophilia and *B. lactis* were cultivated anaerobically under H_2/N_2 (5%/95%) in a 20 mL anerobic test tube using specialized medium. The *M. thermophilia* medium was adapted from ATCC (ATCC) and contained, per liter, 4 g NaOH, 2 g yeast extract, 2 g peptone, 1 g NH_4Cl, 0.4 g $K_2HPO_4 \cdot 3H_20$, 1 g $MgCl_2 \cdot 6H_2O$, 0.4 g $CaCl_2$, 10 mL of 100x trace mineral solution and 30 mM methanol as substrate. This trace mineral solution contains, per 100 mL, 50 mg Na·EDTA dehydrate, 15 mg $CoCl_2 \cdot H_2O$, 10 mg $NgCl_2 \cdot 4H_2O$, 10 mg $FeSO_4 \cdot 7H_2O$, 10 mg ZnCl, 2 mg $CuCl_2 \cdot 2H_2O$, 1 mg H_3BO_3, 1 mg Na2, 1 mg $Na_2MoO_4 \cdot 2H_2O$. *B. lactis* was cultivated in a rich medium containing 5 g yeast extract, 15 g tryptone, 5 g glucose, 0.5 g cysteine (added in anerobic chamber), 2.5 g NaCl per liter. Both strains were grown in an air shaker at 200 rpm at 45°C and 37°C, respectively. *B. subtilis* was cultivated in rich SOC medium (20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 0.19 g KCl, 0.95 g MgCl2, and 3.6 g glucose per liter) at 37°C in an air shaker at 200 rpm. After two days

of cultivation, a sterile syringe was used to remove each culture from the vial and gDNA was extracted using a traditional phenol chloroform method (Dymond 2013).

From the gDNA from each of the four species, the two enzyme genes were PCR-amplified with Q5 polymerase (NEB, Ipswich, MA) and primers (**Table 11**) containing compatible restriction sites for cloning into pXP-based vectors (SpeI/XhoI). These fragments were cloned into both low- (CEN/ARS) and multi- (2μ) copy pXP-based *S. cerevisiae* plasmids (pXP316, pXP318, pXP416 and pXP418 (Fang et al. 2011)) using T4 ligase (NEB, Ipswich, MA). Each plasmid contains either *TRP1* (ptas) or *URA3* (xpks) selection markers, a strong, constitutive promoter *TEF1*p and terminator *CYC1*t (**Table 12**). After cloning and amplification in DH5α, the plasmids were sequence-verified using Sanger sequencing (Genewiz, South Plainfield, NJ) to ensure no mutations occurred during PCR.

These *S. cerevisiae* plasmids was also used to construct plasmids for *K. marxianus*. The *S. cerevisiae* plasmids pXP316m and pXP318o served as a basis for cloning via Gibson assembly into *K. marxianus* CEN/ARS plasmids. First, either the pta or xpk cassette containing TEF1p and CYCt was inserted into pCV842 into the multiple cloning site via SpeI/XhoI digestion. Next, the pCS842-PTA plasmid was linearized by KpnI and the PCR amplified xpk cassette, containing TEF1p and CYCt with 40 bp overhangs to the KpnI site, was Gibson assembled into the linearized vector to create pCV-842(KpnI)-PTA-XPK.

Table 11 Primers used to amplify C-saving genes from genomic DNA and construct expression plasmids

Primer Name	Sequence	Purpose
O_oeni_SpeI_Fwd	AAAAactagtCCACCatggcagctaatgcacctgc	removal of xpk gene from *O. oeni* for ease of cloning into pXP-based plasmids
O_oeni_XhoI_Rev	GGGGctcgagttactttaaagcagtccagt	
B_lactis_SpeI_Fwd	GGGGactagtCCACCATGACTAATCCTGTTATTGG	removal of xpk gene from *B. lactis* for ease of cloning into pXP-based plasmids
B_lactis_XhoI_Rev	AAAActcgagTTACTCGTTGTCGCCGGCGG	
M_thermo_SpeI_Fwd	GGGactagtCCACCATGGTAACATTTTTAGAAAAA	removal of pta gene from *M. thermo* for ease of cloning into pXP-based plasmids
M_thermo_XhoI_Rev	AAAActcgag TTATTTTTGCTGAGCTGCGG	
B_subtilis_SpeI_Fwd	GGGGGactagtCCACCATGGCAGATTTATTTTCAACA	removal of pta gene from *B. subtilis* for ease of cloning into pXP-based plasmids
B_subtilis_XhoI_Rev	AAAActcgagTTACAGTGCTTGCGCCGCTG	
KM_PTA_Fwd	ggttccgcgcacatttccccgaaaagtgccacctgacgtCGTCTAAGAAACCATTATTAT	pta amplification for Gibson into *K. marxianus* plasmids
KM_PTA_Rev	GAAGGCTTTAATTTGCGGCCaagcttgcatgcctgcaggtcgactctagaggatcCCCGG	

Table 12 Constructed carbon saving plasmids

Plasmid name	Description
pXP316 (316)	CEN/ARS, *TRP1, TEF1-CYCt*
pXP318 (318)	CEN/ARS, *URA3 TEF1-CYCt*
pXP416 (416)	2µ, *TRP1, TEF1-CYCt*
pXP418 (418)	2µ, *URA3 TEF1-CYCt*
pXP316-B. subtilis (316s)	CEN/ARS, *TRP1, TEF1-*PTA_{BS}*-CYCt*
pXP316-M. thermophilia (316m)	CEN/ARS, *TRP1, TEF1-*PTA_{MT}*-CYCt*
pXP416-B. subtilis (416s)	2µ, *TRP1, TEF1-*PTA_{BS}*-CYCt*
pXP416-M. thermophilia (416m)	2µ, *TRP1, TEF1-*PTA_{MT}*-CYCt*
pXP318-B. lactis (318l)	CEN/ARS, *URA3, TEF1-*XPK_{BL}*-CYCt*
pXP318-O. oeni (318o)	CEN/ARS, *URA3, TEF1-*XPK_{OO}*-CYCt*
pXP418-B. lactis (418l)	2µ, *URA3, TEF1-*XPK_{BL}*-CYCt*

pXP418-O. oeni (418o)	2µ, *URA3, TEF1-XPK$_{OO}$-CYCt*
pCV842-PTA	CEN/ARS$_{KM}$, *URA3, ADH2-PTA$_{MT}$-CYCt*
pCV842-XPK	CEN/ARS$_{KM}$, *URA3, ADH2-PTA$_{OO}$-CYCt*
pCV842-PTA-XPK(KpnI)	CEN/ARS$_{KM}$, *URA3, ADH2-PTA$_{OO}$-CYCt, ADH2-PTA$_{MT}$-CYCt*

S. cerevisiae CRISPR plasmid construction

The Cas9 plasmid p414-TEF1p-Cas9-CYC1t-TRP1 (Addgene, Church 2014) was used to construct p414-TEF1p-Cas9-CYC1t-URA3. The *URA3* gene from pXP318 and the p414 backbone, split into two fragments via PRC amplification with Q5 polymerase (NEB) and the three pieces Gibson assembled to construct p414-TEF1p-Cas9-CYC1t-URA3 (2x Hi Fi Master Mix) **(Table 13)**. This plasmid was sequence verified via GeneWiz.

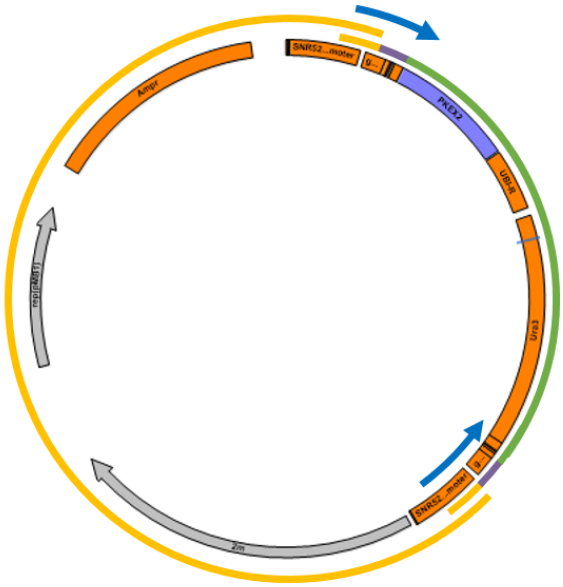


Figure 15 dgRNA with the plasmid backbone (yellow) and gRNAs (purple) and marker (green) amplified with 60bp primers (blue) for a two-part Gibson assembly

Auxotrophic-marker targeting gRNA plasmids were constructed using Gibson assembly (NEB 2x Hi-Fi Master Mix). Starting from a plasmid backbone pBF(URA)-dgRNA which contains two empty regions for creating a 20 bp target (**Figure 15**), a two-piece assembly was constructed through PCR amplification with KOD Hot Start using 60-bp primers containing a unique 20bp targeting region unique to each auxotrophic marker (**Table 13**). Each 20bp sequence was identified using the tool CCTop (Labuhn et al. 2018; Stemmer et al. 2015) to check for nonspecific binding and was chosen based on specificity but also "efficacy score" and distance to the transcription start site.

Table 13 Primers used for *S. cerevisiae* CRISPR plasmid and gRNA construction

Primer Name	Sequence	Purpose
p414_bb_Fwd_piece1	ccaagctgcctttgtgtgcttaatcacgta	amplification of the p414-Cas9 backbone, piece 1
p414_bb_Rev_piece 1	GATCATATGCGCCAGCGCGAG	
p414_bb_Fwd_piece2	CTCGCGCTGGCGCATATGATC	amplification of the p414-Cas9 backbone, piece 2
p414_bb_Rev_piece 2	gtattgtttgtgcacttgcctatgcggtgt	
URA3_for414_Fwd	ACACCGCATAGGCAAGTGCACAAACAATACGACTCTAGAGGATCC	amplification of the *URA3* casette for the p414-Cas9 backbone
URA3_for414_Rev	tacgtgattaagcacacaaaggcagcttggATTCGAGCTCGGTAC	
URA3_dgRNA_Fwd	GcagtgaaagataaatgatcATAAGACAGGACTGTAAAGAGTTTTAGAGCTAGAAATAGC	primers for *URA3* scar targeting
URA3_dgRNA_Rev	GCTATTTCTAGCTCTAAAACTCTTTACAGTCCTGTCTTATgatcatttatctttcactgC	
LEU2_dgRNA_Fwd	GcagtgaaagataaatgatcTCTACATACATTTATCAAGA GTTTTAGAGCTAGAAATAGC	primers for *LEU2* scar targeting
LEU2_dgRNA_Rev	GCTATTTCTAGCTCTAAAACTCTTGATAAATGTATGTAGAgatcatttatctttcactgC	
HIS3_dgRNA_Fwd	GcagtgaaagataaatgatcAACGATGTTCCCTCCACCAAGTTTTAGAGCTAGAAATAGC	primers for *HIS3* scar targeting
HIS3_dgRNA_Rev	GCTATTTCTAGCTCTAAAACTTGGTGGAGGGAACATCGTTgatcatttatctttcactgC	
MET15_dgRNA_Fwd	GcagtgaaagataaatgatcAAGTAAAGCGTCTGTTAGAAGTTTTAGAGCTAGAAATAGC	primers for *MET15* scar targeting
MET15_dgRNA_Rev	GCTATTTCTAGCTCTAAAACTTCTAACAGACGCTTTACTTgatcatttatctttcactgC	
X-1_gRNA_Fwd1	GACTTTagagtgcagacactcaatcg	primers for X-1 target assembly into yTK050 via BsmBI digestion
X-1_gRNA_Rev1	AAACcgattgagtgtctgcactctAA	
X-2_gRNA_Fwd1	GACTTTagagtaagttgagtgtaagg	primers for X-2 target assembly into yTK050 via BsmBI digestion
X-2_gRNA_Rev1	AAACccttacactcaacttactctAA	
X-3_gRNA_Fwd1	GACTTTgtttggaaaagctcactgtg	primers for X-3 target assembly into yTK050 via BsmBI digestion
X-3_gRNA_Rev1	AAACcacagtgagcttttccaaacAA	
X-4_gRNA_Fwd1	GACTTTagtagttggatctttccacg	primers for X-4 target assembly into yTK050 via BsmBI digestion
X-4_gRNA_Rev1	AAACcgtggaaagatccaactactAA	
XI-1_gRNA_Fwd1	GACTTTgctggcaCatgagtcgccgg	primers for XI-1 target assembly into yTK050 via BsmBI digestion
XI-1_gRNA_Rev1	GACTTTgacgctaaaaccgtggccgt	
XI-2_gRNA_Fwd1	AAACacggccacggttttagcgtcAA	primers for XI-2 target assembly into yTK050 via BsmBI digestion
XI-2_gRNA_Rev1	GACTTTaactgttgtcaccgctccag	

XI-3_gRNA_Fwd1	AAACctggagcggtgacaacagttaa	primers for XI-3 target assembly into yTK050 via BsmBI digestion
XI-3_gRNA_Rev1	AAACtgactccctgtatgtattgcaa	
XI-4_gRNA_Fwd1	GACTTTcatcacgatacacgaggtgc	primers for XI-4 target assembly into yTK050 via BsmBI digestion
XI-4_gRNA_Rev1	AAACgcacctcgtgtatcgtgatgaa	
XI-5_gRNA_Fwd1	GACTTTAATTTTCTATAGAACGTGTA	primers for XI-5 target assembly into yTK050 via BsmBI digestion
XI-5_gRNA_Rev1	AAACTACACGTTCTATAGAAAATT	
XII-1_gRNA_Fwd1	GACCTTTCTTTTTGGACCACTTCTTCT	primers for XII-1 target assembly into yTK050 via BsmBI digestion
XII-1_gRNA_Rev1	AAACAGAAGAAGTGGTCCAAAAAG	
XII-2_gRNA_Fwd1	GACCTTTGTGACGCAGCGATAAAACCG	primers for XII-2 target assembly into yTK050 via BsmBI digestion
XII-2_gRNA_Rev1	AAACCGGTTTTATCGCTGCGTCAC	
XII-3_gRNA_Fwd1	GACCTTTATATTGATAGGAATCAGCCG	primers for XII-3 target assembly into yTK050 via BsmBI digestion
XII-3_gRNA_Rev1	AAACCGGCTGATTCCTATCAATAT	
XII-4_gRNA_Fwd1	GACCTTTAATCTTCGAAGCACTCATAC	primers for XII-4 target assembly into yTK050 via BsmBI digestion
XII-4_gRNA_Rev1	AAACGTATGAGTGCTTCGAAGATT	
XII-5_gRNA_Fwd1	GACCTTTAGTAGGAGTCAGAACCTCTG	primers for XII-5 target assembly into yTK050 via BsmBI digestion
XII-5_gRNA_Rev1	AAACCAGAGGTTCTGACTCCTACT	

To target additional integration locations, the MoClo yeast toolkit (M. E. Lee et al. 2015) was used to build "parts" that can be easily mix-and-matched and assembled via Golden Gate Assembly. To construct these plasmids, 20 bp were identified from the sequences of 14 integration sites identified by Mikkleson et al. within chromosomes X, XI and XII the *S. cerevisiae* genome (Mikkelsen et al. 2012). These sites, when completely removed, are flanked by essential genes, preventing chromosomal reorganization due to homologous recombination.

Primers containing the 20 bp targeting site as well as overhangs compatible with assembly into entry vector yTK050 via a BsmBI Golden Gate Assembly were annealed by mixing an equimolar ratio of each primer in duplex buffer (**Table 13**). These oligos were heated to 95°C in a thermocycler and then allowed to slowly cool at a rate of -1.5°C/min until they reach room temperature. These annealed oligos were then assembled into yTK050 via Golden Gate Assembly. Positive (white-colored) clones are selected in chloramphenicol LB medium by the removal of the GFP cassette. This entry vector contains parts numbered 2-4.

After annealing into the entry vector, positive clones were added in equimolar ratios to the remaining parts 1-8 (**Table 14**) into a BsaI Golden Gate Assembly and then transformed into *E. coli* and plated under ampicillin selection. Positive clones (white) were selected via pink/white screening for removal of a RFP casette. Correct gRNA construction was validated by restriction digestion with BsaI and then sequence-verified via GeneWiz using a primer internal to *URA3*. The list of plasmids constructed in shown in **Table 15**.

Table 14 Yeast toolkit "part" plasmids and their description

Part number	Part name
1	Assembly connector-1
2	
3	gRNA entry vector
4	
5	Assembly Connector-2
6	*S. cerevisiae URA3* marker
7	*S. cerevisiae* 2μ origin
8	*E. Coli* AmpR marker and origin

Table 15 *S. cerevisiae* CRISPR plasmids

Plasmid Name	Description	Source
p414-TEF1p-Cas9-CYC1t-TRP1	*TEF1*p, Cas9, *CYC1*t, *TRP1* with CEN/ARS_{SC} origin	Addgene
p414-TEF1p-Cas9-CYC1t-URA3	TEF1p, Cas9, *CYC1*t, *URA3* with CEN/ARS_{SC} origin	this study
pBF(URA)_dgRNA	*SNR52*, *URA3*, 2μ, AmpR with two filler 20bp gRNA	this study
SC_URA3_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for the *URA3* scar	this study
SC_LEU2_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for the *LEU2* scar	this study
SC_HIS3_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for the *HIS3* scar	this study
SC_MET15_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for the *MET15* scar	this study
SC_X-1_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for Chromosome X site 1	this study
SC_X-2_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for Chromosome X site 2	this study
SC_X-3_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for Chromosome X site 3	this study
SC_X-4_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for Chromosome X site 4	this study
SC_XI-1_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for Chromosome XI site 1	this study
SC_XI-2_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for Chromosome XI site 2	this study
SC_XI-3_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for Chromosome XI site 3	this study
SC_XI-4_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for Chromosome XI site 4	this study
SC_XI-5_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for Chromosome XI site 5	this study
SC_XII-1_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for Chromosome XII site 1	this study
SC_XII-2_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for Chromosome XII site 2	this study
SC_XII-3_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for Chromosome XII site 3	this study
SC_XII-4_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for Chromosome XII site 4	this study
SC_XII-5_gRNA	*SNR52*, *URA3*, 2μ, AmpR with gRNA for Chromosome XII site 5	this study

K. marxianus CRISPR-Cas9 plasmid construction

The CRISPR plasmids for *K. marxianus* were constructed in our lab by Danielle Bever (unpublished results) using a codon-optimized Cas9 from pIW601 (Löbs et al. 2017). This plasmid contains a *S. cerevisiae URA3* selection marker, native *K. marxianus* CEN/ARS and a specially designed gRNA scaffold. This gRNA scaffold contains the *S. cerevisiae TDH3* promoter, the *K. marxianus* tRNA_Gly with the HDV downstream of the scaffold, and the *S. cerevisiae CYC1*t.

gRNA plasmids were constructed from pIW t-sg-R backbone via Gibson assembly. First, this plasmid was digested with NheI and gel extracted to remove any undigested plasmid. 60 bp primers (**Table 16**) for gRNA were generated with homology to either side of the NheI cut, removing this restriction site in the process. These 60 bp primers contained 20 bp homology on either end, 20 bp of the unique target site and the forward and reverse primers are complements; therefore, they can be annealed together in duplex buffer by heating to 95°C and gentle cooling to room temperature. These fragments and the linearized backbone were assembled using 2x Hi Fi Master Mix (NEB). Negative (original backbone) clones were identified by successful digestion of the NheI, and positive clones were sequence verified via GeneWiz (**Table 17**).

Table 16 Primers for CRISPR gRNA and integration in *K. marxianus*

Primer Name	Sequence	Purpose
KM_ChIV-1 F1	cgaatcccgtcagtAAACTGGCTTCAAATCATAAgttttagagctagaaat	gRNAs for site IV-1 via Gibson
KM_ChIV-1 F2	cgaatcccgtcagtATCTAGCTAATATTCCTAACgttttagagctagaaat	
KM_ChIV-1 R1	atttctagctctaaaacTTATGATTTGAAGCCAGTTTactgacgggattcg	
KM_ChIV-1 R2	atttctagctctaaaacGTTAGGAATATTAGCTAGATactgacgggattcg	
KM_ChIV-2 F1	cgaatcccgtcagtTCTGGGGGAAGTTACTCTGGgttttagagctagaaat	gRNAs for site IV-2 via Gibson
KM_ChIV-2 F2	cgaatcccgtcagtCCCGGCTTTCCCTCTGTGCGgttttagagctagaaat	
KM_ChIV-2 R1	atttctagctctaaaacCCAGAGTAACTTCCCCCAGAactgacgggattcg	
KM_ChIV-2 R2	atttctagctctaaaacCGCACAGAGGGAAAGCCGGGactgacgggattcg	
KM_IV-1_TEF1_donor_F	AGCTCCATGTACATAATAAATATAGACTAATAAGATTCGCgcgacacggaaatgttgaat	donor for *TEF1/CYCt* at site IV-1
KM_IV-1_CYCt_donor_R	CGGAATATTGAACAACCTTGAATTCGCTGTTCCAATGCGCTCATGAGACAATAACCC	
KM_IV-2_TEF1_donor_F	AAAGAAACAGAGCCAGAAAAAAGGCAGCAACGAATCACAAgcgacacggaaatgttgaat	donor for *TEF1/CYCt* at site IV-2
KM_IV-2_CYCt_donor_R	CTATCTATAGCTCTATATCTCTACTATAACTCTATATCTCCGCTCATGAGACAATAACCC	
KM_IV-1_Check_F	*GAATACACAGCGCGGAAGAA*	site IV-1 integration check
KM_IV-1_Check_R	*GCCCACCAAATCATTTCTCC*	
KM_IV-2_Check_F	ATCAAAGCAACAAGTTCACG	site IV-2 integration check
KM_IV-2_Check_R	GGCCAAGCTAGGCAAAATTA	
KM_RHR2_F	cgaatcccgtcagtCAAGGCAGCGCACAGCTTGAgttttagagctagaaat	gRNAs for site *RHR2* promoter via Gibson
KM_RHR2_R	atttctagctctaaaacTCAAGCTGTGCGCTGCCTTGactgacgggattcg	
KM_RHR2_donorF	TCTATTCTACACACAAACAGAAAACACATATCTATATCACAAGATGTCTGCCGGACGATC	donor for *RHR2* KO
KM_RHR2_donorR	ATTAAAATTTCGTTGTTTGGGATTTTATAATTTACCATTCCAACAGATCGTCCGGCAGAC	
KM_RHR2_ExCheckF	GTCCTGTCCCCCCATGGATC	*RHR2* KO check
KM_RHR2_ExCheckR	GCCAGCCATGTAGGACATCC	
KM_PDC1_F	cgaatcccgtcagtTGATGGTTACGCCAGATTAAgttttagagctagaaat	gRNAs for site *PDC1* KO via Gibson
KM_PDC1_R	atttctagctctaaaacTTAATCTGGCGTAACCATCAactgacgggattcg	
KM_PDC1_Fv2	cgaatcccgtcagtGATGTCAGTGATCATAGCAGgttttagagctagaaat	
KM_PDC1_Rv2	atttctagctctaaaacCTGCTATGATCACTGACATCactgacgggattcg	

KM_PDC1_Fv3	cgaatcccgtcagtTGACTTCAACTTGTCCCTATgttttagagctagaaat	
KM_PDC1_Rv3	atttctagctctaaaacATAGGGACAAGTTGAAGTCAactgacgggattcg	
KM_PDC1_donorF	TCGTAATTTATCTCTTTATCCTCTCCCTCTCTATTCTTGCTTGGCGTTGATGGTAACGAC	donor for *PDC1* KO
KM_PDC1_donorR	CTCCACACCCAAACCAAATAATTGCAATGTCTGAAATTACTCTAGGTCGTTACCATCAAC	
KM_PDC1_checkF	GCCCATACGCTTATAATTCC	*PDC1* KO check
KM_PDC1_checkR	TATAAGTGGAGTGTCTGGAT	

Table 17 *K. marxianus* CRISPR plasmids

Plasmid Name	Description	Source
pIW601	TEF1p-Cas9-CYCt, gRNA, URA3, AmpR	Löbs et al., 2017
piW t-sg-R	TEF1p-Cas9-CYCt, TDH3p-tRNAgly-gRNA-HDV-CYCt, C/A, URA3, AmpR	This study
KM_IV-1.1_gRNA	piW targeting Chromosome IV-1 gRNA 1	This study
KM_IV-1.2_gRNA	piW targeting Chromosome IV-1 gRNA 2	This study
KM_IV-2.1_gRNA	piW targeting Chromosome IV-2 gRNA 1	This study
KM_IV-2.2_gRNA	piW targeting Chromosome IV-2 gRNA 2	This study
KM_RHR2_gRNA	piW targeting RHR2	This study
KM_PDC1_gRNA_v1	piW targeting PDC1, PAM site 1	This study
KM_PDC1_gRNA_v2	piW targeting PDC1, PAM site 2	This study
KM_PDC1_gRNA_v2	piW targeting PDC1, PAM site 3	This study

2-PS expression plasmids

The multi-copy plasmid pXP842-2PS (Cardenas & DaSilva, 2014) was used for all TAL screens. The *K. marxianus* CEN/ARS and pKD1 plasmids pCA-P2PS and pKD-P2PS (McTaggart et al. 2019) were used as backbones for the plasmids developed in this study. The pKD-P2PS backbone was amplified to remove the *URA3* marker and then Gibson assembled with *HIS3* amplified from pXP320 to form pKD1-2PS-HIS3 (**Table 18**).

Table 18 Primers used to create pKD1-2PS-HIS3

Primer Name	Sequence	Purpose
pCV842_bb_Fwd	atagcatacattatacgaagttatCCC	amplify the pCV842 backbone without *URA3*
pCV842_bb_Rev	CCCGGGgatcctctagagtcgac	
HIS3_Fwd	tcgactctagaggatcCCCGGGataacttcgtatagcatacattatac gaagttatCGTT	amplify *HIS3* from pXP320
HIS3_Rev	GGGataacttcgtataatgtatgctatacgaagttatTCGAGTTCA AGAGAAAAAAAAAG	

Plasmid recovery from *E. coli* was performed using the GeneJet™ Plasmid Miniprep Kit (Thermo Scientific, Waltham, MA), and DNA sequence analysis confirmed the correct sequence of all PCR-amplified inserts (GeneWiz, South Plainfield, NJ; Eton Biosciences, San Diego, CA). Q5® Hot Start High-Fidelity DNA Polymerase, T4 DNA ligase, and deoxynucleotides were purchased from New England Biolabs (Ipswich, MA). Oligonucleotide primers were purchased from IDT DNA (San Diego, CA).

S. cerevisiae integration

For CRISPR-mediated integration, we used gRNAs for the chromosomal integrations sites X-1, X-2, XI-1, XI-2, XI-3 and XI-5. Integration donors were constructed containing 45 bp homology

to the yeast genome at the region of integration, upstream and downstream of the double strand break. The donors for the pta and xpk integration were amplified from pXP316m and pXP318o and the donor for 2-PS was amplified from pXP842-2PS. Successful amplification was verified via gel electrophoresis on an agarose gel using a small aliquot of the donor DNA PCR, and the DNA was excised and then purified using Zymo Clean and Concentrator Kit (Zymo Research, Irvine, CA).

A standard yeast transformation protocol (Gietz et al. 1992) was performed using 1 ug donor DNA and 500 ng of each plasmid, and transformants were selected on 1% SDC(A) plates. Clones were screened using primers external to the integration site via colony PCR and size-verified via gel electrophoresis on a 1% agarose gel. All positive or indeterminate clones were screened again using one external primer and one primer internal to the integrated gene (**Table 19**).

Following integration, the gRNA plasmid was cured using 5-FOA in SDC(AU), the cells were plated on SDC(AU), and then colonies are streaked on SDC(A).The Cas9 plasmid (*TRP1*-markered) was retained for the next integration. Once strain modification was complete, the cells were grown in 1% YPD in test tubes, changing media every 24hrs for 2-4 days. Cells were plated on YPD and streaked on SDC(AU) and SDC(AT) to verify plasmid loss of both the gRNA and Cas9 plasmids.

Table 19 Donor and check primers for CRISPR integration in *S. cerevisiae*

Primer Name	Sequence	Purpose
X-1_donor_ADH2_F	gttatctctgtgtccagatccctttgaagtaaagtttattCATATGATTTAGCGGCC	donor DNA for ADH2p genes, site X-1
X-1_donor_CYCt_R	CAGGCATGCAAGCTTagctgcttagatatataggtagatataaataagcacaatg	donor DNA for CYCt genes in site X-1
X-2_donor_ADH2_F	taccgtctatgaggagactgttagttggatatcagtaatgCATATGATTTAGCGGCC	donor DNA for ADH2p genes, site X-2
X-2_donor_CYCt_R	CAGGCATGCAAGCTTccaaggcattaccatcccatgtaagaacggaataaaacag	donor DNA for CYCt genes, site X-2
XI-1_donor_ADH2_F	ggaatagtgacgttgtgatgcggtgagttcggcggttaggCATATGATTTAGCGGCC	donor DNA for ADH2p genes, site XI-1
XI-1_donor_CYCt_R	CAGGCATGCAAGCTTgaaccgagtccccatcaggtcaatggtagatatagccgcc	donor DNA for CYCt genes, site XI-1
XI-2_donor_ADH2_F	tttggagcaggatgaggagaaatagtaccacatgtatataCATATGATTTAGCGGCC	donor DNA for ADH2p genes, site XI-2
XI-2_donor_CYCt_R	CAGGCATGCAAGCTTatcgacccaacgatatgaacaagccaagacctctatgggg	donor DNA for CYCt genes, site XI-2
XI-3_donor_ADH2_F	CTATCATCTTGTCCAATCAAAGAAGCATCGGTTCAGATCGGCATATGATTTAGCGGCCGC	donor DNA for ADH2p genes, site XI-3
XI-3_donor_CYCt_R	AATACAGAAGAGGAAGCGCCCCGATTTCTTTTCTGTTCCGCAAATTAAAGCCTTCGAGCG	donor DNA for CYCt genes, site XI-3
X-1_donor_TEF1_F	gttatctctgtgtccagatccctttgaagtaaagtttattgcgacacggaaatgttgaat	donor DNA for TEF1p genes, site X-1
X-2_donor_TEF1_F	taccgtctatgaggagactgttagttggatatcagtaatggcgacacggaaatgttgaat	donor DNA for TEF1p genes, site X-2
XI-1_donor_TEF1_F	ggaatagtgacgttgtgatgcggtgagttcggcggttagggcgacacggaaatgttgaat	donor DNA for TEF1p genes, site XI-1
XI-2_donor_TEF1_F	tttggagcaggatgaggagaaatagtaccacatgtatatagcgacacggaaatgttgaat	donor DNA for TEF1p genes, site XI-2
XI-3_donor_TEF1_F	ATCTCGGTTTTGTAGTTTGGATGTCATTAGAGATCTACCACCACACACCATAGCTTCAAA	donor DNA for TEF1p genes, site XI-3
XI-5_donor_TEF1_F	AGCTCTTGTTGTCGATGAATTGCTCAAAATGTGGCCATTCCCACACACCATAGCTTCAAA	donor DNA for TEF1p genes, site XI-5
XI-5_donor_CYCt_R	TGTCTTTCTAAAAAAGTCCATTACCCTTAAGGTTGTTGTCGCAAATTAAAGCCTTCGAGC	donor DNA for CYCt genes, site XI-5
X-1_check_F	CGTTGATGTCTGGAGAAGGAC	check primer for gene integration, site X-1
X-1_check_R	AGCAGCTCAATCGAAATACGTGCAG	
X-2_check_F	GGGCAGAAAATGCGACAGAA	check primer for gene integration, site X-2
X-2_check_R	CTGTGAGCCTCTTACCTGTTTGGAG	
XI-1_check_F	CCACGGATTCGAGCTTACTG	check primer for gene integration, site XI-1
XI-1_check_R	TCAAAACCACTACGTACGAGAATCCG	
XI-2_check_F	TGCGAAGCCCTACTCTAACA	check primer for gene integration, site XI-2
XI-2_check_R	AGGAGTAGACTATCACGCTATTCCGG	

XI-3_check_F	GGACCATGATTGCGTCAACT	check primer for gene integration, site XI-3
XI-3_check_R	GCGATGCGACAATAAACGCG	
XI-5_check_F	ACACTAGCCTTCGATTTGACAC	check primer for gene integration, site XI-5
XI-5_check_R	CCCAAGTATGCACAGCAC	
SC_GPP1_F	GACTTTAAATGGGCTGTCGCCACCTC	GPP1 gRNA assembly into yTK050 via BsmBI digestion
SC_GPP1_R	AAACGAGGTGGCGACAGCCCATTTAA	
SC_GPP2_F	GACTTTCAGCCATTGCTGCATTCTGG	GPP2 gRNA assembly into yTK050 via BsmBI digestion
SC_GPP2_R	AAACCCAGAATGCAGCAATGGCTGAA	
SC_GPP1_donorF	AATCCGTATCATTTTCTCGCATACACGAACCCGCGTGCGCCTGGTAAATTGCGGATGATC	GPP1 KO donor
SC_GPP1_donorR	AATTCTCTAAACCAGCTTGATTTGCGCGAACCACCTGTCACGGCAGATCATCCGCAATTT	
SC_GPP2_donorF	TTTCCGTGTAAGCCGTCAAGTGAGGACTTTTCGGATGCTGAAAGAAAGTACGCCGTTATC	GPP2 KO donor
SC_GPP2_donorR	TTTAATCCGTTGTGGCTCCTGTCACTTTCAAGTTGCTAATAACCTGATAACGGCGTACTT	
SC_GPP1_ExCheckF	GAATGTGTGGGTGCGGAAGC	GPP1 KO check
SC_GPP1_ExCheckR	CGGTGGTTCCAGATGTGCCA	
SC_PDC1_F	GACTTTTGACTTCAACTTGTCCTTGT	PDC1 gRNA assembly for yTK050
SC_PDC1_R	AAACACAAGGACAAGTTGAAGTCAAA	
SC_PDC5_F	GACTTTGTTAGCAATATCAGTGATCA	PDC5 gRNA assembly for yTK050
SC_PDC5_R	AAACTGATCACTGATATTGCTAACAA	
SC_PDC6_F	GACTTTTCTGTGCTGGTAACTACTTT	PDC6 gRNA assembly for yTK050
SC_PDC6_R	AAACAAAGTAGTTACCAGCACAGAAA	
SC_MTH1_F	GACTTTAGATCAAGCGTGGCTGAAAG	MTH1 gRNA assembly for yTK050
SC_MTH1_F	AAACCTTTCAGCCACGCTTGATCTAA	
SC_MTH1_donor_F	CACGATGAGTGGCAGTGATAATGCTTCTTTTCAAAGTTTGCCACTATCAATGTTTTCTGC	MTH1 KO donor
SC_MTH1_donor_R	GAGAACGAAGAGTCATTAGTTAGTTGCGTGTGCACAGTAGAGGGGGCAGAAAACATTGAT	

K. marxianus integrations

Unlike *S. cerevisiae*, the *K. marxianus* gRNA and Cas9 enzyme were localized on one plasmid. Donor DNA was generated in the same manner as *S. cerevisiae*. The donors for the pta and xpk integration were amplified from pXP316m and pXP318o (**Table 16**). Successful donor amplification was verified via gel electrophoresis on an agarose gel using a small aliquot of the donor DNA PCR, and if the correct size, the remaining was cleaned up using Zymo Clean and Concentrator Kit.

K. marxianus cultures were grown overnight in 3mL of 2% YPD or 2% SDC(A) at 30°C. Cultures were then reinoculated in fresh media and allowed to grow to OD 2.0. Transformation of plasmids were conducted using the Zymo EZ Freeze Yeast Transformation II Kit (Zymo Research, Irvine, CA). The protocol was modified slightly in the case of transformations for gene disruptions and gene integrations by adding 10 uL sheared salmon sperm DNA (Invitrogen, Carlsbad, CA) after the addition of EZ Solution II. Each transformation used 1 ug donor DNA and 1 ug of plasmid. After EZ Solution III, transformations were incubated at 30°C for 1.5 hours, instead of the prescribed 45 minutes, before plating onto selective media. Clones were screened and plasmids removed in the same manner as for *S. cerevisiae*.

Colony PCR

As an alternative to gDNA extraction using phenol:chloroform, colony PCR enables rapid screening of clones from a CRISPR-based experiment. Colonies from a transformation plate were picked using a toothpick and suspended in 50 uL water. These aliquots were rapidly frozen at -

80°C and then heated to 95°C in a thermocycler for 30 min. 1 uL of this colony suspension was then used for PCR in 2x OneTaq alongside the PCR check primers that are external to the integration site.

TAL assay

To measure TAL concentration, samples were centrifuged at 3500 rpm (2400 x g) for 5 min (Beckman Coulter Allegra X-22R Centrifuge, Brea, CA), followed by collection of the supernatants and storage at 4°C. The concentration of TAL was measured by HPLC using a Shimadzu HPLC system: LC-10AT pumps (Shimadzu, Kyoto, Japan), UV–Vis detector at 280 nm (SPD-10A VP, Shimadzu Kyoto, Japan), Zorbax SB-C18 reversed-phase column (2.1×150 mm, Agilent Technologies, Santa Clara, CA). Acetonitrile buffered in 1% acetic acid was used as the mobile phase, while HPLC grade water buffered in 1% acetic acid was used as the aqueous phase. A gradient program using a 95–85% Pump B gradient (H_2O with 1% acetic acid) provided an elution time of approximately 12 min (flow rate 0.25 mL/min, column temperature 25 °C).

When using minimal medium for *K. marxianus* cultures, a spectrophotometric assay was used. After centrifugation, the supernatant was diluted 20-fold in water into a 96-well, UV-transparent flat-bottom plate (Corning, Corning, NY). Samples are measured in a SpectraMax M3 plate reader at an absorbance of 277 nm (Molecular Devices, San Jose, CA).

Results and discussion

Design of C-saving pathway

Improving carbon efficiency in yeast requires introduction of heterologous enzymes which can reroute carbon losses – primarily in the form of CO_2 – toward acetyl-CoA. One of the simplest ways to achieve this is through the non-oxidative pentose phosphate pathway, which can be achieved by the addition of only two heterologous genes, a pta and xpk. Our collaborators at Penn State predicted this carbon-saving pathway prior to the published work of Meadows (2016) and Bergman (2016), modeling the production of TAL from glucose with the addition of these heterologous enzymes using the COBRA package optStoic (Chowdhury and Maranas 2015) (**Figure 16A**). From this model, they predicted an increase in the maximum theoretical yield of TAL from 0.25 mol/mol glucose to 0.58 mol/mol. This pathway, however, requires two additional ATP per mol of product.

Using the optStoic toolkits minRxn and minFlux, our collaborators also generated alternative pathways (**Figure 16B,C**) that also reduce CO_2 production but do not require additional ATP. The minRxn and minFlux toolkits search from a library of enzymes and can string together reactions from organisms with very diverse backgrounds. Unfortunately, many of these enzymes have not been expressed in yeast, and may be infeasible due to the numerous oxygen-sensitive enzymatic steps.

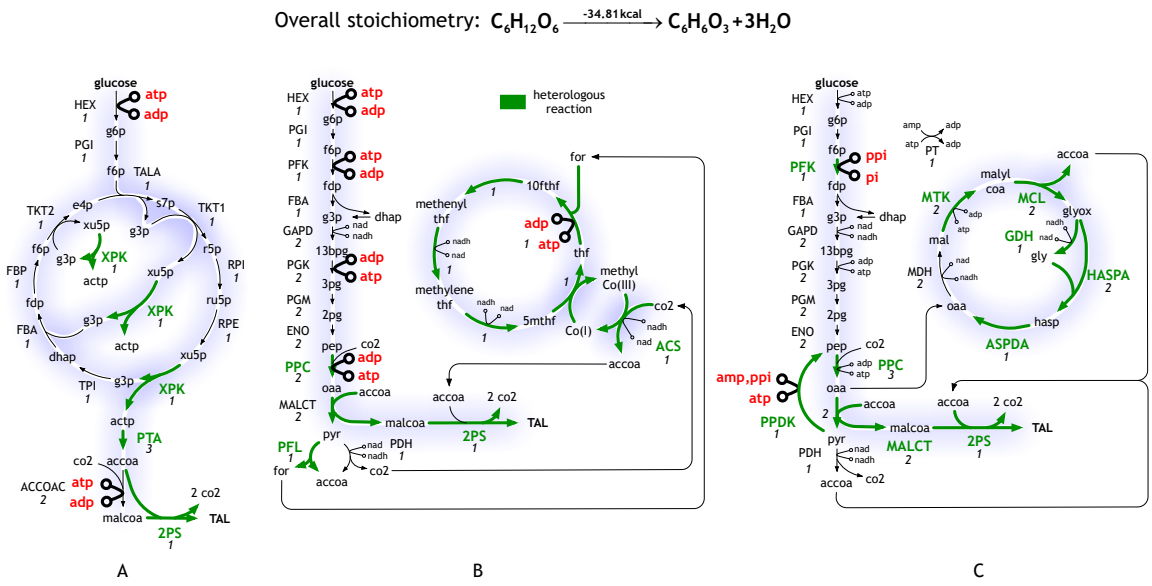


Figure 16 Heterologous pathways in *S. cerevisiae* to reduce carbon loss

We chose scheme (A) due to the simplicity of a two-enzyme system, and considered various microbial Xpk and Pta enzymes. Screening Xpk enzymes from various microbial hosts, Bergman et al. (Bergman et al. 2016) identified the Xpk enzyme from *Leuconostoc mesenteroides* as having the highest specificity for catalysis of xylulose 5-phosphate in crude cell lysates of *S. cerevisiae*. However, introduction of the xpk gene alone resulted in significant acetate accumulation that was likely detrimental to cell viability in *S. cerevisiae* (Sonderegger, Schümperli, and Schu 2004). Using the BRENDA enzyme database, we identified promising phosphoketoase (Xpk) and phosphotransacetylase (Pta) variants. Each enzyme type was sorted by observed kinetic activity (low K_m and high specific activity), then prioritized by previous use in *S. cerevisiae* or eukaryotic hosts (**Table 10**). We preferred enzymes with low K_m in order to accelerate rate of reaction and be competitive with native pathways even at low substrate concentrations.

From this screen, we chose two microbial sources per enzyme – one enzyme with the best overall characteristics and one with previous application in metabolic engineering. The top candidates for each category are the Xpk of *Oenococus oeni* (XPK_{OO}) and the Pta of *Methanosarcina thermophilia* (PTA_{MT}). For the sake of comparison, Xpk of *Bifidobacterium lactis* (XPK_{BL}) and Pta of *Bacillus subtilis* (PTA_{BS}) utilized in a previous study were also included (Sonderegger, Schümperli, and Schu 2004) (**Table 20**).

Table 20 C-saving enzyme K_m and specific activity as reported in literature

Gene	Enzyme Function	Host Organism	Domain	K_m [mM]	Specific Activity [umol/min/mg]
XPK	phosphoketolase	*Oenococcus oeni* (*Leuconostoc oenos*) (Veiga-Da-Cunha, Santos, and Van Schaftingen 1993)	Bacteria gram+	1.6	2.3
		Bifidobacterium lactis (Meile et al. 2001)	Bacteria gram+	45	4.28
PTA	phosphate acetyltransferase	*Methanosarcina thermophila* (Lundie and Ferry 1989; Lawrence and Ferry 2006)	Eukarya	0.17	2469
		Bacillus subtilis (Shin, Choi, and Park 1999; Rado and Hoch 1973)	Bacteria gram+	——	1150

We removed these genes from extracted genomic DNA of each organism and cloned them downstream of $TEF1_{SC}$ promoters, for screening in single- and multi-copy plasmids, as well as for ease of donor generation for CRISPR-based integrations.

The design for *S. cerevisiae* relies on glucose as a substrate, and provides an alternative to production that is typically routed through acetaldehyde. In **Figure 17**, heterologous enzyme reactions are highlighted, Xpk (orange) and Pta (purple). To further encourage carbon flux through this more carbon efficient pathway, we proposed gene knockout and knockdown as indicated by the red crosses. In previous work with *S. cerevisiae,* we found that removal of *ZWF1* had no noticeable effects on growth and improved production of TAL when coupled with a heterologous PDH pathway (Cardenas and Da Silva 2016). Removing *ZWF1* in our system will further reduce CO_2 losses in the pentose phosphate pathway. Furthermore, downregulation of one or more of the

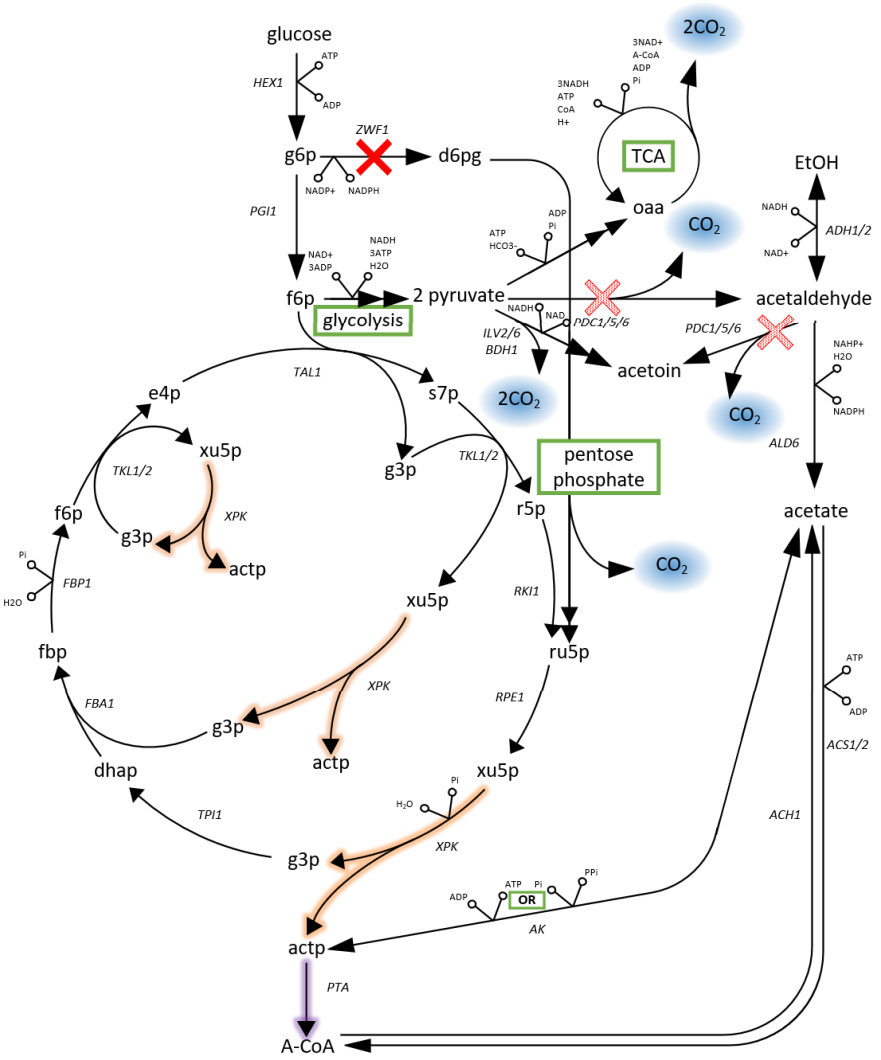


Figure 17 Proposed pathway for 100% C-efficiency in the production on acetyl-CoA. All pathway elements that produce CO_2 are shown, with heterologous Xpk and Pta enzymes highlighted in orange and purple, respectively. Areas to target for further CO_2 reduction are indicated by solid red arrows (deletion) or shaded red (downregulation or partial distruption).

PDC genes may also improve production. In *S. cerevisiae*, *PDC1/5/6* are responsible for a large amount of carbon loss in the form of CO_2, in part due to *the* natively high EtOH production in *S. cerevisiae* under anaerobic conditions. We thus designed CRISPR gRNAs to remove *ZWF1* and *PDC1/5/6*. Since removal of *PDC1/5/6* is associated with poor growth on ethanol, we also designed CRISPR gRNAs to target and mutate the *MTH1* gene to $MTH1^{81D}$, which is demonstrated to restore growth on glucose without *PDC* (Y. Zhang et al. 2015).

The pathway design for *K. marxianus* (**Figure 18**) differs from that for *S. cerevisiae*, primarily due to the differences in substrate utilization. *K. marxianus* is known to not only consume but also produce TAL from a range of nonconventional carbon sources

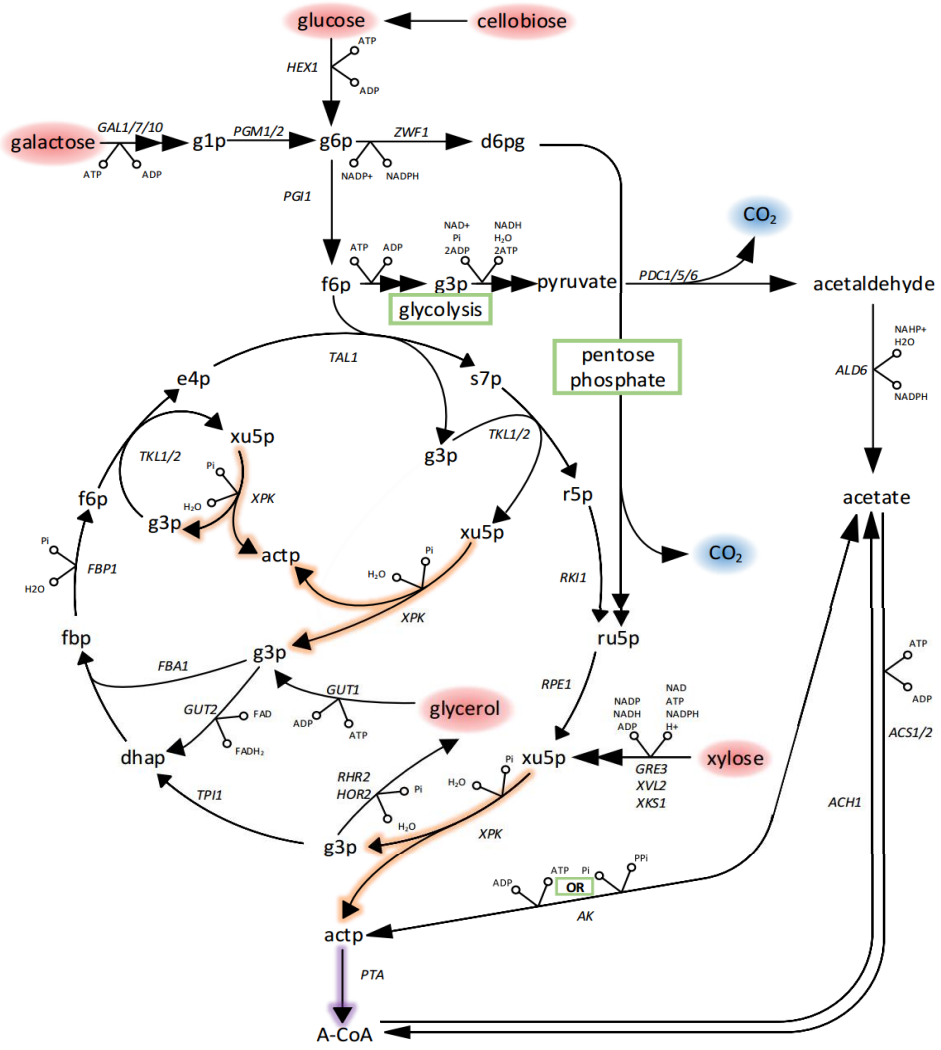


Figure 18 Carbon-saving pathway design for *K. marxianus* from a range of substrates with heterologous enzymes highlighted in orange (xpk) and purple (pta).

(McTaggart et al. 2019). When considering these alternative substrates such as xylose or glycerol, we are less likely to lose carbon through pentose phosphate, since it would require conversion back to glucose 6-phosphate. Further evidence for low ethanol production in xylose (McTaggart et al. 2019) also indications that TAL production from xylose or glycerol utilizing pta and xpk will likely have less carbon losses through *PDC* than *S. cerevisiae.* Furthermore, some of the key issues in the previous implementations of this pathway in *S. cerevisiae*, such as the accumulation of acetate (Bergman et al. 2016; Meadows et al. 2016), is less likely to affect *K. marxianus*, due to strong suppression of acetate production in aerobic conditions (Sakihama et al. 2019).

Testing the efficacy of C-saving enzyme variants in *S. cerevisiae*

All four selected Pta and Xpk enzymes were assembled into low- (CEN/ARS) and high- (2µ) copy plasmids. The pta genes were maintained on plasmids with the *TRP1* marker, and the xpk genes utilized *URA3* so both plasmids can be maintained in selective medium. The BYt-2PS int strain, with ADH2p-2PS-CYCt integrated at the *LEU2* site, was used to evaluate the performance of carbon saving in *S. cerevisiae*.

After transformation and selection on SDC(A), three independent colonies were selected, grown overnight in SDC(A), then cultivated for 48hrs in non-selective 1% YPD. Complex medium

was necessary to maximize 2-PS synthesis from the single integrated copy under the control of *ADH2p*, since production from selective medium is undetectable on the HPLC. The cells were spun down and extracellular TAL levels were measured from the supernatant (**Figure 19**).

Although titers improved only marginally, introduction of these two enzymes improved TAL titer and specific production. From specific titers (g/L/OD) the two genes identified with better kinetics in the enzyme survey (from *M. thermophilia* and *O. oeni*) outperformed the enzymes identified previously (Sonderegger, Schümperli, and Schu 2004) by ~10%. Minor improvements are expected, as titers from single copies of the 2-PS gene are typically low and significantly limited by 2-PS enzyme availability as opposed to precursor pools.

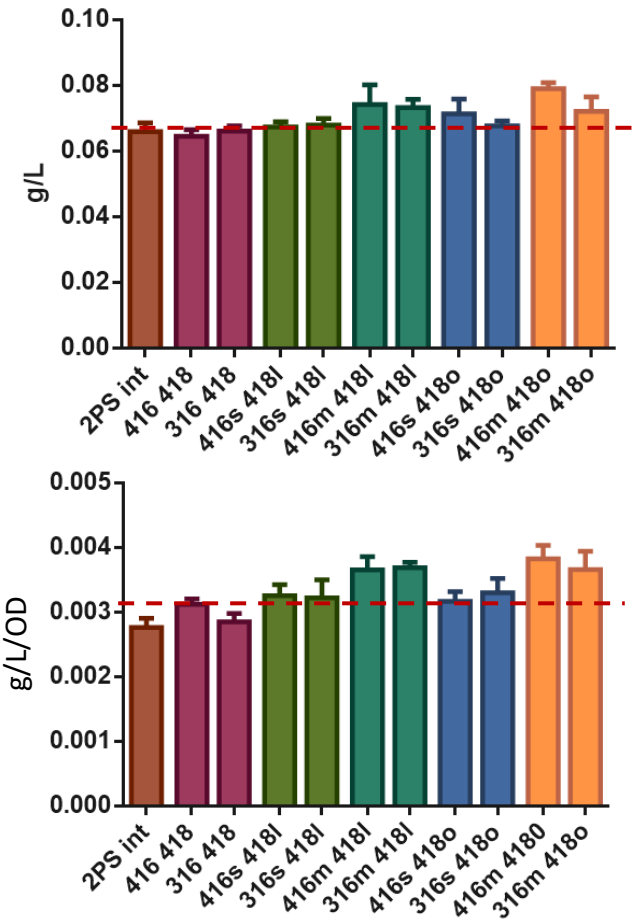


Figure 19 TAL titer and specific production in BYtΔleu2::ADH2p-2PS-CYCt (2PS int). Control without plasmids (2PS int) and with empty plasmids (416 or 316, and 418). Others are harboring two C-saving enzymes, Pta and Xpk, each on individual pXP-based plasmids as listed in **Table 20**. Two PTA genes from *B. subtilis* (s) and *M. thermophilia* (m). XPK genes from *B. lactis* (l) and *O. oeni* (o). 31X/41X are CEN/ARS and 2µ, respectively. XX6 and XX8 are *TRP1* and *URA3* markers, respectively. The dashed line represents the highest TAL titer observed in the control strains without the PTA or XPK genes.

From these screens, it is likely that integration of XPK_{OO} and PTA_{MT} genes in single copies would improve TAL titers, however additional copies of both will likely be advantageous. Therefore, we chose these two genes for integration into the genome.

In order to integrate genes into *S. cerevisiae*, we developed a gRNA plasmid library that provides sites for genomic integration. These include the four scars of the auxotrophic markers *URA3*, *HIS3*, *LEU2* and *MET15* as well as fourteen sites within the genome with high stability and flanked by essential genes (Mikkelsen et al. 2012). The details of this work and the CRISPR tools created can be found in **Appendix A**.

Sites XI-1 and X-2 which were successful for 2-PS integration were used to sequentially integrate the pta and xpk genes from *O. oeni* and *M. thermophilia* into BYt. These genes were integrated individually as well as together to determine if integration of one and not the other influenced TAL titers, generating strains BY4741Δtrp1-OO, BY4741Δtrp1-MT and BY4741Δtrp1-OO-MT. All three strains were PCR and sequence-verified via GeneWiz and then transformed with pXP842-2PS. Unfortunately, integration of single copies of pta and xpk did not show improved levels of TAL production, and the cause of this is unclear.

Improved TAL titers in *K. marxianus* with plasmid-based carbon-saving pathway genes

The carbon-saving pathway may also be useful in *K. marxianus*. Integration of this pathway may be complementary to xylose as a substrate and may further improve the ability of *K. marxianus* to produce TAL in minimal medium. Furthermore, the lowered oxygen needs of organisms with this pathway may be beneficial to *K. marxianus* for large scale fermentation in the future.

To evaluate the effect of these carbon-saving enzymes in *K. marxianus*, we ran a screen using a two-plasmid system in *CBS6556Δura3Δhis3* (Wheeldon group, unpublished). Strains were transformed first with 2-PS on the multi-copy plasmid pKD-2PS-HIS3 and with either pta or xpk on CEN/ARS plasmids (pCV842-PTA or pCV842-XPK); however, neither of these demonstrated increased TAL production individually, relative to the strain with only pKD-2PS-HIS. Although there was some speculation that *K. marxianus* might have a phosphoketolase gene or promiscuous enzyme (Evans and Ratledge 1984) to catalyze the reaction from xylulose 5-phosphate to acetylphosphate, a lack of improvement in TAL titer with the pCV842-XPK alone indicates otherwise.

Next, a bigenic, low copy plasmid pCV842-PTA-XPK(KpnI) was transformed in combination with pKD-2PS-HIS3, and the strains cultivated in minimal medium without casamino acids. TAL production in both xylose and glycerol media increased with expression of both genes, with a 4.4-fold increase in TAL production in xylose and a 5.6-fold increase in glycerol over the wild-type strain (**Figure 20**). This indicates that integration of only 1-2 of each gene into *K. marxianus* is very promising for improving TAL titers.

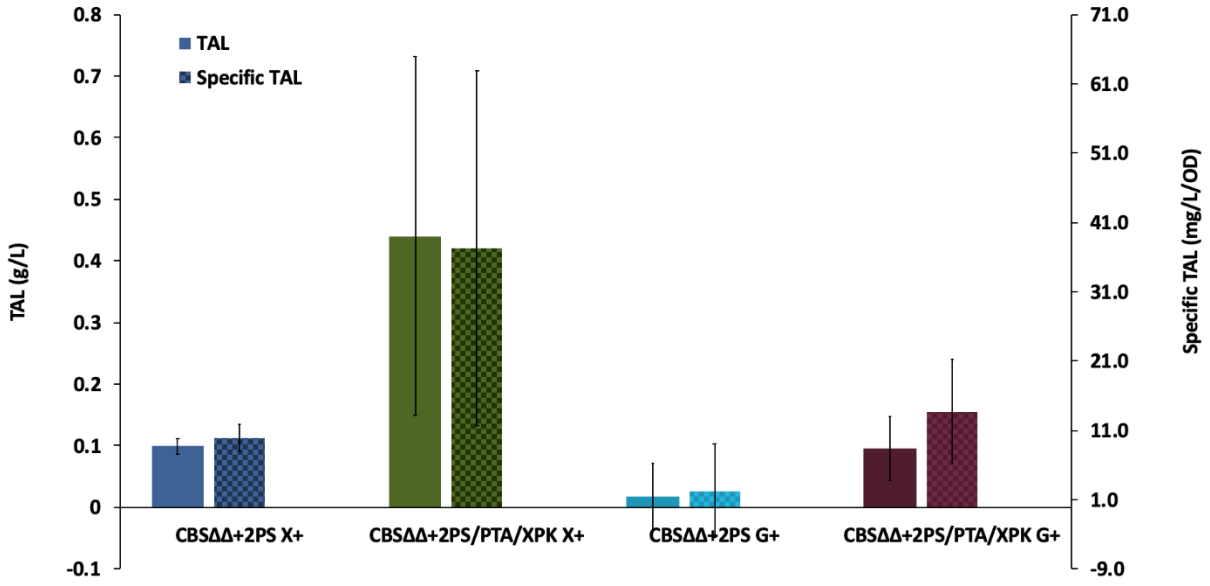


Figure 20 TAL production in *K. marxianus* strain CBS6556Δ*ura3*Δ*his3* expressing pta and xpk on low-copy plasmids in either xylose (X+) or glycerol (G+) medium. Samples were taken after 48 hr cultivation in a 37°C water bath shaker. Note that TAL production here is lower than in previous studies due to true minimal medium used for *HIS3* selection. (n=6)

We also observed very high error bars. This is due to one of the three colonies producing significantly more TAL than the other two – and this observation was validated even after selecting six more colonies (3/9 colonies had titers ~0.7 g/L while the other two produced ~0.4g/L). This indicates there is likely a copy number, plasmid stability or other effect occurring that may be remedied by gene integration.

K. marxianus TAL titers with integrated carbon-saving pathway genes

In order to integrate efficiently into *K. marxianus*, we needed to remove nonhomologous end joining (NHEJ). CRISPR-Cas9 is very efficient in *S. cerevisiae*, in large part from the lack of efficient NHEJ (Haber 1995). This allows homology directed repair to be the dominant mechanism for double strand break repair, increasing the likelihood that surviving cells performed the

desired modification. Therefore, the *KU70* gene, responsible for the first step in double strand break recognition in NHEJ, was disrupted in the genomes of strains CBS712*Δura3*, CBS6556*Δura3* and KM1*Δura3* and then sequence verified.

To confirm that deletion of *KU70* reduces NHEJ and improves homology-directed repair, we tested three different integration donors with large (~400bp) and small (40bp) homology to the genome. All donors had significantly improved knockin rates in the *KU70*-deficient strains, with no colonies observed without the presence of donor DNA. This indicated that NHEJ was eliminated in these strains (Figure 21).

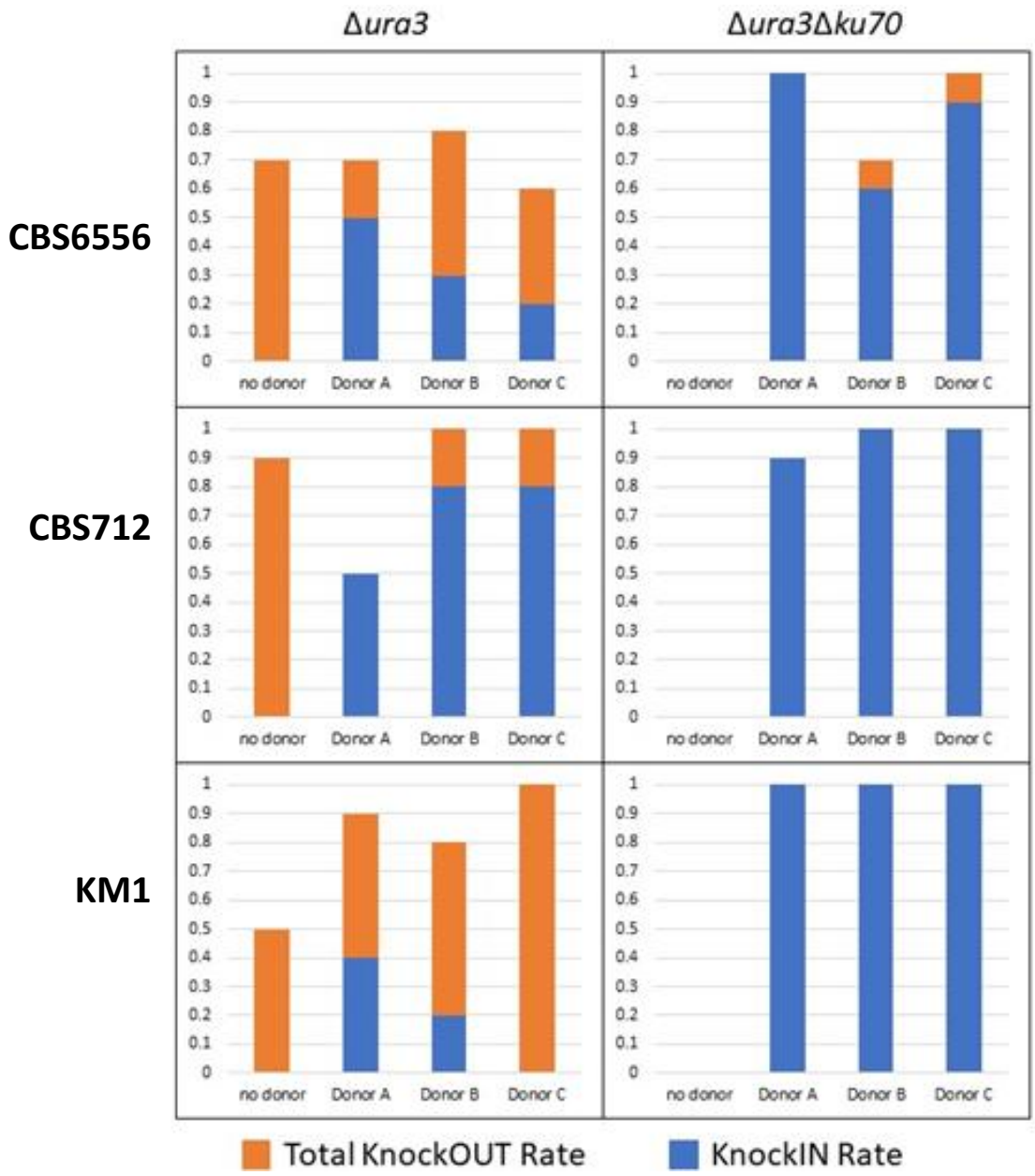


Figure 21 Total Knockout and knockin rates in CBS6556, CBS712, and KM1 (top to bottom row) *Δura3* (left column) and *Δura3Δku70* (right column) strains using pDBtgr_Cas9_leu2 and three different donors of varying homology. Transformations were conducted by standard methods except for a lengthened incubation, 1.5 hours, before plating. Data shown are from a single experiment with n=10.

Using this optimized CRISPR system in NHEJ-free *K. marxianus*, we were able to integrate the pta and xpk genes at chromosomal sites IV-1 and IV-2. After integration at these sites, the strains were transformed with the high copy pKD-P2PS plasmid, demonstrating a 1.6-fold

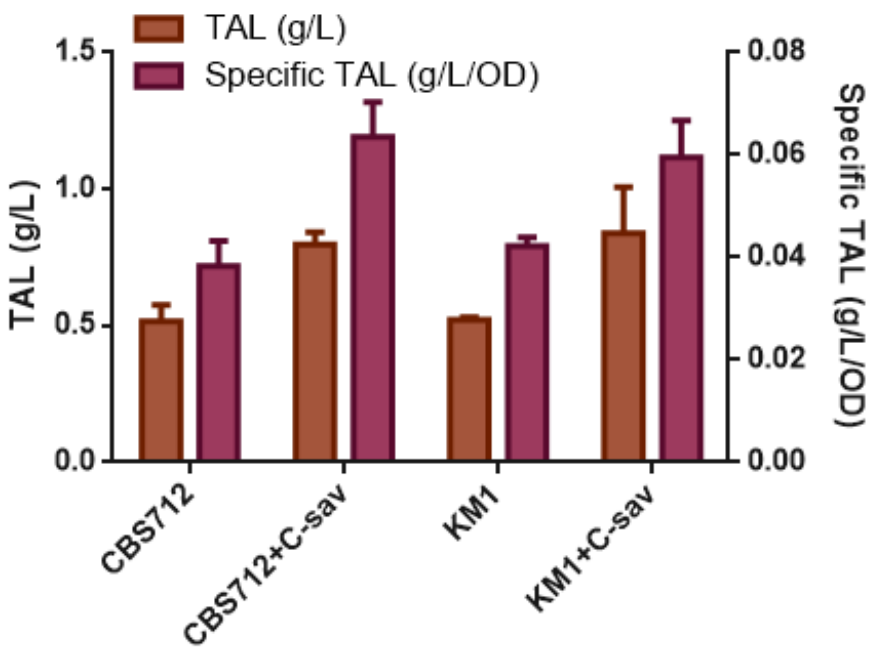


Figure 22 TAL titer (g/L, orange) and Specific Titer (g/L/OD) are reported in *K. marxianus* strains CBS712*Δura3Δku70* and KM1*Δura3Δku70* with and without the integrated C-saving pathway. All strains were transformed with the high copy, pKD-P2PS plasmid and grown in 1% xylose in a water bath shaker at 37C for 48hrs prior to HPLC assay.

improvement in titer (g/L) and specific titer (g/L/OD) in xylose medium in both CBS712 and KM1 strains (**Figure 22**). This promising increase in production may be complemented by additional C-saving pathway gene integrations.

Interestingly, low copy expression of 2-PS (pCA-P2PS) in xylose as well as production from both high and low copy vectors in glucose did not result in significant TAL improvements over the wild-type. The lack of observable increase in TAL production in a low-copy system is likely an indication that 2-PS enzyme availability is the limiting factor in low-copy expression, instead of loss of C-efficiency via CO_2. Furthermore, the lack of improvement in glucose with this system is not unexpected. Since *K. marxianus* makes very little ethanol relative to *S. cerevisiae*, we speculate that the acetyl-CoA that is routed towards TAL in *K. marxianus* is not primarily generated from acetaldehyde but instead from fatty acid catabolism (McTaggart et al. 2019). Therefore, the carbon lost through CO_2 in glucose metabolism is likely less than that of *S. cerevisiae*, and less than CO_2 lost from *K. marxianus* on xylose and glycerol.

Building on this, previous studies have demonstrated significant acetate accumulation is *S. cerevisiae* when C-saving is implemented. In order to test if this is the case in *K. marxianus* as well, we removed the *RHR2* gene responsible for acetate production from acetyl-CoA in the strains with integrated pta and xpk. Removal of *RHR2* did not improve TAL titers with pKD-P2PS high-copy expression (**Figure 23**), indicating that acetate production is either lower or less detrimental to *K. marxianus* than it is to *S. cerevisiae*.

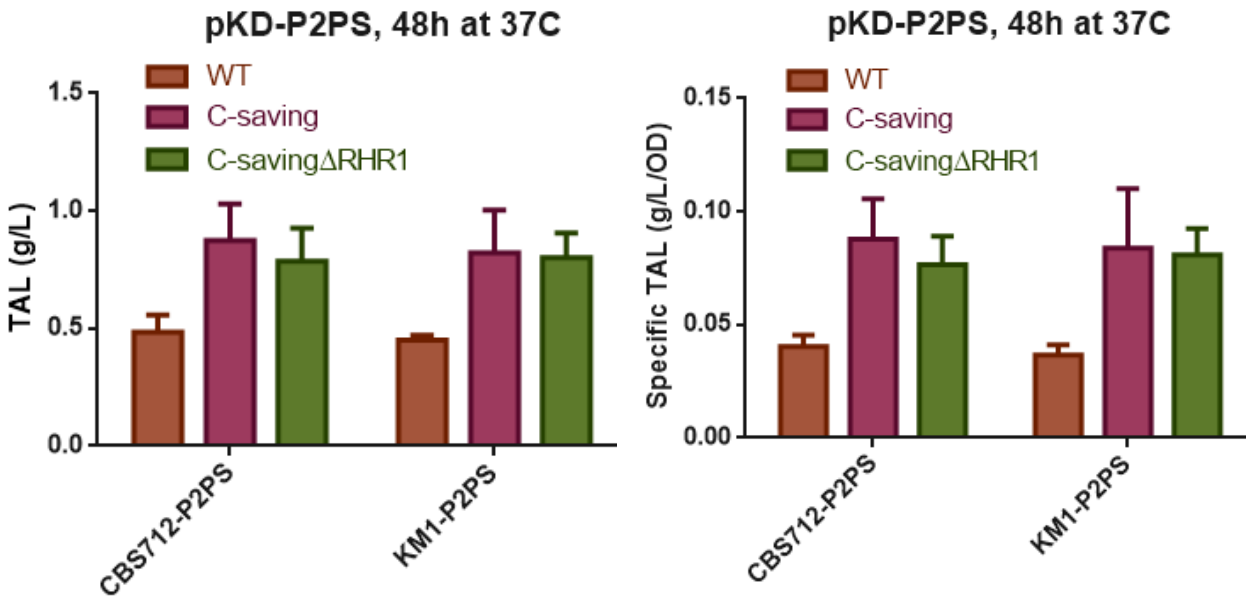


Figure 23 *K. marxianus* strains CBS712Δ*ura3*Δ*ku70* and KM1Δ*ura3*Δ*ku70* with single copies of integrated pta and xpk are grown at 37°C for 48 hr in a gyratory water bath. These are compared to the wild-type (WT) Δ*ura3*Δ*ku70* strain as well as *Δrhr1*. (n=3)

COBRA toolbox modelling for improved TAL titers

Genome scale models contain a list of all known chemical and transport reactions within an organism and can be useful tools for understanding and modelling the basic stoichiometry of an organism's metabolism. Tools such as the COBRA toolbox (Heirendt et al. 2019; Burgard, Pharkya, and Maranas 2003) built in Python enable a process called flux balance analysis (FBA)

that can predict titer and product formation using a set of inputs such as feedstock and oxygen availability. In particular, the OptKnock package utilizes flux balance analysis and genome scale models to systematically remove reactions within a pathway to optimize the desired output – such as the formation of TAL. This system was used to adapt a genome scale model of *K. marxianus* to include the carbon saving pathway genes and then identify pathway genes that may improve the yields of the key TAL precursors, pyruvate and acetyl-CoA.

The tutorial on the COBRA github page (Mendoza, n.d.) for OptKnock was adapted for the *K. marxianus* model iSM996 (Marcišauskas, Ji, and Nielsen 2019). When performing OptKnock, we set the parameters such that final biomass and growth rate could not be reduced below 75%. From this, we varied the objective and carbon source to find knockout targets. We tested glucose as a substrate then varied the objective between pyruvate and acetyl-CoA. From the original model, we also introduced the heterologous enzymes 2-PS, Pta and Xpk into the model using Notepad++ and repeated the optimization. This adapted genome scale model we developed for *K. marxianus* includes the carbon-saving pathway and can be used for prediction of precursor and product formation, as well as predict CO_2 production as well as oxygen and substrate utilization. This model and example python code are provided in **Appendix B**.

Future directions

Future work with this carbon saving pathway should focus on development of the *K. marxianus* platform. Increases of 1.6-fold with single copy integration is promising; however there are many points where this can be refined. First, low copy number may be a factor. To enable higher copy expression of these enzymes, development and screening of new integration targets and gRNAs will be required.

In the literature, acetate accumulation was demonstrated to negatively impact *S. cereivisae* after the carbon saving pathway was introduced (Bergman et al. 2016; Meadows et al. 2016); however this effect was not observed in *K. marxianus*. Nevertheless, this effect indicates that there may be accumulation of acetylphosphate (that is preferrably converted to acetate in *S. cerevisiae*). Improving the activity and conversion rate of the Pta enzyme, which converts acetylphosphate into acetyl-CoA, could improve the rate of TAL production in *K. marxianus* as well as reduce acetate accumulation in *S. cerevisiae*. Previous studies on the activity and essential sites of the *M. thermophilia* Pta enzyme have identified mutations that decrease the K_m by 4x, as well as others that increase k_{Cat} by >2x (Rasche, Smith, and Ferry 1997). These particular mutations at residues R87Q and C277A may improve activity and conversion of acetyl-phosphate, and combinations of these mutations have not yet been tried.

Codon optimization of the pta and xpk genes might also further improve expression in *K. marxianus*. These genes were codon optimized using OPTIMIZER (Puigbò et al. 2007) but were not yet constructed. The closely related *K. lactis* codon usage was used for optimization for *K.*

marxianus, however this should be adjusted as more *K. marxianus* sequences become available. These codon optimized sequences are included in **Appendix C**.

Carbon dioxide is also produced in acetoin byproduct formation during fermentation to ethanol. Reduction of acetoin and ethanol formation via downregulation of *BDH1* and *PDC1/5/6* in *S. cerevisiae* will further reduce CO_2 formation, however extreme throttling of these enzymes is detrimental to growth, cofactor balance and NADH formation during ethanol fermentation (Flikweert et al. 1996, 1999). Despite these detriments, removal of *PDC1* and *PDC5* genes improved pyruvate pools and lactic acid production (ISHIDA et al. 2006). Reduced growth rate due to *PDC* knockout was only observed in *S. cerevisiae* on glucose and was not observed on acetate or ethanol (Tokuhiro et al. 2009). Furthermore, *PDC* knockout coupled with a point mutation in Mth1 contributed to increased 2,3-butanediol titer on glucose in *S. cerevisiae* (S. J. Kim et al. 2013), and partial deletion of the *MTH1* gene further restored glucose growth (Oud et al. 2012). In this work, I've designed gRNA and primers for the removal of the *PDC1/5/6* genes in *S. cerevisiae* and *PDC1* in *K. marxianus*. Implementing these genomic changes, especially when coupled with the *MTH1* mutation in *S. cereivisiae,* might further improve the ability to divert flux through this heterologous pathway.

Furthermore, evaluating the carbon-saving pathway in *K. marxianus* in larger volume culture would also be valuable. In previous experiments in the lab, we've observed *K. marxianus* to need significantly more oxygen than *S. cerevisiae*, such that growth rate in large volumes is oxygen-limited (unpublished results). One benefit overserved with the carbon-saving pathway is

the reduction in oxygen utilization in glucose. Evaluating the growth of this carbon-saving strain in a bioreactor could be promising, and if oxygen needs are truly lesser, this pathway may be essential for production of products on a large scale.

Lastly, in this work a carbon-saving model for *K. marxianus* was designed and built within the COBRA toolbox, however refinements in this model for utilization of xylose as a substrate, as well as production of TAL (as opposed to acetyl-CoA) will need to be evaluated. Further implementation of this modeling tool with OptKnock, OptStoic and OptForce can provide guidance for further metabolic engineering and identification of genetic targets that complement the heterologous carbon-saving pathway.

Conclusions

This study is the first to systematically evaluate polyketide production in metabolically engineered *K. marxianus*, achieving 1.6-fold higher titers than the baseline strain with one copy integration of a heterologous carbon saving pathway. We have demonstrated the introduction of a heterologous pathway that reduces CO_2 loss in the system and coupled this with a robust genome scale model for flux balance analysis simulation. We have also implemented an efficient CRISPR-Cas9 system that expedites metabolic integrations. These developments in a thermotolerant, rapidly growing organism demonstrate the breadth of tools that have now become available for nonconventional organisms and expand our ability to generate high titers of polyketides sustainably and at low-cost.

Acknowledgements

This research was supported by the National Science Foundation: Grant No. CBET-1803677 and EEC-0813570 (through the Engineering Research Center CBiRC (Center for Biorenewable Chemicals)) and the Department of Energy/Office of Science: Grant No. DE-SC0019093. The authors thank Shane Bassett for his introductory *K. marxianus* plasmid screening, Danielle Bever for constructing the *K. marixnaus* CRISPR-Cas9 plasmids and Anupam Chowdhury for the carbon-saving models.

References

ATCC. n.d. “ATCC Medium: 1043.” https://www.atcc.org/~/media/D92466B5E4084EA9AF62938EE401035E.ashx.

Banat, I. M., P. Nigam, and R. Marchant. 1992. “Isolation of Thermotolerant, Fermentative Yeasts Growing at 52-Degree C and Producing Ethanol at 45- Degree C and 50-Degree C.” World Journal of Microbiology & Biotechnology 8 (3): 259–63. https://doi.org/10.1007/BF01201874.

Bergman, Alexandra, Verena Siewers, Jens Nielsen, and Yun Chen. 2016. “Functional Expression and Evaluation of Heterologous Phosphoketolases in Saccharomyces Cerevisiae.” AMB Express 6 (1). https://doi.org/10.1186/s13568-016-0290-0.

Bogorad, Igor W., Tzu Shyang Lin, and James C. Liao. 2013. “Synthetic Non-Oxidative Glycolysis Enables Complete Carbon Conservation.” Nature 502 (7473): 693–97. https://doi.org/10.1038/nature12575.

Branda, Steven S., José Eduardo González-Pastor, Sigal Ben-Yehuda, Richard Losick, and Roberto Kolter. 2001. “Fruiting Body Formation by Bacillus Subtilis.” Proceedings of the National Academy of Sciences of the United States of America. https://doi.org/10.1073/pnas.191384198.

Burgard, Anthony P., Priti Pharkya, and Costas D. Maranas. 2003. “OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain

Optimization." Biotechnology and Bioengineering 84 (6): 647–57. https://doi.org/10.1002/bit.10803.

Cardenas, Javier, and Nancy A. Da Silva. 2016. "Engineering Cofactor and Transport Mechanisms in Saccharomyces Cerevisiae for Enhanced Acetyl-CoA and Polyketide Biosynthesis." Metabolic Engineering 36: 80–89. https://doi.org/10.1016/j.ymben.2016.02.009.

Chia, Mei, Thomas J. Schwartz, Brent H. Shanks, and James A. Dumesic. 2012. "Triacetic Acid Lactone as a Potential Biorenewable Platform Chemical." Green Chemistry 14 (7): 1850. https://doi.org/10.1039/c2gc35343a.

Chowdhury, Anupam, and Costas D. Maranas. 2015. "Designing Overall Stoichiometric Conversions and Intervening Metabolic Reactions." Scientific Reports 5. https://doi.org/10.1038/srep16009.

Deaner, Matthew, Julio Mejia, and Hal S. Alper. 2017. "Enabling Graded and Large-Scale Multiplex of Desired Genes Using a Dual-Mode DCas9 Activator in Saccharomyces Cerevisiae." ACS Synthetic Biology 6: 1931–43. https://doi.org/10.1021/acssynbio.7b00163.

Dudášová, Zuzana, Andrej Dudáš, and Miroslav Chovanec. 2004. "Non-Homologous End-Joining Factors of Saccharomyces Cerevisiae." FEMS Microbiology Reviews. https://doi.org/10.1016/j.femsre.2004.06.001.

Dymond, Jessica S. 2013. "Preparation of Genomic DNA from Saccharomyces Cerevisiae." In Methods in Enzymology. https://doi.org/10.1016/B978-0-12-418687-3.00012-4.

Evans, Christopher Thomas, and Colin Ratledge. 1984. “Induction of Xylulose-5-Phosphate Phosphoketolase in a Variety of Yeasts Grown on d-Xylose: The Key to Efficient Xylose Metabolism.” Archives of Microbiology 139 (1): 48–52. https://doi.org/10.1007/BF00692711.

Fang, Fang, Kirsty Salmon, Michael W Y Shen, Kimberly a Aeling, Elaine Ito, Becky Irwin, Uyen Phuong C Tran, G Wesley Hatfield, Nancy A. Da Silva, and Suzanne Sandmeyer. 2011. “A Vector Set for Systematic Metabolic Engineering in Saccharomyces Cerevisiae.” Yeast 28 (2): 123–36. https://doi.org/10.1002/yea.1824.

Flikweert, Marcel T., Martin De Swaaf, Johannes P. Van Dijken, and Jack T. Pronk. 1999. “Growth Requirements of Pyruvate-Decarboxylase-Negative Saccharomyces Cerevisiae.” FEMS Microbiology Letters 174 (1): 73–79. https://doi.org/10.1016/S0378-1097(99)00124-X.

Flikweert, Marcel T., Linda Van Der Zanden, Wouter M.Th M. Janssen, H. Yde Steensma, Johannes P. Van Dijken, and Jack T. Pronk. 1996. “Pyruvate Decarboxylase: An Indispensable Enzyme for Growth of Saccharomyces Cerevisiae on Glucose.” Yeast 12 (3): 247–57. https://doi.org/10.1002/(SICI)1097-0061(19960315)12:3<247::AID-YEA911>3.0.CO;2-I.

Gao, Jiaoqi, Wenjie Yuan, Yimin Li, Ruijuan Xiang, Shengbo Hou, Shijun Zhong, and Fengwu Bai. 2015. “Transcriptional Analysis of Kluyveromyces Marxianus for Ethanol Production

from Inulin Using Consolidated Bioprocessing Technology." Biotechnology for Biofuels. https://doi.org/10.1186/s13068-015-0295-y.

Gietz, Daniel, Andrew St Jean, Robin A. Woods, and Robert H. Schiestl. 1992. "Improved Method for High Efficiency Transformation of Intact Yeast Cells." Nucleic Acids Research 20 (6): 1425.

Groeneveld, Philip, Adriaan H. Stouthamer, and Hans V. Westerhoff. 2009. "Super Life - How and Why 'cell Selection' Leads to the Fastest-Growing Eukaryote." FEBS Journal 276 (1): 254–70. https://doi.org/10.1111/j.1742-4658.2008.06778.x.

Haber, James E. 1995. "In Vivo Biochemistry: Physical Monitoring of Recombination Induced by Site-specific Endonucleases." BioEssays. https://doi.org/10.1002/bies.950170707.

Heirendt, Laurent, Sylvain Arreckx, Thomas Pfau, Sebastián N. Mendoza, Anne Richelle, Almut Heinken, Hulda S. Haraldsdóttir, et al. 2019. "Creation and Analysis of Biochemical Constraint-Based Models Using the COBRA Toolbox v.3.0." Nature Protocols. https://doi.org/10.1038/s41596-018-0098-2.

Hua, Yan, Jichao Wang, Yelin Zhu, Biao Zhang, Xin Kong, Wenjie Li, Dongmei Wang, and Jiong Hong. 2019. "Release of Glucose Repression on Xylose Utilization in Kluyveromyces Marxianus to Enhance Glucose-Xylose Co-Utilization and Xylitol Production from Corncob Hydrolysate." Microbial Cell Factories. https://doi.org/10.1186/s12934-019-1068-2.

ISHIDA, Nobuhiro, Satoshi SAITOH, Toru ONISHI, Kenro TOKUHIRO, Eiji NAGAMORI, Katsuhiko KITAMOTO, and Haruo TAKAHASHI. 2006. “The Effect of Pyruvate Decarboxylase Gene Knockout in Saccharomyces Cerevisiae on L -Lactic Acid Production.” Bioscience, Biotechnology, and Biochemistry 70 (5): 1148–53. https://doi.org/10.1271/bbb.70.1148.

Juergens, Hannes, Javier A Varela, Arthur R Gorter de Vries, Thomas Perli, Veronica J M Gast, Nikola Y Gyurchev, Arun S Rajkumar, et al. 2018. “Genome Editing in Kluyveromyces and Ogataea Yeasts Using a Broad-Host-Range Cas9/GRNA Co-Expression Plasmid.” FEMS Yeast Research, no. February: 1–16. https://doi.org/10.1093/femsyr/foy012.

Kim, Soo Jung, Seung Oh Seo, Yong Su Jin, and Jin Ho Seo. 2013. “Production of 2,3-Butanediol by Engineered Saccharomyces Cerevisiae.” Bioresource Technology 146: 274–81. https://doi.org/10.1016/j.biortech.2013.07.081.

King, Zachary A., and Adam M. Feist. 2014. “Optimal Cofactor Swapping Can Increase the Theoretical Yield for Chemical Production in Escherichia Coli and Saccharomyces Cerevisiae.” Metabolic Engineering 24: 117–28. https://doi.org/10.1016/j.ymben.2014.05.009.

Labuhn, Maurice, Felix F. Adams, Michelle Ng, Sabine Knoess, Axel Schambach, Emmanuelle M. Charpentier, Adrian Schwarzer, Juan L. Mateo, Jan Henning Klusmann, and Dirk Heckl. 2018. “Refined SgRNA Efficacy Prediction Improves Largeand Small-Scale CRISPR-Cas9 Applications.” Nucleic Acids Research. https://doi.org/10.1093/nar/gkx1268.

Lawrence, Sarah H., and James G. Ferry. 2006. “Steady-State Kinetic Analysis of Phosphotransacetylase from Methanosarcina Thermophila.” Journal of Bacteriology 188 (3): 1155–58. https://doi.org/10.1128/JB.188.3.1155-1158.2006.

Lee, Michael E., William C. DeLoache, Bernardo Cervantes, and John E. Dueber. 2015. “A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly.” ACS Synthetic Biology 4 (9): 975–86. https://doi.org/10.1021/sb500366v.

Lee, Wonkyu, and Nancy A. DaSilva. 2006. “Application of Sequential Integration for Metabolic Engineering of 1,2-Propanediol Production in Yeast.” Metabolic Engineering. https://doi.org/10.1016/j.ymben.2005.09.001.

Löbs, Ann Kathrin, Ronja Engel, Cory Schwartz, Andrew Flores, and Ian Wheeldon. 2017. “CRISPR-Cas9-Enabled Genetic Disruptions for Understanding Ethanol and Ethyl Acetate Biosynthesis in Kluyveromyces Marxianus.” Biotechnology for Biofuels 10 (1). https://doi.org/10.1186/s13068-017-0854-5.

Lu, Hongzhong, Feiran Li, Benjamín J. Sánchez, Zhengming Zhu, Gang Li, Iván Domenzain, Simonas Marcišauskas, et al. 2019. “A Consensus S. Cerevisiae Metabolic Model Yeast8 and Its Ecosystem for Comprehensively Probing Cellular Metabolism.” Nature Communications. https://doi.org/10.1038/s41467-019-11581-3.

Lundie, L. L., and J. G. Ferry. 1989. “Activation of Acetate by Methanosarcina Thermophila. Purification and Characterization of Phosphotransacetylase.” Journal of Biological Chemistry 264 (31): 18392–96.

Marcišauskas, Simonas, Boyang Ji, and Jens Nielsen. 2019. “Reconstruction and Analysis of a Kluyveromyces Marxianus Genome-Scale Metabolic Model.” BMC Bioinformatics. https://doi.org/10.1186/s12859-019-3134-5.

McTaggart, Tami L., Danielle Bever, Shane Bassett, and Nancy A. Da Silva. 2019. “Synthesis of Polyketides from Low Cost Substrates by the Thermotolerant Yeast Kluyveromyces Marxianus.” Biotechnology and Bioengineering. https://doi.org/10.1002/bit.26976.

Meadows, Adam L., Kristy M. Hawkins, Yoseph Tsegaye, Eugene Antipov, Youngnyun Kim, Lauren Raetz, Robert H. Dahl, et al. 2016. “Rewriting Yeast Central Carbon Metabolism for Industrial Isoprenoid Production.” Nature 537 (7622): 694–97. https://doi.org/10.1038/nature19769.

Meile, L., L. M. Rohr, T. A. Geissmann, M. Herensperger, and M. Teuber. 2001. “Characterization of the D-Xylulose 5-Phosphate/D-Fructose 6-Phosphate Phosphoketolase Gene (Xfp) from Bifidobacterium Lactis.” Journal of Bacteriology 183 (9): 2929–36. https://doi.org/10.1128/JB.183.9.2929-2936.2001.

Mendoza, Sebastián N. n.d. “OptKnock Tutorial.” https://opencobra.github.io/cobratoolbox/stable/tutorials/tutorialOptKnock.html.

Mikkelsen, Michael Dalgaard, Line Due Buron, Bo Salomonsen, Carl Erik Olsen, Bjarne Gram Hansen, Uffe Hasbro Mortensen, and Barbara Ann Halkier. 2012. “Microbial Production of Indolylglucosinolate through Engineering of a Multi-Gene Pathway in a Versatile Yeast

Expression Platform." Metabolic Engineering 14 (2): 104–11. https://doi.org/10.1016/j.ymben.2012.01.006.

Nambu-Nishida, Yumiko, Keiji Nishida, Tomohisa Hasunuma, and Akihiko Kondo. 2017. "Development of a Comprehensive Set of Tools for Genome Engineering in a Cold- A Nd Thermo-Tolerant Kluyveromyces Marxianus Yeast Strain." Scientific Reports 7 (1). https://doi.org/10.1038/s41598-017-08356-5.

Oud, Bart, Carmen Lisset Flores, Carlos Gancedo, Xiuying Zhang, Joshua Trueheart, Jean Marc Daran, Jack T. Pronk, and Antonius J A van Maris. 2012. "An Internal Deletion in MTH1 Enables Growth on Glucose of Pyruvate-Decarboxylase Negative, Non-Fermentative Saccharomyces Cerevisiae." Microbial Cell Factories 11. https://doi.org/10.1186/1475-2859-11-131.

Puigbò, Pere, Eduard Guzmán, Antoni Romeu, and Santiago Garcia-Vallvé. 2007. "OPTIMIZER: A Web Server for Optimizing the Codon Usage of DNA Sequences." Nucleic Acids Research. https://doi.org/10.1093/nar/gkm219.

Rado, Thomas A., and James A. Hoch. 1973. "Phosphotransacetylase from Bacillus Subtilis: Purification and Physiological Studies." BBA - Enzymology 321 (1): 114–25. https://doi.org/10.1016/0005-2744(73)90065-X.

Rajkumar, Arun S., Javier A. Varela, Hannes Juergens, Jean Marc G. Daran, and John P. Morrissey. 2019. "Biological Parts for Kluyveromyces Marxianus Synthetic Biology." Frontiers in Bioengineering and Biotechnology. https://doi.org/10.3389/fbioe.2019.97.

Rasche, Madeline E., Kerry S. Smith, and James G. Ferry. 1997. “Identification of Cysteine and Arginine Residues Essential for the Phosphotransacetylase from Methanosarcina Thermophila.” Journal of Bacteriology 179 (24): 7712–17. https://doi.org/10.1128/jb.179.24.7712-7717.1997.

Ryan, Owen W., Jeffrey M. Skerker, Matthew J. Maurer, Xin Li, Jordan C. Tsai, Snigdha Poddar, Michael E. Lee, et al. 2014. “Selection of Chromosomal DNA Libraries Using a Multiplex CRISPR System.” ELife 3 (August2014): 1–15. https://doi.org/10.7554/eLife.03703.

Sakihama, Yuri, Ryota Hidese, Tomohisa Hasunuma, and Akihiko Kondo. 2019. “Increased Flux in Acetyl-CoA Synthetic Pathway and TCA Cycle of Kluyveromyces Marxianus under Respiratory Conditions.” Scientific Reports. https://doi.org/10.1038/s41598-019-41863-1.

Schabort, Du Toit W.P., Precious Letebele, Laurinda Steyn, Stephanus G. Kilian, and James C. Du Preez. 2016. “Differential RNA-Seq, Multi-Network Analysis and Metabolic Regulation Analysis of Kluyveromyces Marxianus Reveals a Compartmentalised Response to Xylose.” PLoS ONE. https://doi.org/10.1371/journal.pone.0156242.

Shin, Byung Sik, Soo Keun Choi, and Seung Hwan Park. 1999. “Regulation of the Bacillus Subtilis Phosphotransacetylase Gene.” Journal of Biochemistry 126 (2): 333–39. https://doi.org/10.1093/oxfordjournals.jbchem.a022454.

Sonderegger, Marco, Michael Schümperli, and Michael Schu. 2004. “Metabolic Engineering of a Phosphoketolase Pathway for Pentose Catabolism in Saccharomyces Cerevisiae

Metabolic Engineering of a Phosphoketolase Pathway for Pentose Catabolism in Saccharomyces Cerevisiae." Applied and Environmental Microbiology 70 (5): 2892–97. https://doi.org/10.1128/AEM.70.5.2892.

Stemmer, Manuel, Thomas Thumberger, Maria Del Sol Keyer, Joachim Wittbrodt, and Juan L. Mateo. 2015. "CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool." PLoS ONE. https://doi.org/10.1371/journal.pone.0124633.

Tokuhiro, Kenro, Nobuhiro Ishida, Eiji Nagamori, Satoshi Saitoh, Toru Onishi, Akihiko Kondo, and Haruo Takahashi. 2009. "Double Mutation of the PDC1 and ADH1 Genes Improves Lactate Production in the Yeast Saccharomyces Cerevisiae Expressing the Bovine Lactate Dehydrogenase Gene." Applied Microbiology and Biotechnology 82 (5): 883–90. https://doi.org/10.1007/s00253-008-1831-5.

Varela, Javier A., Martina Puricelli, Noemi Montini, and John P. Morrissey. 2019. "Expansion and Diversification of MFS Transporters in Kluyveromyces Marxianus." Frontiers in Microbiology. https://doi.org/10.3389/fmicb.2018.03330.

Veiga-Da-Cunha, M., H. Santos, and E. Van Schaftingen. 1993. "Pathway and Regulation of Erythritol Formation in Leuconostoc Oenos." Journal of Bacteriology 175 (13): 3941–48. https://doi.org/10.1128/jb.175.13.3941-3948.1993.

Zhang, Yiming, Guodong Liu, Martin K.M. Engqvist, Anastasia Krivoruchko, Björn M. Hallström, Yun Chen, Verena Siewers, and Jens Nielsen. 2015. "Adaptive Mutations in Sugar

Metabolism Restore Growth on Glucose in a Pyruvate Decarboxylase Negative Yeast Strain.” Microbial Cell Factories. https://doi.org/10.1186/s12934-015-0305-6.

Chapter 4

Development of a novel CRISPRi growth screen for selection of malonyl-CoA overproducers

Abstract

Rationally and computationally predicted metabolic engineering approaches have been successful in the past to increase yeast-based production of polyketides; however it is often difficult to find ideal gene knockdown combinations in a streamlined and high-throughput manner. To address current bottlenecks, CRISPR-based libraries can be used to screen for desirable properties. Further pairing of these large libraries with metabolite sensors can improve throughput. In this work, we coupled a CRISPRi library method with a growth-paired malonyl-CoA sensor to improve the ability to screen and select desirable knockdowns in large, pooled cultures. This process has an advantage over single-modification isolated cultures, as we are not required to maintain, culture and assay a large number of unique cultures separately – a process which can be time, space and material prohibitive. We conducted a proof-of-concept pilot study in *Saccharomyces cerevisiae* and then extended and developed the approach for the thermotolerant, fast-growing yeast *Kluyveromyces marxianus* using a full-coverage functional gene gRNA library. The population of this enriched library provides insight into novel gene knockouts and combinations for future development and optimization for malonyl-CoA based products. This versatile process can be adapted into a range of strains and organisms for the discovery of cryptic pathway bottlenecks in strain engineering.

Introduction

Production of many biobased chemicals in yeast is dependent on two key metabolic building blocks – acetyl-CoA and malonyl-CoA. These two metabolites are essential in a number of cellular processes and are the first steps in fatty acid biosynthesis. Increasing malonyl-CoA is particularly useful for complex polyketides that require many malonyl-CoA extender molecules. For example, lovastatin, a valuable cholesterol lowering drug, requires 8 malonyl-CoA extender molecules for every acetyl-CoA molecule (Chan et al. 2009).

Since acetyl-CoA is converted into malonyl-CoA by acetyl-CoA carboxylase (Acc1), interventions that increase acetyl-CoA can also increase malonyl-CoA. In addition, extensive work on mutating the *ACC1* gene as well as introducing heterologous *ACC1* genes have resulted in higher conversion and substantial increases in production of a variety of desirable chemicals (Choi and Da Silva 2014; X. Li et al. 2014; Shi et al. 2014; Chen et al. 2018). In addition to improving conversion to malonyl-CoA via *ACC1*, other groups have increased malonyl-CoA pools by manipulating transcriptional regulators of fatty acid synthesis (Chen et al. 2017).

Malonyl-CoA is readily used within the cell as an elongating molecule during fatty acid synthesis; therefore, quantifying intracellular malonyl-CoA is difficult to do in a dynamic or high-throughput manner. Techniques for measuring malonyl-CoA include LC-MS/MS (HAYASHI and SATOH 2006) and colorimetric kits (MyBioSource); however these methods require testing of samples individually which can be slow and also sensitive to cell lysing methods. Recently, researchers have taken advantage of an operator/repressor set that interacts with malonyl-CoA and can act as an indirect sensor of free malonyl-CoA (S. Li et al. 2015; David, Nielsen, and Siewers

2016). A transcription factor native to *Bacillus subtilis*, FapR, regulates fatty acid synthesis (Diomandé et al. 2015) by binding to an operator, FapO, in the promoter upstream of the *FabHAF* operon (**Figure 24**). The *FabHAF* operon codes for the enzymes FabHA and FabF, both of which are β-ketoacyl-ACP synthases responsible for β-ketoacyl-ACP formation. The FapR transcription factor binds to malonyl-CoA in addition to FapO; however, in the presence of increased malonyl-CoA it does not bind to FapO and transcription of the *FabHAF* operon occurs, producing fatty acids. When malonyl-CoA is low, transcription of the *fab* genes is blocked and fatty acid synthesis does not occur (Schujman et al. 2003).

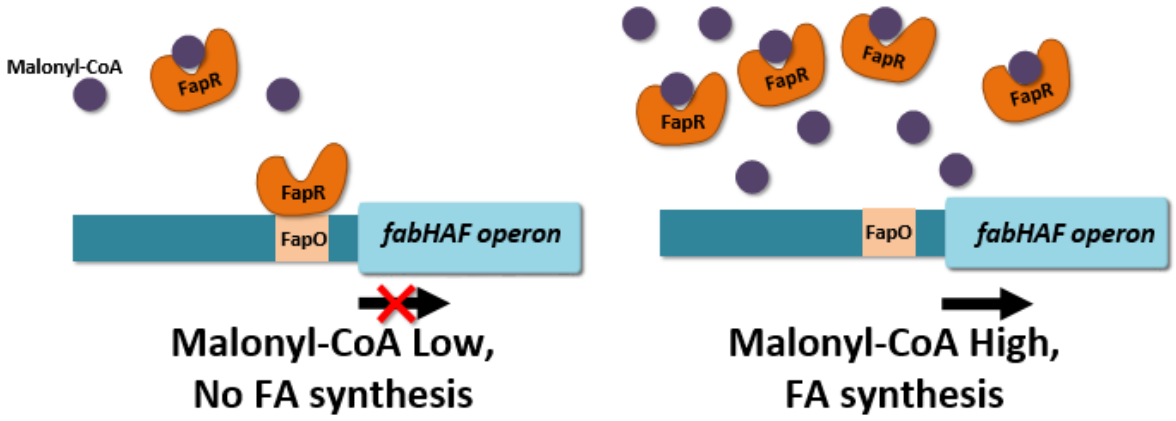


Figure 24 Native *Bacillus subtilis* FapR regulation of the fabHAF operon, responsible for β-ketoacyl-ACP formation and fatty acid synthesis, in both high and low malonyl-CoA conditions

This FapR/FapO system was codon-optimized and successfully utilized in *Saccharomyces cerevisiae* for screening of high malonyl-CoA producing strains and applied for the overproduction of 3-hydroxypropionic acid (3-HPA) (S. Li et al. 2015). The FapO region was cloned into the promoter controlling a fluorescent reporter, tdTomato, such that stains with increased malonyl-CoA would result in higher fluorescence. This sensor was further optimized by

interspersing 3 copies of FapO within the *TEF1* promoter (termed TEF123), which resulted in a more precise response in fluorescent protein production. Application led to further increases in 3-HPA formation (David, Nielsen, and Siewers 2016); however, single-cell screening was required to achieve these results.

Another new tool for metabolic engineering is the development of large knockout or knockdown libraries generated using CRISPR-Cas9. Due to plummeting costs for sequencing and oligo synthesis, it is now easier than ever to synthetically construct large plasmid sets that can target a wide range of genes or genomic regions. Traditional CRISPR-Cas9 strain modification libraries can be implemented for gene knockouts in organisms with nonhomologous end joining (NHEJ), but the rate of knockout is still dependent on random mutation at the cut site to a stop codon, and still provides a chance for up-regulation in case of a favorable mutation. In organisms without NHEJ but very good homologous recombination (HR) machinery (e.g., *S. cerevisiae*), we can instead provide the organism with a "repair donor" that provides a template for repairing the Cas9 cut in a guided manner. Unfortunately, making repair donors for each cut within a very large library can be impractical and limited by the efficiency of transformation of both the correct gRNA and repair donor in one cell.

One way to avoid the need for repair donors altogether is by design of a deactivated Cas9 (dCas9) enzyme that targets and binds to a specific genomic region but does not cleave the DNA (Jinek et al. 2012). These dCas9 proteins will bind and prevent efficient transcription of the region surrounding the dCas9 binding site independent of the host's recombination machinery, but will otherwise not reduce cell viability, thereby making it possible to screen downregulation of a large

system of genetic targets. Recent improvements to the dCas9 enzyme include fusion to the *MxiI* domain, which results in decreased expression of the binding region relative to the standard dCas9 (Gilbert et al. 2013).

In this work, we systematically pair the FapR/FapO-based malonyl-CoA sensor with growth selection by the integration of the FapO upstream of a key feedstock metabolism gene, preventing the need for single-cell sorting. We further extend this tool with a dCas9-based genome-scale library that can be used to generate a set of strains to be screened for improved malonyl-CoA production using our sensor. This system was designed for two yeasts that have demonstrated success for the production of malonyl-CoA based products: traditional Baker's yeast *S. cerevisiae* and a nonconventional, thermotolerant yeast *Kluyveromyces marxianus*. Development in these organisms provides a new tool for enabling rapid identification of complementary strain interventions, which has been a key bottleneck for improving production. A coupled sensor-repression method can be easily transferred to and expanded for other organisms or sensors, which can be useful in both the discovery as well as refining steps of strain engineering.

Materials and methods

Strains and cultivation

Escherichia coli strain DH5α (Invitrogen, Carlsbad, CA) was used for plasmid maintenance and amplification. The *K. marxianus* strains employed were CBS 712 (ATCC 200963; ATCC®, Manassas, VA), and KM1Δ*ura3* (Pecota, Rajgarhia, and Da Silva 2007). In CBS 712 and KM1, we

removed the *URA3* and/or *KU70* genes as previously described (**Chapter 3**). The *S. cerevisiae* strain used was BY4741Δtrp1 (Open Biosystems, Huntsville, AL). All three yeast strains were then used for the integration of FapO, FapR and dCas9 genes.

E. coli strains were cultivated in Luria-Bertani media (LB) at 37°C in an air shaker at 220 RPM (New Brunswick Scientific Co. Excella E25, Edison, NJ). Ampicillin (150 mg/L) was supplemented to select for plasmids. Prior to transformation, both *S. cerevisiae* and *K. marxianus* were cultivated in 1% YPD (1% Bacto yeast extract, 2% Bacto peptone, 1% dextrose) at 30°C in an air shaker at 250 RPM. The transformed yeast strains were cultivated on 1% SD (1% dextrose, 0.67% yeast nitrogen base, 0.5% ammonium sulfate, 1% bacto agar and 100 mg/L adenine sulfate) or SDC(A) medium (1% dextrose, 0.67% yeast nitrogen base, 0.5% ammonium sulfate, 0.5% casamino acids, and 100 mg/L adenine sulfate) depending on auxotrophic needs.

Transformation

Single vectors constructed using standard cloning methods were transformed into chemically competent *E. coli* using a 45-second heat shock at 42°C followed by a 2 h recovery in SOC (20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 0.19 g KCl, 0.95 g MgCl2, and 3.6 g glucose per liter). Plasmid libraries, which require significantly higher transformation efficiency, were transformed via electroporation. Electrocompetent *E. coli* were prepared by centrifuging 10 mL exponentially growing cells (~OD600=1), washing twice with 10 mL ice-cold water, washing twice with 10 mL ice-cold 10% glycerol and then resuspending to a total of 400 uL. 500 ng of drop-dialyzed plasmid DNA was added to the electrocompetent cells, aliquoted into a 0.2 cm

electroporation cuvette and pulsed at 2.1 kV, 100 Ω, and 25 µF in a BTX Electro cell manipulator 600 (BTX, Holliston, Massachusetts), and then recovered in SOC for 30 min. An aliquot was plated on LB-Amp to determine efficiency and the rest was grown in liquid.

K. marxianus cultures were grown overnight in 3 mL of 2% YPD or 2% SDC(A) at 30°C. Cultures were then reinoculated in fresh media and allowed to grow to OD 2.0. Transformation of plasmids was conducted using the Zymo EZ Freeze Yeast Transformation II Kit (Zymo Research, Irvine, CA). The protocol was modified slightly in the case of transformations for gene disruptions and gene integrations: 10 uL sheared salmon sperm DNA (ref) was added after the EZ Solution II, and the transformations were incubated at 30°C for 1.5 hours, instead of the prescribed 45 minutes, before plating onto selective media.

S. cerevisiae cultures were grown overnight in 5 mL of 2% SDC(A) at 30°C and then reinoculated in fresh media to OD_{600}=0.3. Transformation was performed using the standard protocol for the Zymo EZ Freeze Yeast Transformation II Kit (Zymo Research, Irvine, CA).

Electrocompetent cells for both *S. cerevisiae* and *K. marxianus* were prepared by spinning down 100 mL overnight cultures (in YPD at OD_{600}~20) gently at 3000 rpm for 3 min, washing twice in 50 mL ice cold water, and washing once with ice cold electroporation buffer (1 M Sorbitol, 1 mM $CaCl_2$), resuspending in 20 mL of LiAc solution (0.1 M LiAc, 10 mM DTT) and incubating at 30°C for 20 mins. The cells were then washed again with 50 mL ice cold electroporation buffer, spun down, and resuspended in a final volume of 1 mL electroporation buffer (Benatuil et al. 2010).

1 ug of drop-dialyzed plasmid DNA was added to the electrocompetent cells, aliquoted into a 0.2 cm electroporation cuvette and pulsed at 2 kV, 25 μF for 3.0 to 4.5, and then recovered in YPD for 1 hr.

Sensor and strain construction

In *S. cerevisiae,* the genes FapR and FapO were integrated into the genome using CRISPR-Cas9. The 794 bp FapR region was synthesized as a gBlock from IDT (**Appendix D**) and Gibson assembled into pUC18, an integrating vector with 100 bp homology on each side of the *URA3* locus, to generate pUC18-FapRNLS (**Table 21**). This plasmid was used as a template for amplification of both FapR and the backbone with overhangs on either side of the *TEF1* promoter and *ADH1* terminator. The *TEF1* promoter was amplified from pKUTP-6MN and the *ADH1* terminator from pTDH3-dCas9-Mxi1. These four pieces were Gibson assembled into pTFapR, an integrating plasmid with the TEF1p-FapR-ADH1t cassette flanked by 100 bp of the *URA3* loci. This plasmid was digested with NdeI/EcoRI to excise the donor and then 2 ug of this donor DNA was transformed in BY4741 alongside Cas9 and a gRNA targeting the *URA3* locus.

The *TEF1* promoter interspersed with three FapO regions (TEF123) (David, Nielsen, and Siewers 2016) was synthesized as a 639 bp gBlock (**Appendix D**) that included homology to the start of *ADH2*. This was PCR amplified and then Gibson assembled into the backbone of pXP822 to form pMSAv2, a plasmid with 100 bp homology up- and down-stream of the *ADH2* promoter. This fragment was digested with NruI and SwaI and 2 ng of this linear piece was transformed alongside an *ADH2*-targeting gRNA and Cas9 into the strain with verified FapR integration.

Lastly, the dCas9-MxiI region was PCR amplified using Q5 polymerase from pTDH3-dCas9-Mxi1 (Addgene) and was integrated into BY4741 ura3::TEF1p-fapR adh2p::TEF123 at the *LEU2* locus using a *LEU2*-specific gRNA plasmid and Cas9 to form BY4741 ura3::TEF1p-fapR adh2p::TEF123 leu2::dCas9-MxiI (BY-FapO/R-dcas9).

In *K. marxianus* the FapR gene was PCR amplified from pTFapR with Q5 polymerase using primers with homology to the *HIS3* gene (**Table 22**). This fragment was transformed into CBS712 and KM1 alongside KM_HIS3_gRNA to form CBS712-FapR and KM1-FapR. The TEF123 gene was then amplified with homology up- and downstream of the *XYL1* promoter region (500bp upstream of *XYL1*) with Q5 polymerase and then transformed alongside KM_XYL1_gRNA into CBS712-FapR and KM1-FapR to form CBS712-FapR-FapO and KM1-FapR-FapO. Lastly, a codon optimized dCas9 was generated from the Cas9 in piW t-sg-r through the introduction of two point mutations D10A and H840A (Jinek et al. 2012) using overlap extension PCR. dCas9 was cloned into the piW t-sg-r backbone to replace the Cas9 and this plasmid was used to create donor DNA for integration into site IV-1. Transformation of this donor alongside KM_IV-1_gRNA generate strains CBS712-FapR-FapO-dCas9 and KM1-FapR-FapO-dCas9.

All donor DNA fragments were purified via gel electrophoresis on a 1% agarose gel and then excised. The DNA fragment was extracted from the excised gel slice using the Zymo Gel Purification Kit, and the eluted DNA was quantified via Nanodrop. All integrations were verified by PCR with one primer internal to the integration donor and one external to the locus

(internal/external check) and then verified via Sanger sequencing (Genewiz) from PCR products using primers external to the donor (external/external).

Enrichment growth screen

Strains were transformed with the plasmid library via electroporation and recovered in 2 mL YPD for 2 hrs. An aliquot of this recovered library was stored at -80°C while the rest were inoculated directly into enrichment medium. For *S. cerevisiae*, we added 500 ul of the recovered transformation into 3 mL 1% SEC(A) (1% ethanol, 0.67% yeast nitrogen base, 0.5% ammonium sulfate, 0.5% casamino acids, and 100 mg/L adenine sulfate) as enrichment medium, choosing 3 mL to improve oxygen transfer. *K. marxianus* will be recovered in YPD and then inoculated into SXC(A). The next day, an aliquot of this culture will be stored and then reinoculated to OD_{600}=0.2 in 3 mL fresh medium.

The *S. cerevisiae* culture was periodically observed to evaluate growth rate, and once the culture approached an OD_{600} of 4-6, the culture was reinoculated to OD_{600}=0.2 and the remaining volume frozen in 20% glycerol at -80C. As the culture began to grow more quickly due to the enrichment, the time it took to reach this density shortened gradually from 4 days to overnight. Once the growth rate neared the growth rate of the wild type culture, a final sample was taken and frozen.

Table 21 FapR and FapO primers and plasmids for *S. cereivisiae* integration

Primer Name	Sequence	Description
pUC18_bb_Fwd	TTCATGGCCTTTATAAAAAGGAACTATCCAATACCTCGCCgaattcgtaatcatggtcat	amplification of pUC18 for Gibson of FapR gBlock with *URA3* overhangs
pUC18_bb_Rev	CATATATAGAATTACATTATTTATGAAATATCATCACTATaagcttggcactggccgtcg	
pUC18-FapR_bb_Fwd	AAAGGAGTAGAAACATTTTGGGAATCTCGGTCGTAATGAT	amplification of pUC18-FapR to Gibson in promoter and terminator
pUC18-FapR_bb_Rev	agcaaatgcctgcaaatcgctccTTCTTTGATGGTCAAAACTTAT	
TEF1p_fwd	TTTCTCTTCTTCTTAGGCATTTTGTAATTAAAACTTAGAT	amplification of *TEF1p* to Gibson upstream of FapR
TEF1p_Rev	ATCATTACGACCGAGATTCCCAAAATGTTTCTACTCCTTT	
ADH1t_Fwd	ATAAGTTTTGACCATCAAAGAAggagcgatttgcaggcatttgct	amplification of ADH1t to Gibson downstream of FapR
ADH1t_Rev	ACAGATCAAAACACTCCTAAgcgaatttcttatgatttat	
FapR_Fwd	ataaatcataagaaattcgcTTAGGAGTGTTTTGATCTGT	amplification of *URA3*-flanked FapR from pUC18-FapR
FapR_Rev	ATCTAAGTTTTAATTACAAAATGCCTAAGAAGAAGAGAAA	
TEF123_Fwd	AGGGTGTCGTTAATTACCCGTACTAAAGGTTTGGAAAAG	amplification of TEF123 for Gibson into pXP822
TEF123_Rev	gtgagctgatacGTTTAAACTGGCAGACACCAGAGTACTT	
pXP822_Fwd	CTTTTCCAAACCTTTAGTACGGGTAATTAACGACACCCT	amplification pXP822 backbone for TEF123 addition
pXP822_Rev	AAGTACTCTGGTGTCTGCCAGTTTAAACgtatcagctcac	
p426_bb_Fwd	gcaagttaaaataaggctagtccg	amplification of the p426 backbone for Gibson assembly with oPools
p426_bb_Rev	catttatctttcactgcggagaag	

Plasmid Name	Description	Source
pUC18	AmpR, lacZ	Addgene
pUC18-FapRNLS	AmpR, FapR	This study
pTFapR	TEF1p-FapR-ADH1t flanked with *URA3*, AmpR	This study
pTDH3-dCas9-Mxi1	pTDH3-dCas9-Mxi1	Addgene
pXP822	2µ, *URA3*, AmpR	Shen and Da Silva, 2012
pMSAv2	CEN/ARS, TEF123, *MET15*, dCas9, AmpR	This study
p426	p426-SNR52p-gRNA.CAN1.Y-SUP4t	Addgene
SC_ADH2_gRNA	p426-based *SNR52*, *URA3*, 2µ, AmpR with gRNA for the *ADH2* promoter	This study

SC_URA3_gRNA	p426-based *SNR52*, *URA3*, 2µ, AmpR with gRNA for the *URA3* scar	This study
SC_LEU2_gRNA	p426-based *SNR52*, *URA3*, 2µ, AmpR with gRNA for the *LEU2* scar	This study
p414-TEF1p-Cas9-CYC1t-URA3	p426-based *TEF1*p, Cas9, *CYC1*t, *URA3*, CEN/ARS origin	This study

Table 22 FapR and FapO primers and plasmids for *K. marxianus* integration

Primer Name	Sequence	Description
KM_XYL1p F1	cgaatcccgtcagtATATGGGTTTGATATGGCGGgttttagagctagaaat	
KM_XYL1p F2	cgaatcccgtcagtCCTGGGCCTGGAACTATGCCgttttagagctagaaat	gRNAs for site *XYL1* promoter via Gibson
KM_XYL1p R1	atttctagctctaaaacCCGCCATATCAAACCCATACtgacgggattcg	
KM_XYL1p R2	atttctagctctaaaacGGCATAGTTCCAGGCCCAGGactgacgggattcg	
KM_XYL1p_donor_F	TGGATCCATTGTTCAAGGTAACTGTTGGTGCGAGGTATGTCATTTGTAATTAAAACTTAG	donor for integration of FapO into *XYL1*p
KM_XYL1p_donor_R	TATGCCTGGGCGCAGCCCAGGCATAGGTTGGTTTAGGCCCAGGGTGTCGTTAATTACCCG	
KM_FapR_donor_IV-1_F	GCTCCATGTACATAATAAATATAGACTAATAAGATTCGCggagcgatttgcaggcatttg	FapR donor for site IV-1
KM_FapR_donor_IV-1_R	CGGAATATTGAACAACCTTGAATTCGCTGTTCCAATGCATTACGACCGAGATTCCCA	
KM_FapR_donor_IV-2_F	AAGAAACAGAGCCAGAAAAAAGGCAGCAACGAATCACAAggagcgatttgcaggcatttg	FapR donor for site IV-2
KM_FapR_donor_IV-2_R	CTATCTATAGCTCTATATCTCTACTATAACTCTATATCTCCATTACGACCGAGATTCCCA	
KM_FapR_donor_HIS3_F	CCCAGAAAGGAAAGCTTTTGTGTCTAGAATAACAAATGAggagcgatttgcaggcatttg	FapR donor for site *HIS3*
KM_FapR_donor_HIS3_R	CCTAATGCTTCTTTAAAAGCCTGTCCTAGAGCAATACCACCATTACGACCGAGATTCCCA	
KM_XYL1_Check_F	ACGATTCCTTCCTTGATGGC	site *XYL1*p integration check
KM_XYL1_Check_R	CCGTGTTTCCAGTGTGTGAT	
KM_XYL2p_F	cgaatcccgtcagtTCTTTACACCGTCTTGTTTGgttttagagctagaaat	
KM_XYL2p_R	atttctagctctaaaacCAAACAAGACGGTGTAAAGAactgacgggattcg	gRNAs for site *XYL2* promoter via Gibson
KM_XYL2p_Fv2	cgaatcccgtcagtTCTGAAGATGTGCTATACTGgttttagagctagaaat	
KM_XYL2p_Rv2	atttctagctctaaaacCAGTATAGCACATCTTCAGAactgacgggattcg	
KM_XYL2_donorF	AAGAACTTTTTCGCCGCATGGTAGGCTAGATTTCGGTTAGAGGGTGTCGTTAATTACCCG	donor for integration of FapO into *XYL2*p
KM_XYL2_donorR	GTTGATAATTTGTATTTTTGTTATTGGTAGCGCTCGCTCCTTTTTGTAATTAAAACTTAG	
KM_XYL2_checkF	CGGAAACCGAAGAACATATT	

KM_XYL2_checkR	TCTTTAATTTCTGGGACAGG	check for FapO integration into XYL2p

Plasmid Name	Description	Source
piW t-sg-r	TEF1p-Cas9-CYCt, TDH3p-tRNAgly-gRNA-HDV-CYCt, C/A, *URA3*, AmpR	Chapter 3
KM_XYL1_gRNA	piW t-sg-r targeting the *XYL1* promoter	This study
KM_HIS3_gRNA	piW t-sg-r targeting *HIS3*	This study
piW t-sg-r dCas9	TEF1p-dCas9-CYCt, TDH3p-tRNAgly-gRNA-HDV-CYCt, C/A, *URA3*, AmpR	This study
KM_IV-1_gRNA	piW t-sg-r targeting Chromosome IV site 1	This study
KM_IV-2_gRNA	piW t-sg-r targeting Chromosome IV site 2	This study
KM_XYL2_gRNA	piW t-sg-r targeting the *XYL2* promoter	This study

oPool gRNA library construction

Oligo pools (oPool) were designed with overhangs for Gibson into the gRNA plasmid p426. These oPools were made double-stranded by a 10-cycle Q5 pcr using a primer homologous to the 3' end. After PCR, the mixture was cleaned with a Zymo Clean and Concentrator kit and then assembled with the p426 backbone via Gibson assembly. This library was transformed into *E. coli* and cultivated in a 20 mL flask of LB-Amp. Aliquots of this cultivation were frozen, and plasmids were extracted from 5 mL.

Next generation sequencing

The original population of the yeast transformation library with the gRNAs as well as the final population after growth selection were assayed using next generation sequencing (NGS). After transformation and recovery, 1 mL of a frozen aliquot of the culture was removed from the -80°C freezer and spun down on a benchtop centrifuge for 5 min at 16000xg to pellet the cells. This pellet was resuspended in buffer and digested with Zymolase at 37°C for 1 hour, after which the standard extraction protocol was performed using the Zymo Yeast Plasmid Extraction Kit II.

Extracted plasmid DNA was then processed for sequencing. First, primers that add 1-4 bp to each end of the sequence of interest were prepared to add sequence diversity, along with 33 bp illumina-specific adapters (**Table 23**). These primers were verified beforehand to demonstrate no pairwise bias. The plasmid library was PCR amplified using KAPA HiFi 2x and a mix of staggering forward and reverse primers to separate the sequencing region from the plasmid backbone. The DNA was purified from the PCR mix using AMPure XP beads, and then PCR amplified again to add

illumina-specific Nextera XT barcodes to the ends of each sample. This library was then purified again with AMPure XP beads and the samples were pooled and denatured in preparation for the MiSeq.

Table 23 Staggering primers for the gRNA library that include the illumina adapters

Primer Name	Sequence
SC-NSG-Lib-Fwd-1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATcggcgttcgaaacttctccgcag
SC-NSG-Lib-Fwd-2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCATcggcgttcgaaacttctccgcag
SC-NSG-Lib-Fwd-3	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCATcggcgttcgaaacttctccgcag
SC-NGS-Lib-Rev-1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGtcaagttgataacggactagcc
SC-NGS-Lib-Rev-2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATtcaagttgataacggactagcc
SC-NGS-Lib-Rev-3	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCATtcaagttgataacggactagcc

Table 24 gRNA sequences used to construct an oPool library of gRNAs in *S. cerevisiae*

Each targeting sequence is added to the following primer: gcagtgaaagataaatgatc[20bp target]gttttagagctagaaatagc			
Gene Loci	**Targeting Sequence**	**Gene Loci**	**Targeting Sequence**
YSP1	TAAGGCAAAACTACTACATT	*POT1(YIL160C)*	ACATGAGGAGGGTAATGATG
PYC2	AAAAATCACAGAAACACAAT	*ARO1*	ACATTTGCGGTTAGAATCTA
GIS1	AGGAAAAACGATGAAATATG	*ARO9(YHR137W)*	CGGTTATCCGCGGCAAATCT
YEA6	CGTAGAAAATGTTGACAGAG	*LPD1(YFL018C)*	GCGGAGACTTCTCGCCAATG
NTE1	TATAAGGATCTTAAAAATCT	*ATG32(YIL146C)*	GAAATCTATCATCCCGTTGT
FAS1	AATAAATATACTAAAAAGAG	SER3	AAAAAGCGTTGCGTGATGTT
ZWF1	GCCAATTGGAGAGGGGGAAG	IDP1(YDL066W)	TCCACTTGCCCATATAAAGA
GSY1	TCCTTCTTTTCTTCCCGCAA	HXK2	ACAAAGAGACATCACGGAAT
PRB1(YEL060C)	AAGAGCGCGATGAATATAAA	CTP1(YBR291C)	CTTCAGTAAGACTAACTTAA
YIA6(YIL006W)	AAGGCTTACTTTATAACGGT	*GPD1(YDL022W)*	AGTCTACGTGCGAATTAGGT
POR2(YIL114C)	AAAAGGAAAGGAATTGCTAA	*GDH1(YOR375C)*	AAATCAGCACTGAAAAATTG
MPC1(YGL080W)	GTCGCAAAAGAGAACTTTAC	POX1	TGATTAAACTCCGAAGCGAA
MPC2(YHR162W)	AGGAAACAATGCAGCCCCCA	FAA1	ATTGTCTCACGACAAGTGAG
PDA1(YER178W)	TACCAGAGGGGATTATTGTC	FAA4	GAATGCAGAAGAGAAAGATG
YAT2(YER024W)	TCATCGTGATGATTCATTAT	FAT1	TAGTATTTATAAAAAATTTC
ATO2(YNR002C)	CTCCCTCCGCCGATTTAGGA	GPP1(YIL053W)	AGACGACAATATGAACAAAA
MEP3(YPR138C)	TTTTTAGTTAAGGTACCCGA	GPP2(YER062C)	TCACGTGGGAGGCCCGTTTA
HDA1	CTCGAGATTTCCAAACGGTG	ACS2	GTTTGATCAGGCTAAACGTG
YNG2(YHR090C)	TTTCTATTTCAACTAACGGT	PXA1	TATTCAGCTATATGTTGATC
PLB1	GAAAAATATGCGGGAATAAG	ERG10	TCCAGCGAAAAAACCGGCTT
ACC1	AATTAAGCTAGAAGACGAAT	ANT1	GAGTGACTATTTCAGATAAT
PEP4	GCGGGTGTCGATGGATTAAG	ACB1	TCCGCAAACTCAACAGCTCC
ADH1(YOL086C)	CACTACTCTCTAATGAGCAA	PCK1	TTGGCTGGGGATAGCAACAT
PDH1	TAAAGTTTGTAAACGGCCCG	ALD1	ACACCGTTCGAGGTCAAGCC
ACS1(YAL054C)	CATTTTGTATATAAACTGGG	ERG13	CTCCATTCGGCCTCATCGGG
CIT2(YCR005C)	TTCCCTGAAACTTAATAATG	HMG1	TTAAAAGATCTAATTTTCAA
MLS1(YNL117W)	AGACAAGAAAAGAAAAATAA	DGA1	ACTAAGTTACGGGCCGACAA
MCT1(YOR221C)	CGCATTTCAACCTCCGATGG		

Results and discussion

Design of a growth-based malonyl-CoA sensor for *S. cerevisiae*

The FapR/FapO malonyl-CoA sensor native to *B. subtilis* has been used to optimize malonyl-CoA levels in *S. cerevisiae* in the past (Bergman et al. 2016; Meadows et al. 2016); however all previous methods relied on fluorescent based single-cell or single-mutant screening methods, which significantly limits the rate and scope of identifying favorable genetic modifications. These methods place the FapO binding region upstream of a fluorescent marker, either integrated into the genome or on a plasmid. Clones with high levels of malonyl-CoA produce higher levels of this marker and can therefore be sorted within a population of clones using specialized cell sorting equipment or through mass cultivation in 96-well plates. A disadvantage of this method is that the assay only measures total fluorescent protein production and is sensitive to when the samples are assayed. For example, a clone with high malonyl-CoA in the early exponential phase might produce a similar amount of fluorescent protein over the growth period than a clone with consistent malonyl-CoA levels throughout all growth phases. Furthermore, fluorescent protein production is generally resource-intensive and expression is dependent on more than just transcription, potentially mudding the results further.

An alternative method is to use this FapO binding region upstream of a growth-associated gene, enabling growth rate to be correlated to malonyl-CoA availability throughout the entire growth phase. Growth-correlated screening can enable pooling of samples and elimination of poor mutants using a process of selective enrichment. Mutants with high malonyl-CoA levels will be able to bind more FapR molecules and therefore prevent repression of key metabolism genes,

leading to better substrate utilization and a faster growth rate. Over multiple generations, the mutants with a higher growth rate will slowly outcompete their slower-growing cohort and eventually dominate the population within the mixed culture. One disadvantage of this method would be the rapid elimination of mutants that may have slower growth rates but high malonyl-CoA as a result, but generally these slow-growing phenotypes are undesirable.

S. cerevisiae is typically grown on glucose, however regulation of glucose metabolism is not feasible using FapO due to the multitude of glucose metabolism enzymes. Therefore, we chose ethanol as an alternative substrate. Ethanol can be metabolized readily by *S. cerevisiae*: however, this metabolism and subsequent conversion to acetyl-CoA and biomass is dependent on one initial enzyme, alcohol dehydrogenase, which is transcribed from the gene *ADH2.* Strains with modifications that improve malonyl-CoA will see FapR binding to malonyl-CoA instead of the FapO region in the promoter and therefore have more Adh2 enzyme available to metabolize ethanol, resulting in a correlation between malonyl-CoA levels and growth rate in ethanol. Previous studies have characterized the sensitivity of the sensor using *GFP* and found a strong correlation between protein expression and malonyl-CoA by using a native *TEF1* promoter that is interspaced with three FapO binding regions. This “TEF123” promoter was chosen for integration to replace the *ADH2* promoter (**Figure 25**).

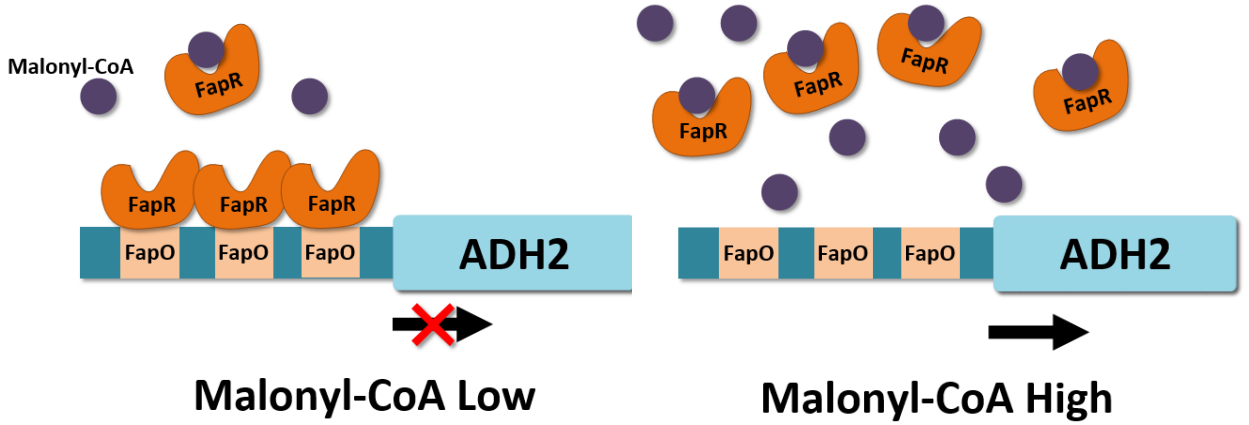


Figure 25 Designed FapR/FapO regulation of *ADH2* in *Saccharomyces cerevisiae* utilizing a TEF123 strategy

Malonyl-CoA sensor validation

After integration and sequencing of the FapO and FapR insertions in *S. cerevisiae*, we tested our hypothesis of malonyl-CoA associated growth rate. We cultivated BY-FapO-FapR relative to BY-FapO and the strain BY-FapO-FapRΔgis1 (the latter constructed in our lab by Anh Pham). The *GIS1* gene is a transcription factor that is responsible for gene regulation during nutrient limitation (Pedruzzi et al. 2000) and was used as a control since it has 2.4-fold higher 6MSA titers over the wild-type with no effect on growth rate (Anh Pham, unpublished results). Since there is higher 6MSA, it is likely due to either higher malonyl-CoA pools or a higher flux towards malonyl-CoA

We expect that the control strain BY-FapO will have the maximum expected growth rate on ethanol, since there is no FapR repression of the *ADH2* gene, and that strain BY-FapO-FapR will have the minimum growth rate, since FapR will be able to bind to the FapO region, repressing *ADH2* expression. Strain BY4741ΔGIS1-FapO-FapR, which has a higher production of 6MSA and likely a higher flux of malonyl-CoA, should have a growth rate between that of the BY-FapO and the BY-FapO-FapR.

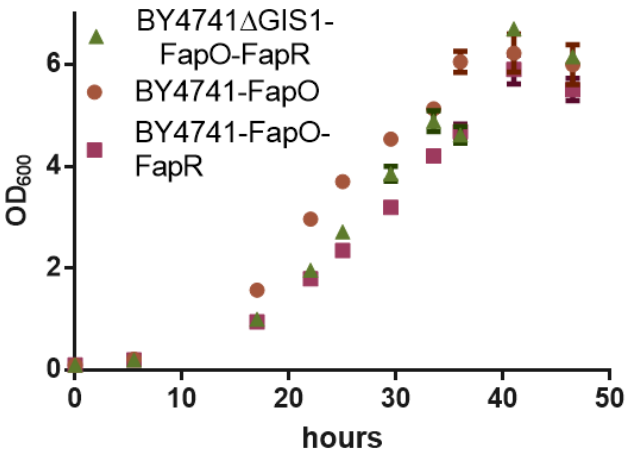


Figure 26 Growth curve of BY-FapO strains in ethanol medium at 30C in a water bath shaker, over 48 hours. (n=3)

The three strains were cultivated in ethanol medium and OD_{600} was recorded and plotted over the course of two days **(Figure 26**). From the exponential growth phase of these cultures, we calculated the maximum specific growth rate and doubling time (**Table 25**). As expected, the fastest growing strain was the BY-FapO strain without any FapR repressor elements. The Δgis strain had a 5% faster growth rate than the BY-FapO-FapR strain without the knockout. This is consistent with our expectation that this strain produces more malonyl-CoA and therefore, is able to bind more FapR molecules, preventing repression of the *ADH2* gene. Furthermore, this Δgis1 strain must not produce sufficient malonyl-CoA to bind all the FapR molecules, as it grew 15% slower relative to the FapO only strain, which has no *ADH2* repression.

Table 25 Maximum specific growth rate and doubling time of BY-strains with and without FapR compared to BY with a *gis1* knockout

	BY4741-FapO	**BY4741-FapO-FapR**	**BY4741ΔGIS1-FapO-FapR**
μ_{MAX} (1/h)	0.150 ± 0.006	0.128 ± 0.002	0.135 ± 0.002
Doubling time (h)	4.64	5.4	5.13

These results are encouraging, and suggest large-scale selection of strains with improved malonyl-CoA levels may be possible. It should be noted that the growth rate range observed is within a very narrow window with potentially small resolution. Strains with significantly more

malonyl-CoA might only improve growth rate by a few percent. Further tuning this sensor dynamic range by increasing FapO binding sites in the promoter, by integrating more copies of the FapR gene, or by running growth enrichment for a larger number of generations may further improve our ability to detect malonyl-CoA associated growth rate changes.

Design of a *S. cerevisiae* gRNA library

With our new growth-based malonyl-CoA sensor, the next step was to design a method to create a large mutant library that we can screen for improved malonyl-CoA levels. A dCas9-based system which binds to the gDNA but does not cut prevents the need for recombination donor DNA in *S. cerevisiae* and also enables long-term maintenance of a gRNA plasmid that can be used as an identifier later in the selection process. The dCas9-MxiI gene was integrated into the genome to enable consistent expression of this enzyme throughout our screens.

We approached designing a CRISPRi gRNA library via two different methods. The first method was to develop a set of synthesized gRNAs from oligos. Based on the sequence information, gRNAs targeting genes of interest within the *S. cereivisiae* genome were designed using a yeast CIRSPRi web tool (Smith et al. 2016) to target 56 sites with known effects on acetyl- or malonyl-CoA levels from studies in our lab (unpublished results and from (Fernandez-Moya and Da Silva 2017; Cardenas and Da Silva 2016, 2014)). These synthetic oligos are double-stranded and then were cloned into a gRNA backbone using Gibson assembly. This small library was designed as a validation experiment for the FapR-FapO system. Using next generation sequencing (NGS), comparing the distribution within the gRNA population at the end of the

growth enrichment to the gRNA population of the original transformation will provide insight into how the FapR-FapO system is influencing growth when in the presence of various gRNAs.

We also sought to design a library using a more randomized, genome-scale method that will allow us to select new genomic modifications that increase malonyl-CoA levels. Modifying a method developed by Lane et al. (A. B. Lane et al. 2015), it is possible to generate a library of donor targets for dCas9 that has good coverage of the whole *S. cerevisiae* genome including regulatory elements, structural regions and regions of unknown function, not just functional genes and associated promoters. This library is generated by digesting and then cloning gRNAs that are sourced from the original *S. cerevisiae* gDNA cut at PAM sites, which are distributed randomly throughout the entire genome.

The method we used to generate this library is outlined in **Figure 27** and a more in-depth methodology can be found in **Appendix E**. An overview of this whole process is in **Figure 29**. Searching for PAM-site digesting restriction enzymes within chromosomal sequences of the *S. cerevisiae* reference genome ("SGD S288C Reference Genome," n.d.), we found the cut frequency of the three enzymes HpaII (NGG), ScrFI (NGG) and BfaI (NAG) to be ~7.6% of the yeast genome, yielding approximately 920,000 target gRNAs (one cut site every 13 bases). Digesting with only

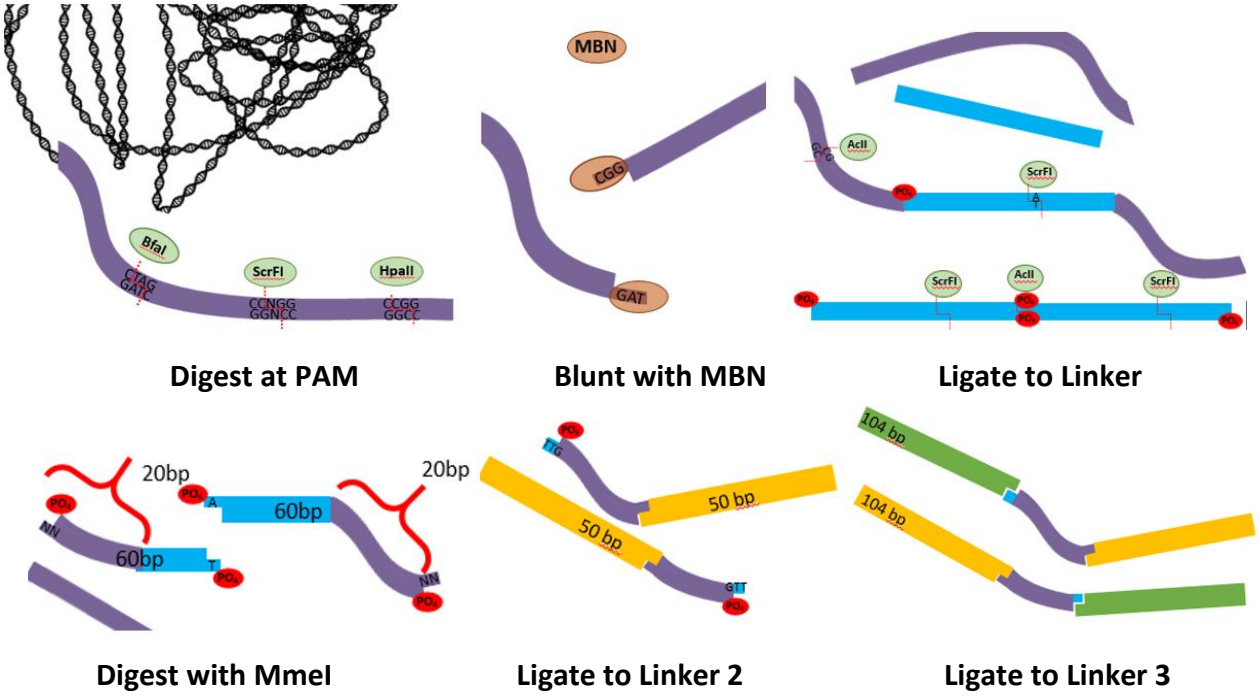


Figure 27 Randomized gRNA library construction method from yeast gDNA for generation of fragments for Gibson cloning into dgRNA plasmids. gDNA (purple), linker 1 (blue), linker 2 (yellow) and linker 3 (green).

NGG-yielding enzymes HpaII and ScrFI cuts about 0.2% of the genome, with the average fragment length being 460 bases. This yields a maximum gRNA plasmid pool of 52,600. We digested ~10 µg of the 12.1 Mb BY4741 genome with the PAM-site enzymes HpaII and ScrFI.

These gDNA fragments were blunted and ligated to Linker 1, which enables digestion 20 bp upstream via the type-II MmeI enzyme. The 20bp gDNA fragment that is generated is then ligated to Linker 2 and Linker 3 which contain overhangs that are homologous to the dgRNA plasmid. These fragments are PCR-amplified to generate a large quantity of DNA for Gibson assembly.

Gibson-assembled dgRNA plasmids (**Figure 28**) were transformed into *E. coli* via high efficiency electroporation ($4*10^{10}$ cfu/µg) and grown in 50 mL LB-Amp media. To ensure a

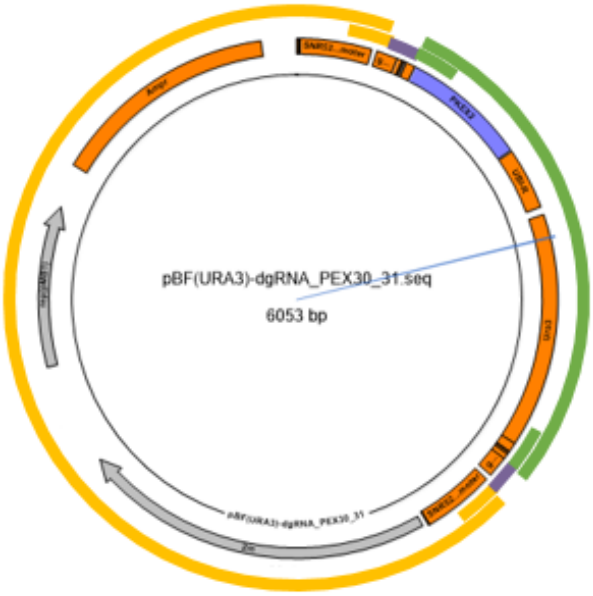


Figure 28 Gibson assembly of dgRNA plasmids, with 20 bp randomized gRNA target (purple)

sufficient population of the library is carried over between each reinoculation step, approximately 6mL of the fully grown culture (or $1.44*10^{10}$ cells) was saved as a freezer stock. With an expected plasmid library size of 52,600 targets, 10^{10} will be enough to have approximately ~200,000-fold coverage, assuming a very conservative one plasmid per cell, and should be sufficient to ensure the full library is maintained. The remaining plasmid library was extracted and saved for transformation into yeast. Coupling a library of this kind with the malonyl-CoA sensor represents a novel method for identifying unknown genes and key bottlenecks that are relevant for polyketide production.

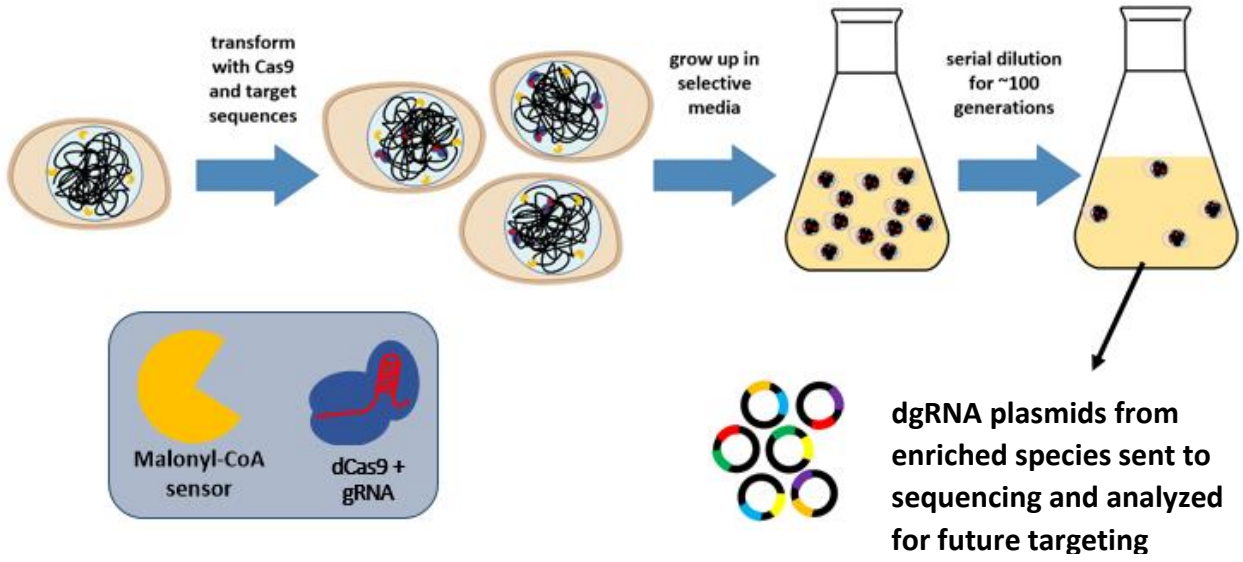


Figure 29 Method overview for combinatorial CRISPR/dCas9 intervention and selective cultivation of the strain library for increased malonyl-CoA production.

S. cerevisiae growth experiments & NGS sequencing results

The *S. cerevisiae* library was generated by transforming the expression strain, BYt-FapO-FapR-dCas9, with the library that was synthetically generated from pooled oligos. We chose this library initially to evaluate the efficacy of this expression strain and our sequencing methodologies with a smaller library. The yeast library was grown and then serially diluted into fresh ethanol medium over the course of three weeks. The long time course of this cultivation was due to the slow growth rate. Over the course of the experiment, the gRNAs that improve malonyl-CoA levels (and therefore the growth rate of the culture in ethanol) should be enriched in this population. At each reinoculation step, aliquots of the previous culture were maintained in glycerol stocks. At the end of the experiment, each culture had been cultivated between 50 and 70 generations.

To evaluate if the malonyl-CoA sensor influenced the population of gRNAs, we sequenced three pools of plasmids. One pool was from the original transformation of BY-FapO-FapR-dCas9 prior to enrichment. This plasmid pool serves as the control library to which the enriched libraries will be compared. The next two pools were from the end of the enrichment experiment, but were cultivated and reinoculated at two different initial optical densities: OD_{600}=0.2 or OD_{600}=0.4. We chose two different starting optical densities due to the extremely slow initial growth rate of the cultures on ethanol (doubling time of 3 days at OD_{600}=0.1) to ensure that doubling time of at least one culture exceeded the rate of cell death. Note that the doubling time of these cultures is significantly slower than that of the sensor validation experiment, in part due to poor recovery after transformation in ethanol medium due to the *ADH2* repression.

We designed primers such that we stagger the alignment of each gRNA region by 1-4 bases, which should prevent issues in complexity during sequencing. These primers also contain the 33-34 bp adapter that will be used to add on the illumina barcodes. Prior to sequencing, we verified these staggering primers did not have significant pairwise bias via PCR. These plasmid libraries were then sequenced with paired-end reads on an illlumina MiniSeq at the UCI Genomics High-Throughput Facility.

Each of the three samples resulted in between 60,000 and 80,000 pairs of reads that were high quality. The paired ends of the fastq files were joined using KBase (Arkin et al. 2018) and then, using the count_spacers.py script (Joung et al. 2017), the frequency of each gRNA in our library was calculated. The first test was to verify that our staggering primers were indeed unbiased. Based on the counts for each of the primers, there is a small bias towards SC-NGS-Lib-Fwd-1 (**Figure 30**) but it is not significant enough to have any influence on sequencing quality. However, for searching for the staggering tag,

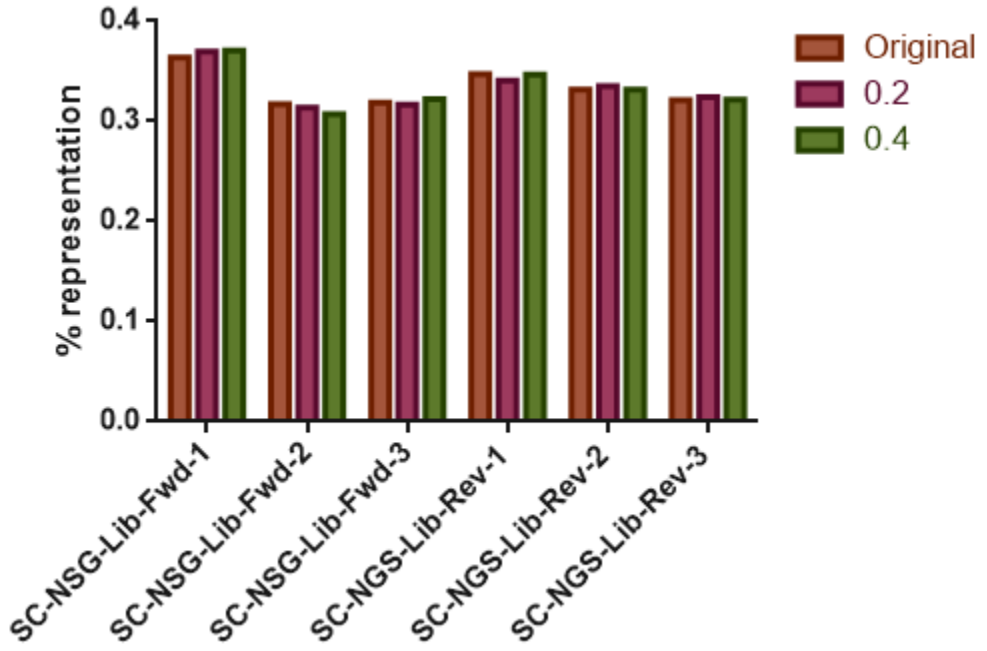


Figure 30 Percent representation of each of the three staggering primers (forward and reverse), over three samples (original, OD=0.2 and OD=0.4) and ~70k paired end reads.

only about ~30,000 of each sample had any of the six tags. This indicates that there likely was an error or contamination in the sequencing or initial PCR steps that contributed to many of our sequences not being properly tagged.

Next, we evaluated all the calls with correct staggering primers and gRNA structural elements and counted the frequency of each gRNA. Of the total ~70k reads and ~35k reads with the correct staggering primer, only ~11k were exact matches for any of the gRNAs. Of these matches, sixteen of the fifty six gRNAs were not found in any of the samples (**Table 26**). Despite the issues with sequencing, the low coverage of our gRNAs in even the original library implies there was a bias from the construction or initial transformation of the gRNA library.

Table 26 gRNAs that were not represented in sequencing in any of the three gRNA pools evaluated

Missing gRNAs	
YSP1	*POT1(YIL160C)*
PYC2	HXK2
YEA6	FAT1
PRB1(YEL060C)	GPP1(YIL053W)
POR2(YIL114C)	GPP2(YER062C)
MPC1(YGL080W)	ANT1
YNG2(YHR090C)	HMG1
ADH1(YOL086C)	DGA1

From the gRNAs that were represented, we plotted the percent representation of each gRNA in each of the three pools (**Figure 31**). From these results, it is clear that there is a strong bias in the original sample towards the *MPC2* gRNA. *MPC2* is a mitochondrial pyruvate carrier and expressed during growth on fermentable carbon sources (Herzig et al. 2012). Since the original library was cultivated and transformed in rich medium, it is unclear why downregulation of this gene would be favorable in this original library. It is most likely a bias that occurred in the original cultivation in *E. coli* prior to transformation.

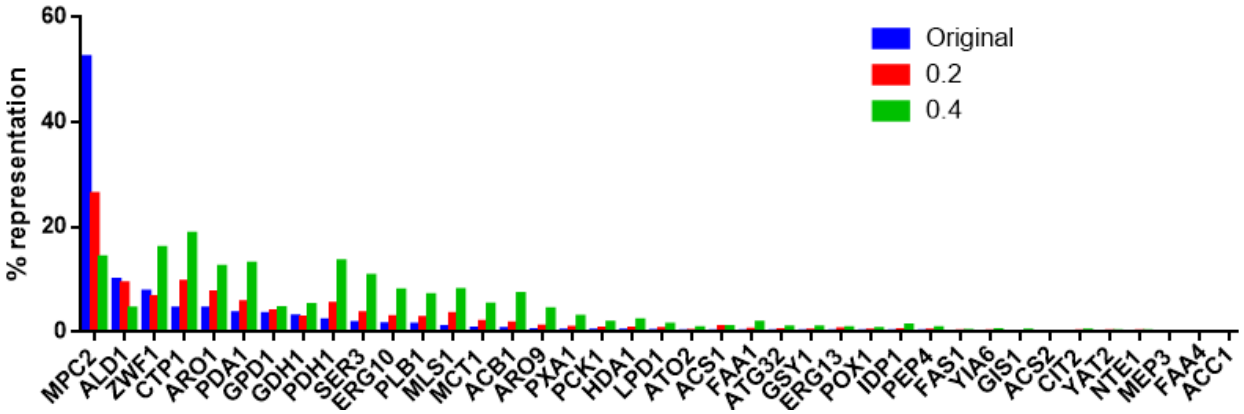


Figure 31 Percent representation of each gRNA in each of the three plasmid pools, out of a total of ~11k exact matches.

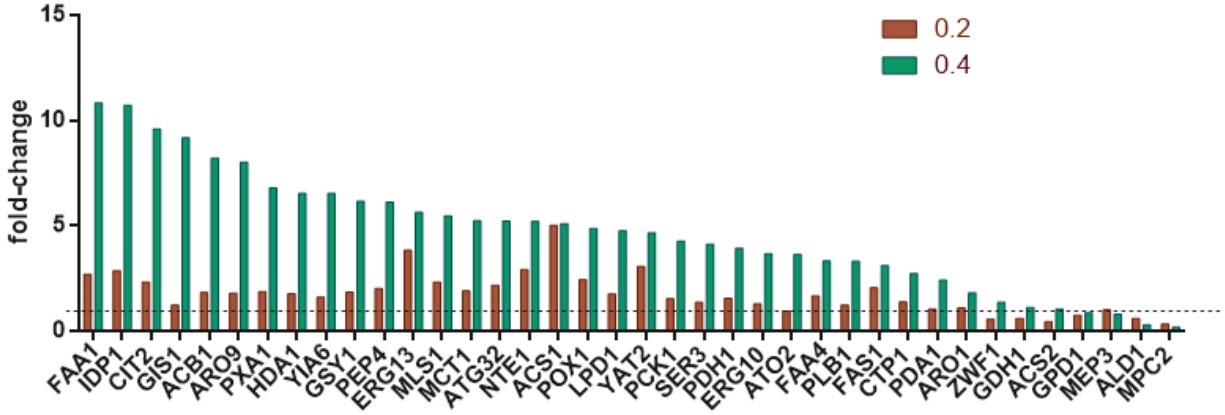


Figure 32 Fold-change in gRNA counts relative to the original gRNA pool

When we replot this data to evaluate the fold-change between the original library and the enriched libraries, there are some potentially interesting results. The guide for *FAA1*, a gene for acyl-CoA synthase, had a high increase in fold-change relative to the original culture. Since acyl-CoA synthase is involved in fatty acid synthesis, and therefore utilization of acetyl- and malonyl-CoA, downregulation of this gene (especially when cells are growing slowly), may result in higher levels of malonyl-CoA. Similarly, *IDP1* and *CIT2*, the next two highest represented targets, are related to citrate metabolism in the TCA cycle. Downregulation of these genes

might contribute to an increase in pyruvate, which is consumed in the TCA cycle, and in a key precursor to acetyl- and malonyl-CoA that has been demonstrated in the past to improve TAL accumulation (Cardenas and Da Silva 2016).

From the first trial of the malonyl-CoA sensor in *S. cerevisiae*, we have some interesting results; however expectations need to be tempered. To fully verify the result of this screen and demonstrate transferability from the FapO-FapR-dCas9 strain to a wild-type host, promising gene targets such as *FAA1* will need to be individually removed in the wild-type strain and tested for malonyl-CoA levels and polyketide productivity. Furthermore, challenges with NGS regarding missing and unexpected DNA tags will need to be addressed with this small library. Following this, a discovery platform using the full, genome-scale gRNA library can be initiated.

Adaptions of a growth-based malonyl-CoA sensor to *K. marxianus*

In Chapter 2, *K. marxianus* was demonstrated to be a promising host for polyketide production and in particular on nonconventional carbon sources. However, *K. marxianus* metabolism is significantly different than that of *S. cerevisiae* and is not well understood. Therefore, a malonyl-CoA sensor would be particularly useful in this yeast species. Utilizing a sensor to screen and select a library of gRNAs in a similar manner to the strategy in *S. cerevisiae* would not only improve the utility of *K. marxianus* as a polyketide platform, but also enable a better understanding of the genes associated with an exceptional xylose metabolism that is not fully understood. We thus designed a similar growth-based malonyl-CoA sensor for this yeast.

In a similar manner to the design in *S. cerevisiae*, the FapR, FapO and dCas9 genes are integrated into the genome. In *K. marxianus,* we have observed efficient growth on xylose, which fortunately is also reliant on just one initial gene – *XYL1*, which encodes for xylose reductase. Therefore, we chose to regulate *XYL1* (instead of *ADH2*) by replacing the native *XYL1* promoter with the artificial FapO-TEF123 promoter, which should result in a correlation between malonyl-CoA availability and growth rate in xylose (**Figure 33**). This has an advantage over the design in *S. cerevisiae*, as the initial growth rate of *K. marxianus* on xylose far exceeds that of *S. cerevisiae* on ethanol, allowing a more rapid enrichment phase as well as less opportunity for contamination or long-term adaption via mutations to occur.

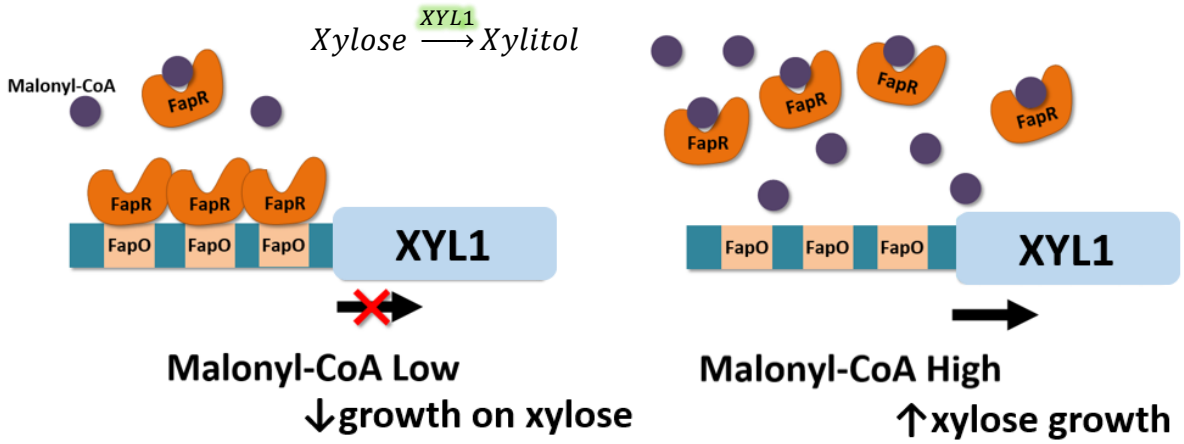


Figure 33 Scheme for implementing a malonyl-CoA sensor in *K. marxianus* that is tied to xylose metabolism

Another change is the method of gRNA generation. Instead of generating this library in-house, we collaborated with the Wheeldon Lab (UC Riverside) and the Joint Genome Institute (JGI) to generate a genome-scale gRNA library. This library was designed to target 4 sites within each known gene and promoter region. The oligos for each gRNA were then synthesized on a chip and Gibson assembled into a gRNA backbone. We now have the library in the lab, and have

initiated experiments to optimize transformation efficiency in the FapO-FapR-dCas9 strains. Future work will focus on the development of this *K. marxianus* based library for the identification of novel metabolic engineering targets, and extension of the sensor into already engineered strains. The initial strains tested in this study will be the *K. marxianus* strains CBS712Δ*ura3*Δ*ku70* and KM1Δ*ura3*Δ*ku70* expressing the carbon-saving pathway (Chapter 3).

Conclusions

Development and subsequent selection of refined strains for improved production yields is often challenging, requiring many combinations of gene interventions, some of which may be unexpected. We developed a malonyl-CoA sensor that helps to address the challenges of this screening process. *S. cerevisiae* and *K. marxianus* strains were designed to incorporate growth-coupled malonyl-CoA sensing through the integration of FapR and FapO genes that regulate the metabolism of ethanol and xylose, respectively. To generate large libraries of engineered strains, CRISPR-dCas9 libraries were also generated for each yeast species. Together, these tools enable a pooled screening method that can be used to identify novel gene interventions. Evaluation of this method using a small pilot library in *S. cerevisiae* yielded promising results that can be further explored using conventional knockout methods. The novel sensor-selection method developed in this work can be adapted for use in other strains or from other substrates, enabling a faster method for screening and identifying genes to improve the production of malonyl-CoA based products.

Acknowledgements

This research was supported by the National Science Foundation: Grant No. CBET-1803677, CBET-1605357 and EEC-0813570 (through the Engineering Research Center CBiRC (Center for Biorenewable Chemicals)) and the Department of Energy/Office of Science: Grant No. DE-SC0019093. The authors thank Anh Pham for the *ΔGIS1* knockout strain, Jin Wook Choi for all the FapR and FapO gBlock design and experimental inspiration, Ruben Fernandez Moya for helping in strain construction and Pam B. Besada Lombana for aiding in next generation sequencing preparation and analysis.

References

Arkin, Adam P., Robert W. Cottingham, Christopher S. Henry, Nomi L. Harris, Rick L. Stevens, Sergei Maslov, Paramvir Dehal, et al. 2018. “KBase: The United States Department of Energy Systems Biology Knowledgebase.” Nature Biotechnology. https://doi.org/10.1038/nbt.4163.

ATCC. n.d. “ATCC Medium: 1043.” https://www.atcc.org/~/media/D92466B5E4084EA9AF62938EE401035E.ashx.

Banat, I. M., P. Nigam, and R. Marchant. 1992. “Isolation of Thermotolerant, Fermentative Yeasts Growing at 52-Degree C and Producing Ethanol at 45- Degree C and 50-Degree C.” World Journal of Microbiology & Biotechnology 8 (3): 259–63. https://doi.org/10.1007/BF01201874.

Benatuil, Lorenzo, Jennifer M. Perez, Jonathan Belk, and Chung Ming Hsieh. 2010. “An Improved Yeast Transformation Method for the Generation of Very Large Human Antibody Libraries.” Protein Engineering, Design and Selection. https://doi.org/10.1093/protein/gzq002.

Bergman, Alexandra, Verena Siewers, Jens Nielsen, and Yun Chen. 2016. “Functional Expression and Evaluation of Heterologous Phosphoketolases in Saccharomyces Cerevisiae.” AMB Express 6 (1). https://doi.org/10.1186/s13568-016-0290-0.

Bogorad, Igor W., Tzu Shyang Lin, and James C. Liao. 2013. “Synthetic Non-Oxidative Glycolysis Enables Complete Carbon Conservation.” Nature 502 (7473): 693–97. https://doi.org/10.1038/nature12575.

Branda, Steven S., José Eduardo González-Pastor, Sigal Ben-Yehuda, Richard Losick, and Roberto Kolter. 2001. “Fruiting Body Formation by Bacillus Subtilis.” Proceedings of the National Academy of Sciences of the United States of America. https://doi.org/10.1073/pnas.191384198.

Burgard, Anthony P., Priti Pharkya, and Costas D. Maranas. 2003. “OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization.” Biotechnology and Bioengineering 84 (6): 647–57. https://doi.org/10.1002/bit.10803.

Cardenas, Javier, and Nancy A. Da Silva. 2014. “Metabolic Engineering of Saccharomyces Cerevisiae for the Production of Triacetic Acid Lactone.” Metabolic Engineering 25: 194–203. https://doi.org/10.1016/j.ymben.2014.07.008.

Cardenas, Javier, and Nancy A. Da Silva. 2016. “Engineering Cofactor and Transport Mechanisms in Saccharomyces Cerevisiae for Enhanced Acetyl-CoA and Polyketide Biosynthesis.” Metabolic Engineering 36: 80–89. https://doi.org/10.1016/j.ymben.2016.02.009.

Chan, Yolande A., Angela M. Podevels, Brian M. Kevany, and Michael G. Thomas. 2009. “Biosynthesis of Polyketide Synthase Extender Units.” Nat. Prod. Rep. 26 (1): 90–114. https://doi.org/10.1039/B801658P.

Chen, Xiaoxu, Xiaoyu Yang, Yu Shen, Jin Hou, and Xiaoming Bao. 2017. “Increasing Malonyl-CoA Derived Product through Controlling the Transcription Regulators of Phospholipid Synthesis in Saccharomyces Cerevisiae.” ACS Synthetic Biology 6 (5): 905–12. https://doi.org/10.1021/acssynbio.6b00346.

Chen, Xiaoxu, Xiaoyu Yang, Yu Shen, Jin Hou, and Xiaoming Bao. 2018. “Screening Phosphorylation Site Mutations in Yeast Acetyl-CoA Carboxylase Using Malonyl-CoA Sensor to Improve Malonyl-CoA-Derived Product.” Frontiers in Microbiology 9 (JAN). https://doi.org/10.3389/fmicb.2018.00047.

Chia, Mei, Thomas J. Schwartz, Brent H. Shanks, and James A. Dumesic. 2012. “Triacetic Acid Lactone as a Potential Biorenewable Platform Chemical.” Green Chemistry 14 (7): 1850. https://doi.org/10.1039/c2gc35343a.

Choi, Jin Wook, and Nancy A. Da Silva. 2014. “Improving Polyketide and Fatty Acid Synthesis by Engineering of the Yeast Acetyl-CoA Carboxylase.” Journal of Biotechnology 187: 56–59. https://doi.org/10.1016/j.jbiotec.2014.07.430.

Chowdhury, Anupam, and Costas D. Maranas. 2015. “Designing Overall Stoichiometric Conversions and Intervening Metabolic Reactions.” Scientific Reports 5. https://doi.org/10.1038/srep16009.

David, Florian, Jens Nielsen, and Verena Siewers. 2016. “Flux Control at the Malonyl-CoA Node through Hierarchical Dynamic Pathway Regulation in Saccharomyces Cerevisiae.” ACS Synthetic Biology 5 (3): 224–33. https://doi.org/10.1021/acssynbio.5b00161.

Deaner, Matthew, Julio Mejia, and Hal S. Alper. 2017. “Enabling Graded and Large-Scale Multiplex of Desired Genes Using a Dual-Mode DCas9 Activator in Saccharomyces Cerevisiae.” ACS Synthetic Biology 6: 1931–43. https://doi.org/10.1021/acssynbio.7b00163.

Diomandé, Sara E, Christophe Nguyen-The, Marie-Hélène Guinebretière, Véronique Broussolle, and Julien Brillard. 2015. “Role of Fatty Acids in Bacillus Environmental Adaptation.” Frontiers in Microbiology 6: 813. https://doi.org/10.3389/fmicb.2015.00813.

Dudášová, Zuzana, Andrej Dudáš, and Miroslav Chovanec. 2004. “Non-Homologous End-Joining Factors of Saccharomyces Cerevisiae.” FEMS Microbiology Reviews. https://doi.org/10.1016/j.femsre.2004.06.001.

Dymond, Jessica S. 2013. “Preparation of Genomic DNA from Saccharomyces Cerevisiae.” In Methods in Enzymology. https://doi.org/10.1016/B978-0-12-418687-3.00012-4.

Evans, Christopher Thomas, and Colin Ratledge. 1984. “Induction of Xylulose-5-Phosphate Phosphoketolase in a Variety of Yeasts Grown on d-Xylose: The Key to Efficient Xylose Metabolism.” Archives of Microbiology 139 (1): 48–52. https://doi.org/10.1007/BF00692711.

Fang, Fang, Kirsty Salmon, Michael W Y Shen, Kimberly a Aeling, Elaine Ito, Becky Irwin, Uyen Phuong C Tran, G Wesley Hatfield, Nancy A. Da Silva, and Suzanne Sandmeyer. 2011. “A Vector Set for Systematic Metabolic Engineering in Saccharomyces Cerevisiae.” Yeast 28 (2): 123–36. https://doi.org/10.1002/yea.1824.

Fernandez-Moya, Ruben, and Nancy A. Da Silva. 2017. “Engineering Saccharomyces Cerevisiae for High-Level Synthesis of Fatty Acids and Derived Products.” FEMS Yeast Research. https://doi.org/10.1093/femsyr/fox071.

Flikweert, Marcel T., Martin De Swaaf, Johannes P. Van Dijken, and Jack T. Pronk. 1999. “Growth Requirements of Pyruvate-Decarboxylase-Negative Saccharomyces Cerevisiae.” FEMS Microbiology Letters 174 (1): 73–79. https://doi.org/10.1016/S0378-1097(99)00124-X.

Flikweert, Marcel T., Linda Van Der Zanden, Wouter M.Th M. Janssen, H. Yde Steensma, Johannes P. Van Dijken, and Jack T. Pronk. 1996. “Pyruvate Decarboxylase: An Indispensable Enzyme for Growth of Saccharomyces Cerevisiae on Glucose.” Yeast 12 (3): 247–57. https://doi.org/10.1002/(SICI)1097-0061(19960315)12:3<247::AID-YEA911>3.0.CO;2-I.

Gao, Jiaoqi, Wenjie Yuan, Yimin Li, Ruijuan Xiang, Shengbo Hou, Shijun Zhong, and Fengwu Bai. 2015. “Transcriptional Analysis of Kluyveromyces Marxianus for Ethanol Production from Inulin Using Consolidated Bioprocessing Technology.” Biotechnology for Biofuels. https://doi.org/10.1186/s13068-015-0295-y.

Gietz, Daniel, Andrew St Jean, Robin A. Woods, and Robert H. Schiestl. 1992. “Improved Method for High Efficiency Transformation of Intact Yeast Cells.” Nucleic Acids Research 20 (6): 1425.

Gilbert, Luke a, Matthew H Larson, Leonardo Morsut, Zairan Liu, A Gloria, Sandra E Torres, Noam Stern-ginossar, et al. 2013. “CRISPR-Mediated Modular RNA-Guided Regualtion of Transcription in Eukaryotes.” Cell 154 (2): 442–51. https://doi.org/10.1016/j.cell.2013.06.044.CRISPR-Mediated.

Groeneveld, Philip, Adriaan H. Stouthamer, and Hans V. Westerhoff. 2009. “Super Life - How and Why ‘cell Selection’ Leads to the Fastest-Growing Eukaryote.” FEBS Journal 276 (1): 254–70. https://doi.org/10.1111/j.1742-4658.2008.06778.x.

HAYASHI, Osamu, and Kiyoshi SATOH. 2006. “Determination of Acetyl-CoA and Malonyl-CoA in Germinating Rice Seeds Using the LC-MS/MS Technique.” Bioscience, Biotechnology, and Biochemistry 70 (11): 2676–81. https://doi.org/10.1271/bbb.60269.

Heirendt, Laurent, Sylvain Arreckx, Thomas Pfau, Sebastián N. Mendoza, Anne Richelle, Almut Heinken, Hulda S. Haraldsdóttir, et al. 2019. “Creation and Analysis of Biochemical Constraint-Based Models Using the COBRA Toolbox v.3.0.” Nature Protocols. https://doi.org/10.1038/s41596-018-0098-2.

Herzig, Sébastien, Etienne Raemy, Sylvie Montessuit, Jean Luc Veuthey, Nicola Zamboni, Benedikt Westermann, Edmund R.S. Kunji, and Jean Claude Martinou. 2012. “Identification and Functional Expression of the Mitochondrial Pyruvate Carrier.” Science. https://doi.org/10.1126/science.1218530.

Hua, Yan, Jichao Wang, Yelin Zhu, Biao Zhang, Xin Kong, Wenjie Li, Dongmei Wang, and Jiong Hong. 2019. “Release of Glucose Repression on Xylose Utilization in Kluyveromyces

Marxianus to Enhance Glucose-Xylose Co-Utilization and Xylitol Production from Corncob Hydrolysate." Microbial Cell Factories. https://doi.org/10.1186/s12934-019-1068-2.

ISHIDA, Nobuhiro, Satoshi SAITOH, Toru ONISHI, Kenro TOKUHIRO, Eiji NAGAMORI, Katsuhiko KITAMOTO, and Haruo TAKAHASHI. 2006. "The Effect of Pyruvate Decarboxylase Gene Knockout in Saccharomyces Cerevisiae on L -Lactic Acid Production." Bioscience, Biotechnology, and Biochemistry 70 (5): 1148–53. https://doi.org/10.1271/bbb.70.1148.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity." Science 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Joung, Julia, Silvana Konermann, Jonathan S. Gootenberg, Omar O. Abudayyeh, Randall J. Platt, Mark D. Brigham, Neville E. Sanjana, and Feng Zhang. 2017. "Genome-Scale CRISPR-Cas9 Knockout and Transcriptional Activation Screening." Nature Protocols. https://doi.org/10.1038/nprot.2017.016.

Juergens, Hannes, Javier A Varela, Arthur R Gorter de Vries, Thomas Perli, Veronica J M Gast, Nikola Y Gyurchev, Arun S Rajkumar, et al. 2018. "Genome Editing in Kluyveromyces and Ogataea Yeasts Using a Broad-Host-Range Cas9/GRNA Co-Expression Plasmid." FEMS Yeast Research, no. February: 1–16. https://doi.org/10.1093/femsyr/foy012.

Kim, Soo Jung, Seung Oh Seo, Yong Su Jin, and Jin Ho Seo. 2013. “Production of 2,3-Butanediol by Engineered Saccharomyces Cerevisiae.” Bioresource Technology 146: 274–81. https://doi.org/10.1016/j.biortech.2013.07.081.

King, Zachary A., and Adam M. Feist. 2014. “Optimal Cofactor Swapping Can Increase the Theoretical Yield for Chemical Production in Escherichia Coli and Saccharomyces Cerevisiae.” Metabolic Engineering 24: 117–28. https://doi.org/10.1016/j.ymben.2014.05.009.

Labuhn, Maurice, Felix F. Adams, Michelle Ng, Sabine Knoess, Axel Schambach, Emmanuelle M. Charpentier, Adrian Schwarzer, Juan L. Mateo, Jan Henning Klusmann, and Dirk Heckl. 2018. “Refined SgRNA Efficacy Prediction Improves Largeand Small-Scale CRISPR-Cas9 Applications.” Nucleic Acids Research. https://doi.org/10.1093/nar/gkx1268.

Lane, Andrew B., Magdalena Strzelecka, Andreas Ettinger, Andrew W. Grenfell, Torsten Wittmann, and Rebecca Heald. 2015. “Enzymatically Generated CRISPR Libraries for Genome Labeling and Screening.” Developmental Cell 34 (3): 373–78. https://doi.org/10.1016/j.devcel.2015.06.003.

Lawrence, Sarah H., and James G. Ferry. 2006. “Steady-State Kinetic Analysis of Phosphotransacetylase from Methanosarcina Thermophila.” Journal of Bacteriology 188 (3): 1155–58. https://doi.org/10.1128/JB.188.3.1155-1158.2006.

Lee, Michael E., William C. DeLoache, Bernardo Cervantes, and John E. Dueber. 2015. “A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly.” ACS Synthetic Biology 4 (9): 975–86. https://doi.org/10.1021/sb500366v.

Lee, Wonkyu, and Nancy A. DaSilva. 2006. “Application of Sequential Integration for Metabolic Engineering of 1,2-Propanediol Production in Yeast.” Metabolic Engineering. https://doi.org/10.1016/j.ymben.2005.09.001.

Li, Sijin, Tong Si, Meng Wang, and Huimin Zhao. 2015. “Development of a Synthetic Malonyl-CoA Sensor in Saccharomyces Cerevisiae for Intracellular Metabolite Monitoring and Genetic Screening.” ACS Synthetic Biology 4 (12): 1308–15. https://doi.org/10.1021/acssynbio.5b00069.

Li, Xiaowei, Daoyi Guo, Yongbo Cheng, Fayin Zhu, Zixin Deng, and Tiangang Liu. 2014. “Overproduction of Fatty Acids in Engineered Saccharomyces Cerevisiae.” Biotechnology and Bioengineering 111 (9): 1841–52. https://doi.org/10.1002/bit.25239.

Löbs, Ann Kathrin, Ronja Engel, Cory Schwartz, Andrew Flores, and Ian Wheeldon. 2017. “CRISPR-Cas9-Enabled Genetic Disruptions for Understanding Ethanol and Ethyl Acetate Biosynthesis in Kluyveromyces Marxianus.” Biotechnology for Biofuels 10 (1). https://doi.org/10.1186/s13068-017-0854-5.

Lu, Hongzhong, Feiran Li, Benjamín J. Sánchez, Zhengming Zhu, Gang Li, Iván Domenzain, Simonas Marcišauskas, et al. 2019. “A Consensus S. Cerevisiae Metabolic Model Yeast8

and Its Ecosystem for Comprehensively Probing Cellular Metabolism." Nature Communications. https://doi.org/10.1038/s41467-019-11581-3.

Lundie, L. L., and J. G. Ferry. 1989. "Activation of Acetate by Methanosarcina Thermophila. Purification and Characterization of Phosphotransacetylase." Journal of Biological Chemistry 264 (31): 18392–96.

Marcišauskas, Simonas, Boyang Ji, and Jens Nielsen. 2019. "Reconstruction and Analysis of a Kluyveromyces Marxianus Genome-Scale Metabolic Model." BMC Bioinformatics. https://doi.org/10.1186/s12859-019-3134-5.

McTaggart, Tami L., Danielle Bever, Shane Bassett, and Nancy A. Da Silva. 2019. "Synthesis of Polyketides from Low Cost Substrates by the Thermotolerant Yeast Kluyveromyces Marxianus." Biotechnology and Bioengineering. https://doi.org/10.1002/bit.26976.

Meadows, Adam L., Kristy M. Hawkins, Yoseph Tsegaye, Eugene Antipov, Youngnyun Kim, Lauren Raetz, Robert H. Dahl, et al. 2016. "Rewriting Yeast Central Carbon Metabolism for Industrial Isoprenoid Production." Nature 537 (7622): 694–97. https://doi.org/10.1038/nature19769.

Meile, L., L. M. Rohr, T. A. Geissmann, M. Herensperger, and M. Teuber. 2001. "Characterization of the D-Xylulose 5-Phosphate/D-Fructose 6-Phosphate Phosphoketolase Gene (Xfp) from Bifidobacterium Lactis." Journal of Bacteriology 183 (9): 2929–36. https://doi.org/10.1128/JB.183.9.2929-2936.2001.

Mendoza, Sebastián N. n.d. “OptKnock Tutorial.” https://opencobra.github.io/cobratoolbox/stable/tutorials/tutorialOptKnock.html.

Mikkelsen, Michael Dalgaard, Line Due Buron, Bo Salomonsen, Carl Erik Olsen, Bjarne Gram Hansen, Uffe Hasbro Mortensen, and Barbara Ann Halkier. 2012. “Microbial Production of Indolylglucosinolate through Engineering of a Multi-Gene Pathway in a Versatile Yeast Expression Platform.” Metabolic Engineering 14 (2): 104–11. https://doi.org/10.1016/j.ymben.2012.01.006.

MyBioSource. n.d. “Mouse Malonyl Coenzyme A ELISA Kit.” https://www.mybiosource.com/prods/ELISA-Kit/Mouse/malonyl-coenzyme-A/malonyl-CoA/datasheet.php?products_id=705127.

Nambu-Nishida, Yumiko, Keiji Nishida, Tomohisa Hasunuma, and Akihiko Kondo. 2017. “Development of a Comprehensive Set of Tools for Genome Engineering in a Cold- A Nd Thermo-Tolerant Kluyveromyces Marxianus Yeast Strain.” Scientific Reports 7 (1). https://doi.org/10.1038/s41598-017-08356-5.

Oud, Bart, Carmen Lisset Flores, Carlos Gancedo, Xiuying Zhang, Joshua Trueheart, Jean Marc Daran, Jack T. Pronk, and Antonius J A van Maris. 2012. “An Internal Deletion in MTH1 Enables Growth on Glucose of Pyruvate-Decarboxylase Negative, Non-Fermentative Saccharomyces Cerevisiae.” Microbial Cell Factories 11. https://doi.org/10.1186/1475-2859-11-131.

Pecota, Douglas C., Vineet Rajgarhia, and Nancy A. Da Silva. 2007. “Sequential Gene Integration for the Engineering of Kluyveromyces Marxianus.” Journal of Biotechnology. https://doi.org/10.1016/j.jbiotec.2006.07.031.

Pedruzzi, Ivo, Niels Bürckert, Pascal Egger, and Claudio De Virgilio. 2000. “Saccharomyces Cerevisiae Ras/CAMP Pathway Controls Post-Diauxic Shift Element-Dependent Transcription through the Zinc Finger Protein Gis1.” EMBO Journal 19 (11): 2569–79. https://doi.org/10.1093/emboj/19.11.2569.

Puigbò, Pere, Eduard Guzmán, Antoni Romeu, and Santiago Garcia-Vallvé. 2007. “OPTIMIZER: A Web Server for Optimizing the Codon Usage of DNA Sequences.” Nucleic Acids Research. https://doi.org/10.1093/nar/gkm219.

Rado, Thomas A., and James A. Hoch. 1973. “Phosphotransacetylase from Bacillus Subtilis: Purification and Physiological Studies.” BBA - Enzymology 321 (1): 114–25. https://doi.org/10.1016/0005-2744(73)90065-X.

Rajkumar, Arun S., Javier A. Varela, Hannes Juergens, Jean Marc G. Daran, and John P. Morrissey. 2019. “Biological Parts for Kluyveromyces Marxianus Synthetic Biology.” Frontiers in Bioengineering and Biotechnology. https://doi.org/10.3389/fbioe.2019.97.

Rasche, Madeline E., Kerry S. Smith, and James G. Ferry. 1997. “Identification of Cysteine and Arginine Residues Essential for the Phosphotransacetylase from Methanosarcina Thermophila.” Journal of Bacteriology 179 (24): 7712–17. https://doi.org/10.1128/jb.179.24.7712-7717.1997.

Ryan, Owen W., Jeffrey M. Skerker, Matthew J. Maurer, Xin Li, Jordan C. Tsai, Snigdha Poddar, Michael E. Lee, et al. 2014. “Selection of Chromosomal DNA Libraries Using a Multiplex CRISPR System.” ELife 3 (August2014): 1–15. https://doi.org/10.7554/eLife.03703.

Sakihama, Yuri, Ryota Hidese, Tomohisa Hasunuma, and Akihiko Kondo. 2019. “Increased Flux in Acetyl-CoA Synthetic Pathway and TCA Cycle of Kluyveromyces Marxianus under Respiratory Conditions.” Scientific Reports. https://doi.org/10.1038/s41598-019-41863-1.

Schabort, Du Toit W.P., Precious Letebele, Laurinda Steyn, Stephanus G. Kilian, and James C. Du Preez. 2016. “Differential RNA-Seq, Multi-Network Analysis and Metabolic Regulation Analysis of Kluyveromyces Marxianus Reveals a Compartmentalised Response to Xylose.” PLoS ONE. https://doi.org/10.1371/journal.pone.0156242.

Schujman, Gustavo E., Luciana Paoletti, Alan D. Grossman, and Diego de Mendoza. 2003. “FapR, a Bacterial Transcription Factor Involved in Global Regulation of Membrane Lipid Biosynthesis.” Developmental Cell. https://doi.org/10.1016/S1534-5807(03)00123-0.

“SGD S288C Reference Genome.” n.d. SGD. https://downloads.yeastgenome.org/sequence/S288C_reference/genome_releases/.

Shi, Shuobo, Yun Chen, Verena Siewers, and Jens Nielsen. 2014. “Improving Production of Malonyl Coenzyme A-Derived Metabolites by Abolishing Snf1-Dependent Regulation of Acc1.” MBio 5 (3). https://doi.org/10.1128/mBio.01130-14.

Shin, Byung Sik, Soo Keun Choi, and Seung Hwan Park. 1999. “Regulation of the Bacillus Subtilis Phosphotransacetylase Gene.” Journal of Biochemistry 126 (2): 333–39. https://doi.org/10.1093/oxfordjournals.jbchem.a022454.

Smith, Justin D., Sundari Suresh, Ulrich Schlecht, Manhong Wu, Omar Wagih, Gary Peltz, Ronald W. Davis, Lars M. Steinmetz, Leopold Parts, and Robert P. St.Onge. 2016. “Quantitative CRISPR Interference Screens in Yeast Identify Chemical-Genetic Interactions and New Rules for Guide RNA Design.” Genome Biology 17 (1): 45. https://doi.org/10.1186/s13059-016-0900-9.

Sonderegger, Marco, Michael Schümperli, and Michael Schu. 2004. “Metabolic Engineering of a Phosphoketolase Pathway for Pentose Catabolism in Saccharomyces Cerevisiae Metabolic Engineering of a Phosphoketolase Pathway for Pentose Catabolism in Saccharomyces Cerevisiae.” Applied and Environmental Microbiology 70 (5): 2892–97. https://doi.org/10.1128/AEM.70.5.2892.

Stemmer, Manuel, Thomas Thumberger, Maria Del Sol Keyer, Joachim Wittbrodt, and Juan L. Mateo. 2015. “CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool.” PLoS ONE. https://doi.org/10.1371/journal.pone.0124633.

Tokuhiro, Kenro, Nobuhiro Ishida, Eiji Nagamori, Satoshi Saitoh, Toru Onishi, Akihiko Kondo, and Haruo Takahashi. 2009. “Double Mutation of the PDC1 and ADH1 Genes Improves Lactate Production in the Yeast Saccharomyces Cerevisiae Expressing the Bovine Lactate

Dehydrogenase Gene." Applied Microbiology and Biotechnology 82 (5): 883–90. https://doi.org/10.1007/s00253-008-1831-5.

Varela, Javier A., Martina Puricelli, Noemi Montini, and John P. Morrissey. 2019. "Expansion and Diversification of MFS Transporters in Kluyveromyces Marxianus." Frontiers in Microbiology. https://doi.org/10.3389/fmicb.2018.03330.

Veiga-Da-Cunha, M., H. Santos, and E. Van Schaftingen. 1993. "Pathway and Regulation of Erythritol Formation in Leuconostoc Oenos." Journal of Bacteriology 175 (13): 3941–48. https://doi.org/10.1128/jb.175.13.3941-3948.1993.

Zhang, Yiming, Guodong Liu, Martin K.M. Engqvist, Anastasia Krivoruchko, Björn M. Hallström, Yun Chen, Verena Siewers, and Jens Nielsen. 2015. "Adaptive Mutations in Sugar Metabolism Restore Growth on Glucose in a Pyruvate Decarboxylase Negative Yeast Strain." Microbial Cell Factories. https://doi.org/10.1186/s12934-015-0305-6.

Appendix A

S. cerevisiae CRISPR/Cas9 integration plasmid library

CRISPR-Cas9 allows multiplexed knockout and knockin of genes, however effective knockin requires effective guides. To ease gene integrations in *S. cerevisiae*, I developed a CRISPR/Cas9-based plasmid library which enables integration of desirable genes into up to 18 sites within the yeast genome. These sites allow precise optimization and control of copy number at all times and in all media. This versatile plasmid library utilizes sites that are shown to have high expression and have minimal impact on cell fitness (Mikkelsen et al. 2012), in addition to the four auxotrophic marker loci *URA3, MET15, HIS3 and LEU2*.

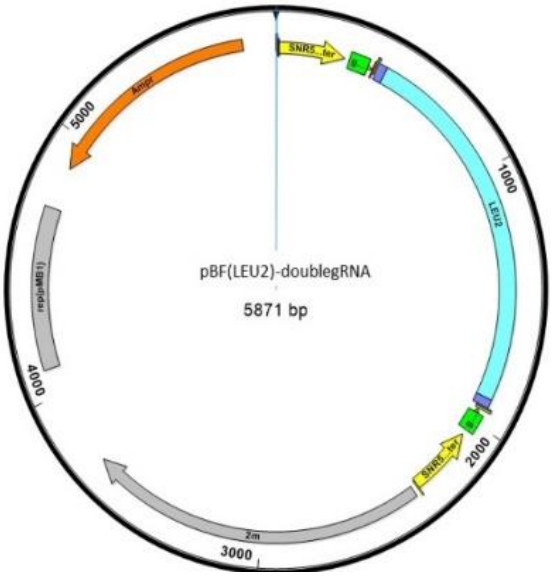


Figure 34 The dgRNA plasmid used for integration of 2PS genes into the *HIS3* and *LEU2* loci. A similar plasmid with different gRNA fragments (green) was constructed for targeting to *URA3* and *MET15*.

The first four integrations into the auxotrophic loci were accomplished using a double gRNA-bearing plasmid (dgRNA, green) (**Figure 34**) that is structurally stabilized by reversal of the structural gRNA sequences which flank the auxotrophic marker, based on p426-*SNR52*p-gRNA.CAN1.Y-SUP4t (from Addgene Church, 2014).

Yeast transformation with dgRNA and Cas9 plasmid coupled with PCR-generated 2-PS donor fragments enabled integration of one and two copies of ADH2p-2PS into both the genome of BYt and BYeng via CRISPR/Cas9 at the *HIS3* and *LEU2* loci.

CRISPR gRNA plasmids were also designed to target 14 additional sites within the genome as identified by Mikkelsen et al (Mikkelsen et al. 2012) using parts generated using the yeast toolkit strategy (M. E. Lee et al. 2015), which combines plasmid "parts" in a modular system using golden gate assembly. The yeast toolkit (M. E. Lee et al. 2015) enables modular design of plasmids that enable straightforward multiplexing of gRNAs into one plasmid. The design of the quadruple guide RNA plasmid (qgRNA) is demonstrated in **Figure 35**. These gRNA parts can be used in future studies to rapidly integrate up to four genes at once. After multiple plasmid curing rounds (performed rapidly with *URA3* selection/5FOA counter-selection), we can integrate genes into up to 18 sites.

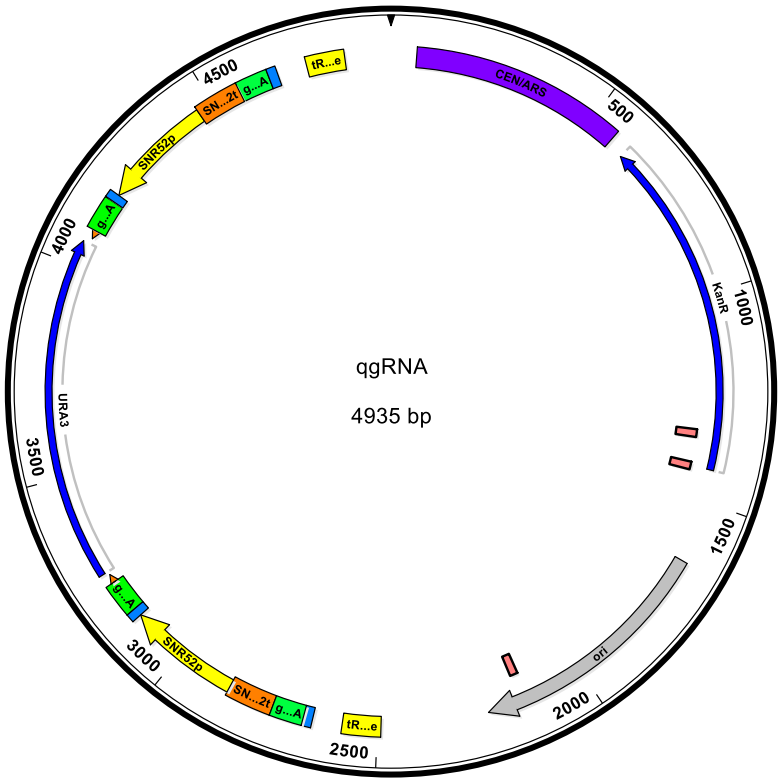
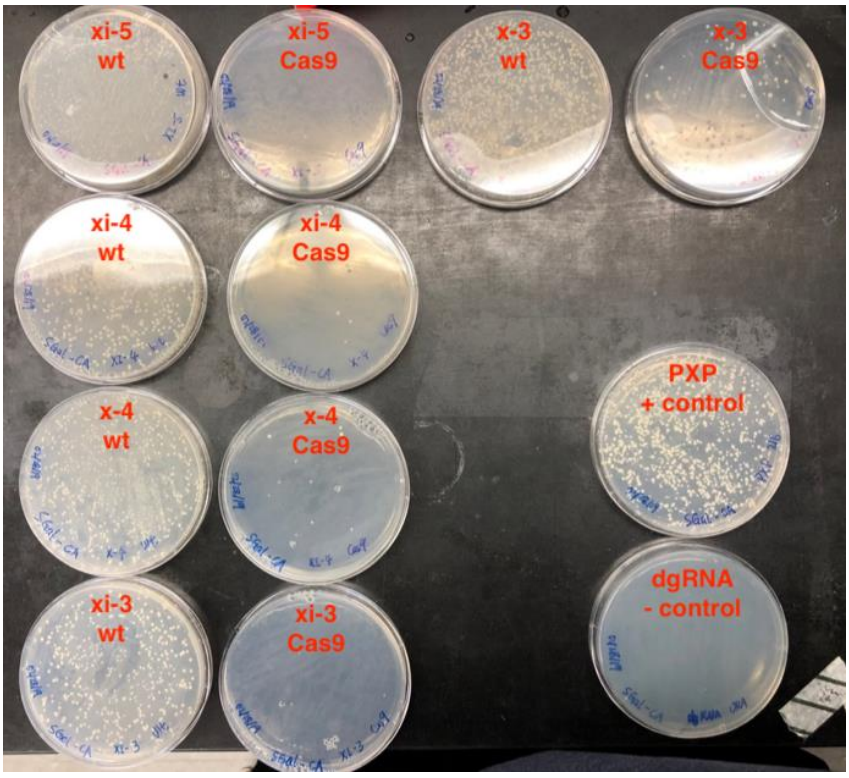


Figure 35 Design of the qgRNA plasmid, with promoters (yellow), 20bp gRNA (med blue), structural gRNA (green), terminators (orange). This plasmid includes CEN/ARS yeast origin, URA3 auxotrophic marker, and E. coli origin and KanR marker.

The single gRNA plasmids for all 14 sites were constructed via primer annealing and then Golden Gate assembly into a gRNA entry vector yTK050. These plasmids were then grouped into sets of four and given different assembly connectors for future construction into

quadruple-gRNA plasmids. Since there is significant homology among the gRNA promoters and structural components, we designed the 4-site plasmids (qgRNA plasmids) with two promoter/terminator sets that were reversed on opposite strands to ensure structural stability and prevent spontaneous recombination of these regions. These cassettes are incorporated to flank the *URA3* auxotrophic marker. Using the two different promoter/terminator pairs (*SNR52*p/SUP4t and tRNAPhep/*SNR52*t), one set from the yeast toolkit (M. E. Lee et al. 2015) and the other from the lab's previous gRNA plasmids, up to 4 gRNA sites can be accommodated stably.

These gRNAs in entry vectors were also used to generate a set of 14 single gRNAs for when integration without multiplexing is desired. All single gRNAs were assembled with the *URA3* selection marker for ease of 5-FOA screening to remove the gRNA, on a 2µ-based plasmid.



To demonstrate the efficacy of these gRNA plasmids, each plasmid was transformed into *S. cerevisiae* with and without Cas9 plasmid p414-TEF1p-Cas9-CYC1t-TRP1 and without any donor DNA. Since *S. cerevisiae* lacks

nonhomologous end joining, an effective gRNA will cause cell death and no colonies to form. From all 18 sites, as well as the 4 auxotrophic marker sites, the single gRNA plasmids demonstrated effective cutting, meaning they are effective gRNAs for CRISPR. I further validated these sites by integrating 2-PS into four of the sites: X-1, X-2, XI-1 and XI-2, which was verified by PCR of gDNA with one primer internal to the 2-PS gene and one directly outside the targeted region. Integration of 2-PS into these sites demonstrates their versatility for generalized use as integration sites into *S. cerevisiae* that extend beyond the traditional auxotrophic marker sites.

Lee, Michael E., William C. DeLoache, Bernardo Cervantes, and John E. Dueber. 2015. "A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly." ACS Synthetic Biology 4 (9): 975–86. https://doi.org/10.1021/sb500366v.

Mikkelsen, Michael Dalgaard, Line Due Buron, Bo Salomonsen, Carl Erik Olsen, Bjarne Gram Hansen, Uffe Hasbro Mortensen, and Barbara Ann Halkier. 2012. "Microbial Production of Indolylglucosinolate through Engineering of a Multi-Gene Pathway in a Versatile Yeast Expression Platform." Metabolic Engineering 14 (2): 104–11. https://doi.org/10.1016/j.ymben.2012.01.006.

Appendix B

COBRA toolbox model and OptKnock code example

Example Python code using a subset of the model's reactions for optimization of Pyruvate

```
initCobraToolbox(false)
changeCobraSolver ('gurobi', 'all');
%model = readCbModel('iSM996');
%model = readCbModel('iSM996_Csav');
model = readCbModel('iSM996_C_NoTAL');


%max number of solutions to find
threshold = 5;

%all rxns in the system
selectedRxnList = model.rxns;

%add acetyl-CoA transport and exchange reactions for original model
%model = addMetabolite(model, 's_0377[e]', 'metFormula', 'C23H34N7O17P3S');
%model = addReaction(model, 'r_1914', 'metaboliteList', {'s_0377[e]'}, 'stoichCoeffList', -1);
%model = addReaction(model, 'r_1915', 'reactionFormula', 's_0377[e] -> s_0373[c]');
%model = changeRxnBounds(model, 'r_1914', 1000, 'u');
%model = changeRxnBounds(model, 'r_1914', 0, 'l');
%model = changeRxnBounds(model, 'r_1915', -1000, 'l');
%model = changeRxnBounds(model, 'r_1915', 1000, 'u');

%add acetyl-CoA transport and exchange reactions for csaving model
model = addMetabolite(model, 's_0377[e]', 'metFormula', 'C23H34N7O17P3S');
model = addReaction(model, 'r_1917', 'metaboliteList', {'s_0377[e]'}, 'stoichCoeffList', -1);
model = addReaction(model, 'r_1918', 'reactionFormula', 's_0377[e] -> s_0373[c]');
model = changeRxnBounds(model, 'r_1917', 1000, 'u');
model = changeRxnBounds(model, 'r_1917', 0, 'l');
model = changeRxnBounds(model, 'r_1918', -1000, 'l');
model = changeRxnBounds(model, 'r_1918', 1000, 'u');

%set reaction bounds
model = changeRxnBounds(model, 'r_1770', -6, 'l'); %beta-D-glucose exchange
model = changeRxnBounds(model, 'r_1725', -11, 'l'); %oxygen exchange
model = changeRxnBounds(model, 'r_1728', -1000, 'l'); %sulphate exchange
model = changeRxnBounds(model, 'r_1727', -1000, 'l'); %ammonium exchange
model = changeRxnBounds(model, 'r_1729', -1000, 'l'); %phosphate exchange
model = changeRxnBounds(model, 'r_1803', 1000, 'u'); %ethanol
model = changeRxnBounds(model, 'r_1759', 1000, 'u'); %acetate
```

```
model = changeRxnBounds(model, 'r_1775', 1000, 'u'); %CO2
model = changeRxnBounds(model, 'r_1814', 1000, 'u'); %glycerol
model = changeRxnBounds(model, 'r_1878', 1000, 'u');  %pyruvate

%calculate and print initial conditions
fbaWT = optimizeCbModel(model);
pyruvateFluxWT = fbaWT.x(strcmp(model.rxns,'r_1878'));
etohFluxWT = fbaWT.x(strcmp(model.rxns,'r_1803'));
acetateFluxWT = fbaWT.x(strcmp(model.rxns,'r_1759'));
glycerolFluxWT = fbaWT.x(strcmp(model.rxns,'r_1814'));
CO2FluxWT = fbaWT.x(strcmp(model.rxns,'r_1775'));
growthRateWT = fbaWT.f;

fprintf('The growth rate before optimization is %.3f \n', growthRateWT);
fprintf('The production of pyruvate before optimization is %.3f \n', pyruvateFluxWT);
fprintf('The production of ethanol before optimization is %.3f \n', etohFluxWT);
fprintf('The production of glycerol before optimization is %.3f \n', glycerolFluxWT);
fprintf('The production of acetate before optimization is %.3f \n', acetateFluxWT);
fprintf('The production of CO2 before optimization is %.3f \n', CO2FluxWT);

% Set optKnock options
% The synthesis of pyruvate will be the objective of the outer problem
options = struct('targetRxn','r_1878','numDel', 2);
%options = struct('targetRxn','r_1878','numDel', 3); not quite working :(


%Impose that biomass be at least 50% of the biomass of WT
constrOpt = struct('rxnList',{{'r_1913'}},'values',0.5*fbaWT.f,'sense','G');

% We will try to find 10 optKnock sets of a maximun length of 2
previousSolutions = cell(10, 1);
contPreviousSolutions = 1;
nIter = 1;
while nIter < threshold
fprintf('...Performing optKnock analysis...\n')
if isempty(previousSolutions{1})
optKnockSol = OptKnock(model, selectedRxnList, options, constrOpt);
else
optKnockSol = OptKnock(model, selectedRxnList, options, constrOpt, previousSolutions, 1);
end

%determine pyruvate production and growth rate after optimization
pyruvateFluxM1 = optKnockSol.fluxes(strcmp(model.rxns,'r_1878'));
```

```
growthRateM1 = optKnockSol.fluxes(strcmp(model.rxns,'r_1913'));
etohFluxM1 = optKnockSol.fluxes(strcmp(model.rxns,'r_1803'));
glycerolFluxM1 = optKnockSol.fluxes(strcmp(model.rxns,'r_1814'));
acetateFluxM1 = optKnockSol.fluxes(strcmp(model.rxns,'r_1759'));
CO2FluxM1 = optKnockSol.fluxes(strcmp(model.rxns,'r_1775'));
setM1 = optKnockSol.rxnList;

if ~isempty(setM1)
previousSolutions{contPreviousSolutions} = setM1;
contPreviousSolutions = contPreviousSolutions + 1;

%printing results
fprintf('optKnock found a optKnock set of large %d composed by ', length(setM1));
for j = 1:length(setM1)
if j==1
fprintf('%s', setM1{j});
elseif j == length(setM1)
fprintf(' and %s', setM1{j});
else
fprintf(', %s', setM1{j});
end
end

fprintf('\n');
fprintf('The production of pyruvate after optimization is %.3f \n', pyruvateFluxM1);
fprintf('The growth rate after optimization is %.3f \n', growthRateM1);
fprintf('The production of ethanol after optimization is %.3f \n', etohFluxM1);
fprintf('The production of glycerol after optimization is %.3f \n', glycerolFluxM1);
fprintf('The production of acetate after optimization is %.3f \n', acetateFluxM1);
fprintf('The production of CO2 after optimization is %.3f \n', CO2FluxM1);
[type, maxGrowth, maxProd, minProd] = analyzeOptKnock(model, setM1, 'r_1878');
fprintf('The solution is of the type: %s\n', type);
fprintf('The maximum growth rate given the optKnock set is %.2f \n', maxGrowth);
fprintf(['The maximum and minimum production of pyruvate given the optKnock set is ' '%.2f
and %.2f, respectively \n\n'], minProd, maxProd);
if strcmp(type, 'growth coupled')
singleProductionEnvelope(model, setM1,
'r_1878','r_1913','savePlot',1,'showplot',1,'fileName',['pyruvate'
num2str(nIter)],'outputFolder','OptKnockResults');
end,
else
if nIter == 1
fprintf('optKnock was not able to found an optKnock set\n');
```

```
else
fprintf('optKnock was not able to found additional optKnock sets\n');
end
break;
end
nIter = nIter + 1;
end
```

The adapted *K. marxianus* model that includes the carbon-saving pathway genes PTA and XPK can be accessed at https://www.dropbox.com/s/k35yqo5swo88bmp/iSM996_Csav.xml?dl=0.

Running the model in *K. marxianus* the OptKnock identified one gene to target when optimizing for the products acetyl-CoA and pyruvate with the integrated carbon saving pathway.

Gene Name	**Substrate**	**Optimizing Product**	**With C-saving?**	**Organism**	**Rationale**
GSY2	Glucose	Acetyl-CoA	Yes	*K. marxianus*	Glycogen synthase
YIA6	Glucose	Pyruvate	Yes	*K. marxianus*	Potential pyruvate transporter

Appendix C

Codon optimized xpk and pta genes for *S. cerevisiae* and *K. marxianus*

O. oeni xpk codon optimized for *K. marxianus*:

ATG GCC GCT AAC GCC CCA GCC GTA GAC TAT GAT TCA ACA GAA TAT TTA GAG CTT GTT GAT
AAA TGG TGG AGG GCT ACG AAT TAT TTA TCA GGC GGT ATG ATT TTC CTA AAG AAT AAT CCG
CTT TTC TCA ATA ACT AAA ACT CCA TTA AAA AGT GAG GAT GTA AAG TTT AAG CCC ATC GGT
CAC TGG GGA ACA ATC TCA GGC CAA ACA TTT ATA TAT GCA CAT GAG AAT CGT TTA ATT AAT
AAG TAT GGG TTA AAC GCC TTT TAC GTG GAA GGT CCC GGG CAT GGG GGA CAG GTA ATG GTA
ACA AAT TCC TAC TTG GAT GGA TCG TAT ACC GAG GAT TAC CCT GAA ATA ACA CAA GAC CTA
AAA GGC ATG TCA AAA CTA TTC AAA AGG TTC TCC TTT CCA GGA GGC ATC GGA TCT CAT GCT
ACC GCT CAG ACG CCA GGC TCA CTT CAC GAA GGT GGA GAA TTA GGC TAC AGT TTA TCA CAT
GGT ATA GGT GCC ATA CTA GAT AAT CCC GAC CAA GTT GGT TTC GTT ACT GTA GGG GAT GGC
GAG GCC GAG ACC GGT CCT GCC ATG ACG GCC TGG CAC GGG ATA AAA TTT ATA AAC CCC AAA
ACA GAT GGC GCC GTC CTA CCT ATA TTG GAC TTG AAC GGA TGG AAG ATC TCA AAC CCG ACC
ATT TTC TCA AGA ATG AGT GAT GAA CAA ATC GCT AAG TTT TTT GAA GGA TTG GGG TGG AGT
CCC CGT TTT TTG GAG AAT GAC GAA ATT CAC GAT TTC ATG ACA TAC CAT AAG AAA GCT GCT
AAA TTG TTT GAC CAG GCT ATA GAA GAT ATT AAG CAA ATA CAA AAG GAT GCC AGA GAG AAT
AAC AAA TAT CAG GAT GGG ACG ATC CCC GCT TGG CCG GTC ATA ATT GCC AGA CTA CCG AAG
GGC TGG GGC GGG CCA AAA TTT GAC GAG GAC GGT AAC CCC ATA GAA AAC TCG TTT CGT GCC
CAT CAA GTG CCA CTT AAC TTT TCG GCT GAG CAT ATG GAG GAA CTT CCG CAG TTC GAG GAG
TGG ATG AAC TCA TAT AAG CCG GAG GAG TTG TTT AAT GAA GAC GGA TCC CTA AAG GCT GAA
ATT TCC GCC ATA GCA CCA AAA GGG TCT AAG AGG ATG GCT GCC AAT CCT TTA GCT AAT GCC
GGT GTC GAT AAT TCT GAC CTA AAG CTA CCA GAC TGG AAG GAG TAC TCC ACT GGT GTC ACG
CCC GAG AAC CGT GGC ACT GAA ATG AAA GAT GCA AAC ATG AAC ATG GAC ATG GTC ACC CTA
TCA GGA TTC TTG GCT GGA GTC GCA AAA CTA AAT CCA ACA AGA TTT AGA TTT TTT GGA CCC
GAT GAA ACT ATG TCC AAT AGG TTA TGG AAA TTA TTT GAT AAG ACC CCT CGT CAG TGG ATG
AGT AAG ATC AAG TTT CCA AAT GAT GCC CTT CTT GCA CCT GAG GGG CGT ATT ATT GAT TCG
CAG TTA TCC GAG CAT GAA GCC GAG GGC TGG TTA GAG GGG TAC ACA TTA ACG GGC CGT GTG
GGC ATG TTT GCT AGT TAC GAA TCC TTT CTT AGA GTC GTG GAC TCG ATG ATT AAC GAA TAC
TTC AAG TGG ATT AGA CAG GCA GAC GCT GAG CCC TGG AGG AAC AAA TAT CAG TCC TTA AAC
TTA ATT TCC ACT AGT ACC GTG TTC CAG CAA GAT CAT AAT GGA TAT ACG CAC CAA GAC CCG
GGT ATG TTA ACA AAC CTT GCT GAG AAA AAG CAA AAC TAT ATC AGA CAA TAT CTT CCC GCA
GAT GGC AAC GAG TTA TTA GCT GTT GCC TCA AGG GCA CTA GTC GAT CGT CAG AAA ATC AAC
CAT ATA GTC GCT TCA AAG CAA CCC AGA CAG CAG TGG TTT ACA GCA GAG GAG GCA GAG AAA
TTA GTA AAC AAT GGA CTT GGG ATC GTT GAT TGG GCC TCT ACG TCT CCT GAC GGA GAC GTA
GAT ATT ACT TTT GCT TCG GCC GGC ACA GAG CCC ACA ATC GAA ACC CTT GCT GCA TTG TGG
CTA GTA AAC CAG AGT TTC CCA GAC GTT AAA TTT CAC TAC GTA AAT GTT GTA GAG CTA GGA
CGT TTG CAA AAA AAG GTT GGA CCC TTA AAC GAT GAT AGA TCA TTG TCA GAC GAA GAG TTC
GAC AAA TTC TTC CCC GCT GGC AAA CCA GTG GTC TTC GGC TTT CAC GGC TTT GAA GAT TTA
ATA GAA GCA ATC TTT TAT GAA AGG AAC CAC CAT AAT TTA CAC GCA CAC GGG TAC AGG GAA
GAT GGG GAC ATA ACG ACG CAG TAC GAC ATG AGG GTG TAT TCA CAT TTA GAC AGG TTC CAC
CAG GCT ATA GAT GCT GTT TCG GCC TTG AAG GAC GAG GGG GTA ATA GAC GGC GAG AAG GCT
GAT TCG TTT ATA GAT GAT ATG AAT AAG ATT TTA GAT AAG CAC TTT GAA GTT ACA CGT AAC
GAG GGT AGG GAT ATT GAA GAG TTC ACA GAT TGG AAT TGG ACT GCT CTA AAA TAA

O. oeni xpk codon optimized for *S. cerevisiae*:

A TG GCT GCG AAC GCG CCC GCT GTG GAC TAT GAC TCT ACT GAG TAC TTG GAA CTG GTG GAT AAG TGG TGG CGT GCT ACG AAT TAC CTT TCC GGC GGG ATG ATA TTC TTA AAA AAC AAT CCC TTG TTT TCC ATA ACG AAA ACC CCA CTT AAA TCT GAA GAT GTG AAG TTT AAG CCG ATA GGT CAC TGG GGA ACA ATA AGT GGT CAG ACC TTC ATT TAT GCA CAT GAA AAC AGG CTT ATT AAC AAA TAC GGT CTA AAT GCC TTT TAT GTA GAG GGA CCA GGA CAT GGG GGC CAG GTT ATG GTT ACT AAC TCC TAC TTG GAC GGC TCT TAC ACG GAG GAT TAC CCG GAG ATA ACA CAG GAT CTG AAA GGT ATG AGC AAA CTG TTT AAA AGG TTT TCT TTT CCC GGG GGC ATT GGC TCT CAT GCC ACT GCT CAA ACG CCA GGT TCT CTA CAC GAA GGA GGG GAA TTG GGC TAC AGC CTA AGT CAT GGC ATA GGG GCA ATC TTA GAC AAC CCA GAT CAA GTC GGA TTT GTG ACT GTG GGT GAC GGT GAG GCC GAG ACT GGT CCA GCA ATG ACA GCC TGG CAC GGA ATT AAA TTC ATA AAC CCG AAG ACC GAT GGA GCT GTT TTG CCT ATA TTG GAC CTG AAC GGG TGG AAA ATT AGT AAC CCT ACC ATT TTT AGT AGG ATG TCT GAC GAA CAA ATT GCG AAA TTT TTC GAG GGA TTG GGC TGG TCT CCC AGG TTT CTT GAG AAC GAT GAG ATC CAC GAC TTC ATG ACA TAC CAC AAG AAG GCG GCT AAA TTA TTT GAC CAG GCC ATA GAA GAT ATC AAG CAA ATC CAA AAG GAT GCC CGT GAG AAC AAT AAA TAT CAA GAT GGG ACC ATC CCA GCT TGG CCT GTA ATC ATC GCA AGG CTA CCG AAA GGG TGG GGC GGG CCA AAG TTC GAT GAA GAT GGG AAC CCA ATA GAA AAC TCT TTC CGT GCG CAT CAG GTG CCG TTG AAC TTT TCT GCA GAG CAT ATG GAG GAG CTA CCA CAG TTT GAA GAA TGG ATG AAT TCT TAT AAA CCT GAA GAG TTG TTT AAC GAG GAT GGT TCT TTG AAG GCG GAG ATA AGC GCG ATC GCT CCC AAG GGC AGC AAA AGA ATG GCA GCA AAC CCC TTA GCG AAC GCG GGT GTC GAC AAT TCA GAT CTA AAG CTA CCT GAC TGG AAG GAG TAC TCT ACC GGC GTG ACC CCT GAG AAT AGG GGG ACC GAA ATG AAA GAC GCG AAT ATG AAC ATG GAT ATG GTA ACC CTT TCA GGG TTT CTA GCT GGC GTT GCC AAA CTT AAC CCC ACG AGA TTT AGG TTT TTT GGA CCC GAT GAA ACA ATG TCC AAC AGG CTT TGG AAG CTT TTC GAT AAA ACA CCC CGT CAG TGG ATG AGT AAG ATT AAG TTC CCT AAT GAC GCA CTT TTG GCT CCT GAG GGA AGA ATA ATT GAC TCC CAG TTA TCT GAA CAC GAA GCA GAG GGG TGG CTT GAA GGG TAT ACT CTT ACT GGC AGA GTT GGC ATG TTT GCG AGC TAC GAA TCA TTC TTG AGA GTA GTG GAC TCA ATG ATA AAC GAG TAC TTC AAG TGG ATA CGT CAG GCA GAT GCG GAG CCA TGG CGT AAC AAG TAT CAA TCT CTT AAT CTA ATT AGC ACG AGC ACA GTC TTT CAG CAG GAT CAT AAC GGC TAT ACG CAC CAG GAC CCG GGT ATG CTA ACA AAC TTG GCC GAA AAG AAA CAA AAT TAC ATA AGA CAG TAC TTG CCG GCC GAC GGA AAC GAA CTA CTT GCA GTC GCA AGC AGG GCA TTA GTA GAT AGG CAG AAG ATC AAC CAC ATT GTT GCG TCC AAG CAA CCA AGG CAG CAA TGG TTC ACT GCA GAG GAG GCA GAG AAA CTG GTT AAC AAT GGC CTT GGT ATT GTA GAC TGG GCT TCA ACT TCT CCT GAT GGC GAC GTA GAT ATT ACT TTC GCC AGC GCC GGA ACA GAA CCG ACA ATC GAG ACG CTA GCC GCT TTG TGG CTT GTT AAT CAG TCA TTT CCT GAC GTT AAG TTC CAC TAC GTT AAC GTA GTT GAG CTA GGG AGA CTT CAG AAA AAG GTG GGA CCC TTG AAT GAC GAT AGG AGC CTG TCT GAC GAA GAG TTT GAC AAG TTT TTT CCG GCG GGC AAG CCC GTC GTA TTC GGC TTT CAC GGT TTT GAA GAC CTA ATA GAG GCC ATA TTT TAT GAA CGT AAT CAC CAC AAC TTG CAC GCT CAT GGC TAT AGG GAG GAT GGG GAT ATC ACG ACG CAG TAT GAT ATG CGT GTA TAC TCT CAC CTG GAT CGT TTT CAC CAG GCG ATT GAC GCC GTG AGC GCC TTA AAA GAC GAG GGC GTG ATA GAC GGC GAG AAG GCA

GAC TCT TTC ATA GAT GAT ATG AAT AAA ATT TTG GAC AAA CAC TTT GAA GTA ACT AGA AAT GAG GGG CGT GAT ATT GAA GAG TTC ACC GAT TGG AAT TGG ACG GCG CTT AAA TAA

M. thermo pta codon optimized for *K. marxianus*

TTG GTC ACT TTT TTG GAA AAA ATA TCC GAG AGA GCC AAG AAA TTA AAC AAA ACC ATC GCA TTA CCT GAG ACT GAA GAT ATC CGT ACG CTT CAA GCT GCT GCT AAG GTC CTT GAA AGG GGT ATA GCA AAC GTG GTG CTT ATT GGC AAA GAA AAA GAT ATC AAA GAG TTA AGT GGG GAC TTG GAC TTG TCG AAA GCC CGT ATA GTC GAC CCG GAG ACT TAT GAA CGT AAA GAC GAG TAC GTG AAG ACG TTC TAT GAG CTT CGT AAG CAT AAA GGT GTA ACA CTT GAC AGT GCA GCC GAA ATC ATG AAG GAC TAC GTC TAC TTC GCT GTG ATG ATG GCC AAA TTG GGT GAG GTA GAC GGA GTG GTT TCT GGA GCT GTG CAT TCG TCC TCC GAT ACC CTA AGA CCA GCA GTT CAA ATC GTA AAG ACC GCA CCA GAC TCG GCT CTT GCT TCA GCC TTT TTT ATT ATT TCA GTA CCT GAC TGT GAG TAT GGA TCC AAT GGT ACC TTT TTA TTT GCC GAC TCA GGC ATG GTC GAG ATG CCG ACA GTT GAA GAG TTA GCC CAT ATA GCC GTT ACC TCC GCC AAG ACT TTC GAG TTG CTA GTA CAA GAT ACC CCT TAT GTG GCT ATG TTG TCA TAT AGT ACG AAG GGT TCA GCT CAC TCT AAG CTA ACA GAG GCA ACC GTT GCT GCT ACT AAA AGA GCA CAA GAG TTA GCT CCG GAT ATT GCC ATC GAC GGA GAA CTA CAA GTC GAT GCT GCC ATT GTT CCA AAG GTT GCC GCA TCA AAG GCT CCT GGT TCA CCC GTT GCC GGC AAA GCT AAC GTA CTA ATC TTC CCT GAC CTA AAT GCA GGA AAT ATA GCC TAT AAA ATT GCA CAT CGT TTA GCC AAA GCA GAG GCT TAT GGC CCG ATT ACG CAG GGC TTG GCT AAA CCT ATA AAC GAC TTG TCG CGT GGC TGT TCG GAC GAA GAC ATA GTA GGA GCA GTG GCC ATT ACC TGT GTC CAA GCC GCT GCA CAA CAG AAG TAA

M. thermo pta codon optimized for *S. cerevisiae*

TTG GTC ACG TTC CTG GAG AAA ATT TCC GAG CGT GCC AAG AAG TTA AAC AAG ACC ATT GCA CTT CCA GAA ACT GAG GAC ATA AGG ACG CTA CAA GCC GCC GCC AAA GTC CTT GAG AGG GGC ATC GCA AAC GTA GTA CTA ATT GGA AAG GAG AAA GAC ATT AAA GAA TTA TCA GGG GAT TTA GAT CTA TCC AAG GCC AGA ATC GTG GAC CCA GAA ACG TAT GAG CGT AAA GAT GAA TAT GTT AAA ACT TTC TAT GAA TTA AGG AAA CAC AAG GGT GTA ACC TTG GAC TCT GCT GCG GAA ATA ATG AAA GAC TAC GTC TAT TTT GCC GTA ATG ATG GCG AAG CTG GGG GAA GTG GAT GGG GTC GTA AGT GGA GCG GTT CAT TCA TCA TCT GAC ACA TTA AGA CCG GCC GTG CAA ATC GTA AAG ACG GCG CCC GAC AGT GCT TTG GCC TCA GCA TTT TTC ATT ATC AGC GTG CCA GAC TGC GAG TAC GGT AGT AAT GGC ACG TTC CTG TTT GCG GAC TCT GGC ATG GTC GAG ATG CCC ACT GTG GAA GAG TTG GCC CAT ATA GCT GTA ACG AGT GCG AAA ACT TTC GAG CTT TTG GTC CAA GAC ACC CCT TAT GTC GCC ATG TTA TCA TAT TCT ACG AAG GGG AGT GCC CAT TCT AAA CTG ACG GAG GCT ACT GTA GCT GCA ACG AAG AGA GCG CAG GAG TTA GCT CCT GAT ATA GCC ATC GAT GGA GAG TTA CAG GTG GAT GCA GCC ATT GTA CCC AAA GTA GCT GCG TCC AAA GCC CCT GGA TCA CCT GTT GCA GGA AAA GCT AAC GTC CTA ATT TTT CCA GAC CTT AAC GCC GGA AAC ATC GCT TAT AAG ATA GCT CAC CGT TTG GCA AAA GCG GAG GCA TAC GGC CCG ATA ACG CAA GGC TTG

GCC AAA CCT ATA AAT GAC CTG TCC CGT GGC TGC TCT GAC GAA GAT ATA GTG GGT GCG GTC GCT ATA ACA TGC GTC CAA GCC GCT GCG CAA CAG AAA TAA

Appendix C

gBlocks for malonyl-CoA sensor

FapR with *URA3* flanking regions:

ATAGTGATGATATTTCATAAATAATGTAATTCTATATATGTTAATTACCTTTTTTGCGAGGCATATTTATG
GTGAAGGATAAGTTTTGACCATCAAAGAAATGccaaagaagaagagaaaggtaggcagcAGAAGAAATAAGAG
AGAACGCCAGGAATTACTTCAGCAGACGATTCAAGCAACCCCCTTTATTACAGATGAAGAACTAGCGGG
TAAATTCGGGGTGAGCATCCAGACGATACGTTTGGACCGCTTAGAGCTTTCCATACCTGAACTGAGAGA
AAGAATTAAGAACGTGGCAGAGAAAACACTTGAGGACGAAGTGAAGTCCCTGTCACTTGATGAAGTTA
TCGGAGAAATTATTGACCTTGAGCTGGATGATCAGGCGATATCCATTTTAGAAATAAAACAGGAGCACG
TGTTCAGCCGGAATCAGATTGCGAGAGGACACCATTTATTTGCACAGGCGAATTCTTTGGCCGTTGCAGT
CATTGATGACGAGCTGGCGCTGACTGCAAGTGCAGACATCCGCTTTACAAGACAGGTAAAGCAGGGTG
AACGTGTCGTAGCAAAAGCGAAAGTGACGGCTGTCGAAAAAGAAAAAGGAAGAACGGTTGTCGAAGT
GAACAGCTACGTTGGCGAAGAAATTGTTTTTTCTGGACGCTTTGACATGTATCGTTCAAAACATTCATAA
GGAATCTCGGTCGTAATGATTTCTATAATGACGAAAAAAAAAAAATTGGAAAGAAAAAGCTTCATGGCC
TTTATAAAAAGGAACTATCCAATACCTCGCC

TEF123 with *ADH2* flanking regions:

CTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGCATGAGGTCGCTCTTATTGACCACACCTCTACCGGc
atgccgagcaaatgcctgcaaatcgctccAGGGTGTCGTTAATTACCCGTACTAAAGGTTTGGAAAAGAAAAAAG
AGACCGCCTCGTTTCTTTTTCTTCGTCGAAAAAGGCAATAAAAATTTTTATCACGTTTCTTTTTCTTGAAAA
TTTTTTTTTTTGATTTTTTTCTCTTTCGATGACCTCCCATTGATATTTAAGTTAATAAACGGTCTTCAATTTCT
CAAGTTTCAGTTTCATTTTTCTTGTTCAATTATATACTACTATTAGTACCTAGTCTTAATTTATTACAACTTT
TTTTACTTCTTGCTCATTAGAAATTATATACTACTATTAGTACCTAGTCTTAATTAAGAAAGCAAATTATAT
ACTACTATTAGTACCTAGTCTTAATTTAGCAATCTAATCTAAGTTTTAATTACAAAATGTCTATTCCAGAAA
CTCAAAAAGCCATTATCTTCTACGAATCAAACGGCAAGTTGGAGCATAAGGATATCCCAGTTCCAAAGCC
AAAGCCCAACGAATTGTTAATCAACGTCAAGTACTCTGGTGTCTGCCACACCGATTTGCA

Appendix D

gRNA library construction method and troubleshooting

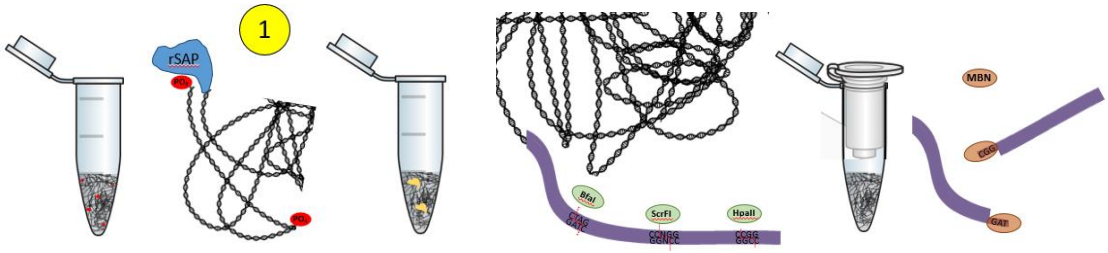
1. Design of primers, and selective phosphorylation of top/bottom strands with T4PNK. Anneal by slowly reducing temperature in CutSmart buffer. Drop dialyze to remove salts.

a. Linker 1 – phosphorylate both top and bottom

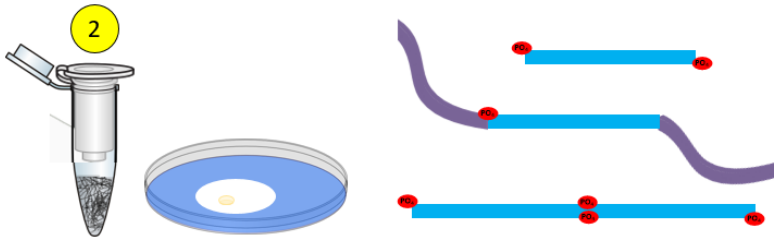
b. Linker 2 – phosphorylate bottom

c. Linker 3 – phosphorylate top

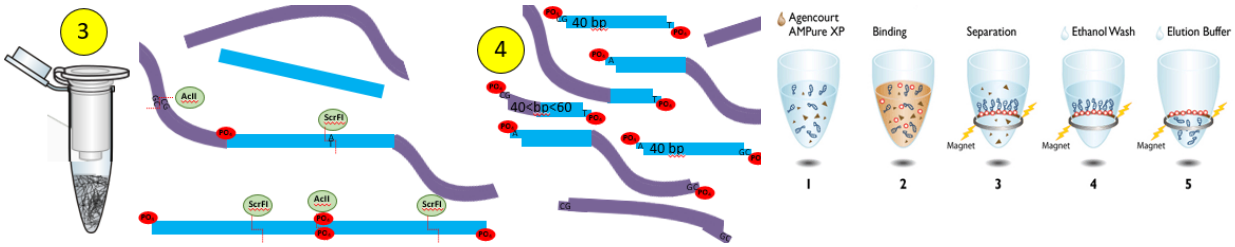
2. Prepare yeast gDNA via phenol/chloroform extraction, dephosphorylate with rSAP for 1hr at 37C to prevent ligation of non-PAM ends and purify over a Zymo Column with oligo binding buffer (OBB). Digest with HpaII, ScrFI, BfaI to cut gDNA at PAM sequences for 1hr at 37C, then purify over a column using OBB. Remove restriction overhangs with mung bean nuclease (MBN) at 25C for 30min to prepare for blunt end ligation.



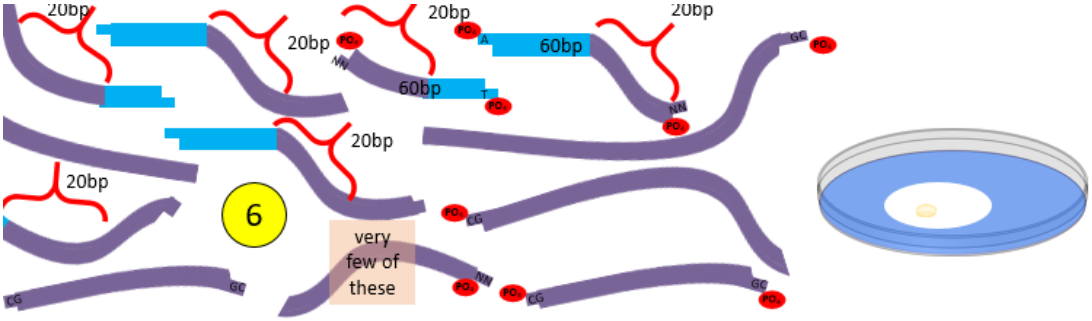
3. Purify over a column with OBB, drop dialyze to desalt then ligate to linker 1 with Quick Ligation Kit for 30min at room temperature.



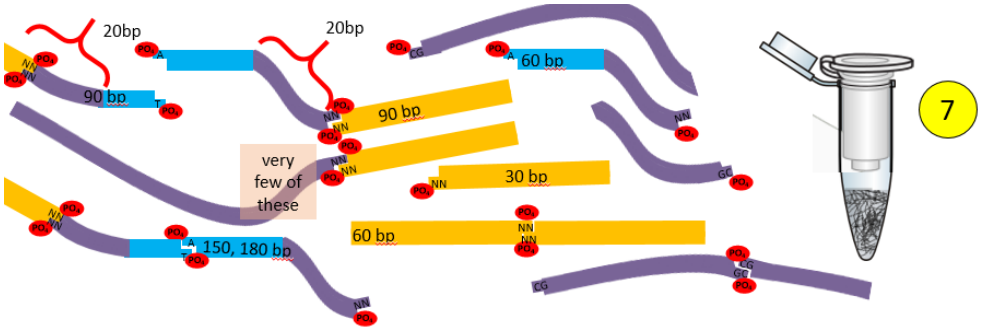
4. Purify over a column with OBB, digest with ScrFI and AclI to break apart linker 1 for 2hr at 37C, purify to remove small fragments <40bp via magnetic Ampure beads



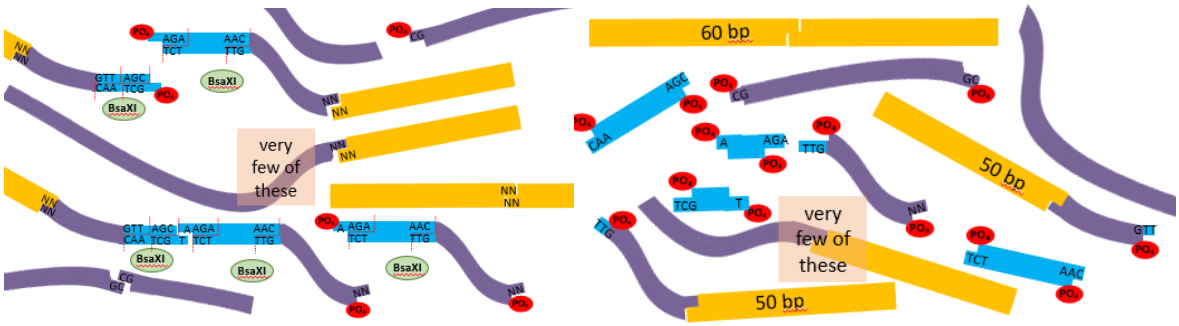
5. Eluent (>60bp) is digested with MmeI to truncate the gRNA to 20bp at 37C for 1hr, then drop dialyzed



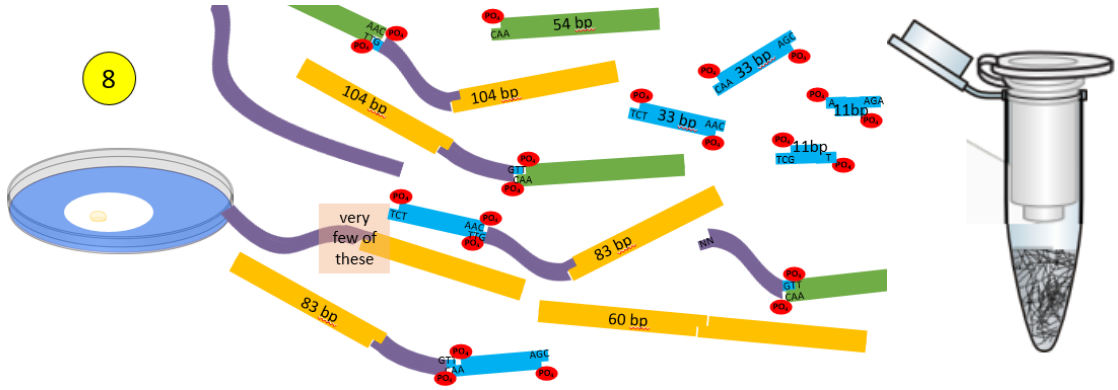
6. Ligate to linker 2 with Quick Ligation Kit for 30min at room temp, then Clean and Concentrate over a column to remove very small and very large fragments



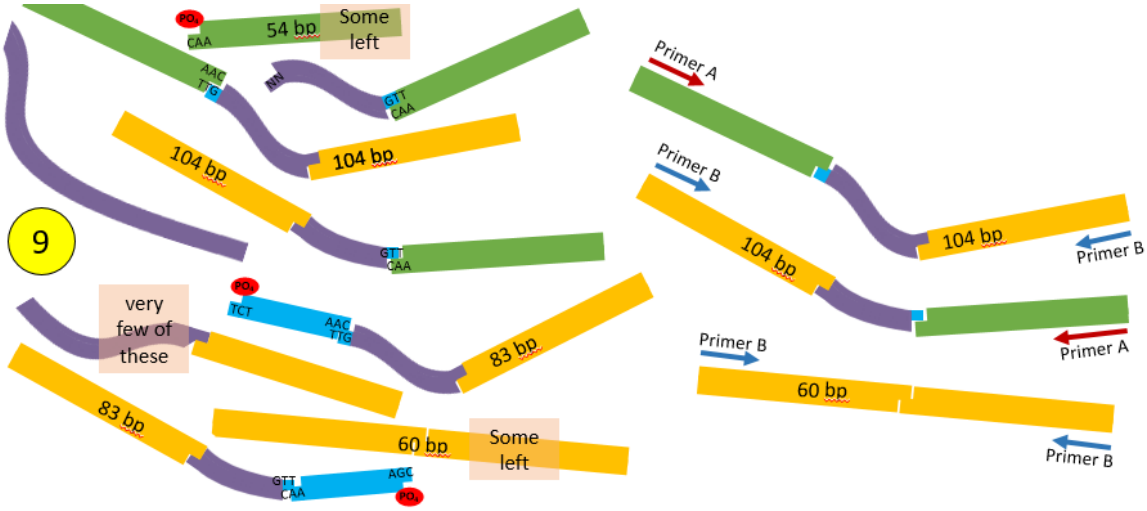
7. Digest with BsaXI at 37C for 1hr



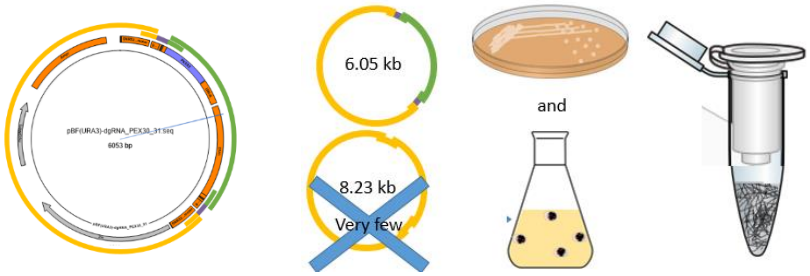
8. Drop dialyze and then ligate to linker 3 with the Quick Ligase Kit for 30min at room temp.



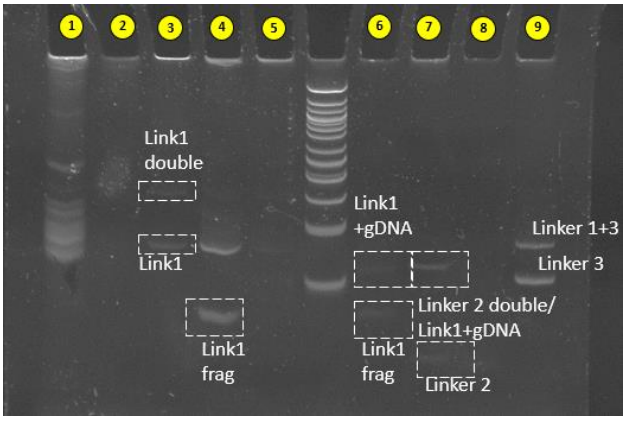
9. PCR amplify clean and concentrated constructs with HotStart 2x master mix for 10 cycles

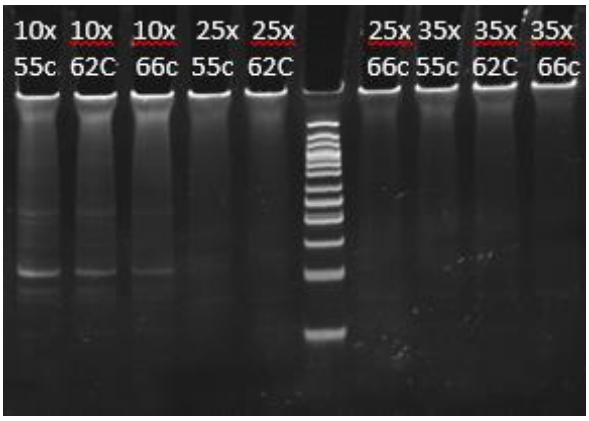


10. Clean and concentrate, then Gibson product into dgRNA backbone, transform into *E. coli*, expand transformation in flasks and extract the plasmid library. Plate a small fraction to evaluate individually and sequence verify.



Troubleshooting steps:





Run aliquots of each step in the process on TBE-Page gels to verify correct fragments are maintained throughout the process. Final 103bp products are PCR amplified using Q5 polymerase at various temeratures and cycles to optimize amplification of the correct-sized fragment. PCR products are run on TBE-Page gels.

Appendix E

Unauthorized access to genetic genealogy services by saliva spoofing

Abstract

Direct-to-consumer (DTC) commercial genetic testing services such as 23andMe and Ancestry.com allow individuals to connect with their ancestry and family members. These services have risen in popularity due to low cost and simple at-home kits, with 23andMe and Ancestry.com boasting over 10 million kits sold. These genealogy services assume honesty from their consumers and do not test their sample intakes to verify saliva composition nor do they verify customer identity. This allows DNA that has been extracted from any source to be used to identify and impersonate someone, even individuals who have otherwise avoided inclusion in biometric databases.

We hypothesize that since saliva is 99% water, extracted DNA could be submitted instead of genuine saliva. We demonstrate this by successfully submitting kits containing pure DNA to the two largest genetic genealogy services, Ancestry.com and 23andMe, and analyzing the genotype. Our results indicate that extracted DNA is indistinguishable from saliva to these companies and is therefore a viable alternative to each service's saliva collection method. A successfully processed kit allows an attacker to both enumerate the relatives of the DNA source and impersonate the owner of that DNA. This exposes anyone with relatives in genetic genealogy services to de-anonymization and users on the platform to manipulation. Our technique constitutes a novel attack on genetic genealogy services with no trivial countermeasure.

Introduction

Genetic genealogy services offer consumers a way to connect with family members and build family trees using direct-to-consumer (DTC) genetic testing. At the time of writing, these genetic databases contain genetic data for over 16 million people ("Ancestry Company Facts," n.d.; "About Us - 23andMe," n.d.) and are expected to grow. These databases are populated with genotype data that includes hundreds of thousands of single nucleotide polymorphisms (SNPs) that are collected from the DNA in each customer's saliva sample. Each new genotype is compared to those already in the database. Individuals with shared DNA are revealed to each other, allowing communication and collaboration between members in a service.

Genetic genealogy providers do not require customers to provide any authentication in order for a sample to be processed. One might assume that it is difficult to collect saliva in sufficient quantities from an unwitting subject. In this way, the nature of saliva can be expected to act as an authenticator. We show that this assumption is false and that, without proper authentication procedures, it is possible for an attacker to access a genetic genealogy database posing as any individual for whom they have genetic material.

We demonstrate this by constructing "spoofed" samples that process successfully at two genealogy providers, but contain no saliva. We chose the two largest providers, Ancestry.com and 23andMe, who together account for over 80% of all DTC kits sold.

In each spoofed kit, we replaced saliva with extracted, purified DNA from oral cells. We hypothesize that because saliva is 99% water (Carpenter 2013), dissolving purified DNA in water will be similar enough to saliva to pass through their system. If this pure DNA is sufficient to

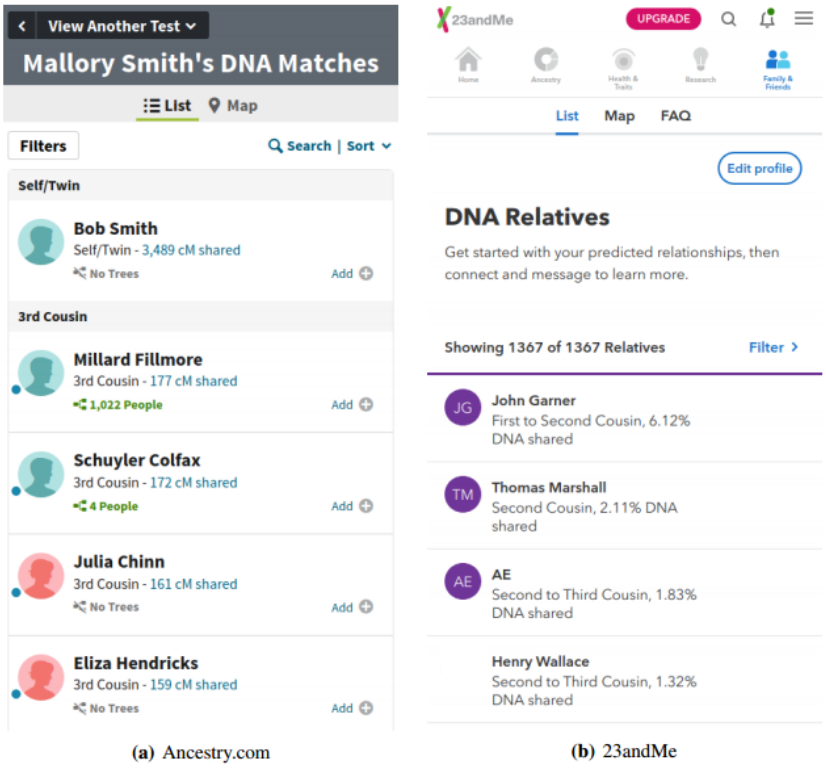
achieve results with commercial providers, then any biological material from which sufficient DNA can be extracted can be used to create a viable sample. This could be discarded material, for example, on a bandage or tissue.

Providing extracted DNA allows an attacker to obtain the genotype of a target individual. Genotypes can provide highly sensitive information, for example both 23andMe and Ancestry.com kits test for at least one gene associated with Alzheimer's disease risk (APOE) (A. M. Saunders et al. 1993). This is not unique to genetic genealogy providers, as there are many services that will genotype or sequence DNA. However, unlike other providers, genetic genealogy DTC services connect an individual to their relatives through the service's extensive database. By disclosing relatives, these services can enable identification of an unknown individual, even if that individual is not in any biometric database (Erlich et al. 2018). Because genetic genealogy services also reveal predicted family members to both parties of a match, an attacker could impersonate an individual's relative, enabling fraud.

Here, we report the first successful processing of spoofed DNA at the two leading DTC providers. We then used the genotypes obtained from these spoofed kits to compare to kits constructed using the expected methods and found that spoofed kits contain true genotyping errors of less than one hundredth of one percent. Furthermore, we find no significant difference in missing SNPs in authentic versus spoofed samples and also report no significant difference in relative matching due to these differences. From this, we demonstrate our hypothesis that extracted DNA dissolved in water is indistinguishable from authentic samples under the current DTC procedures.

The lack of authentication that we have revealed within genetic genealogy services exposes all individuals to unnecessary risk of de-anonymization and fraud. To protect everyone, these organizations must implement new measures for authentication of samples and individuals. As we will explore later, countermeasures fall under two categories: authenticating samples and authenticating individuals.

Background and Related Work



(a) Ancestry.com (b) 23andMe

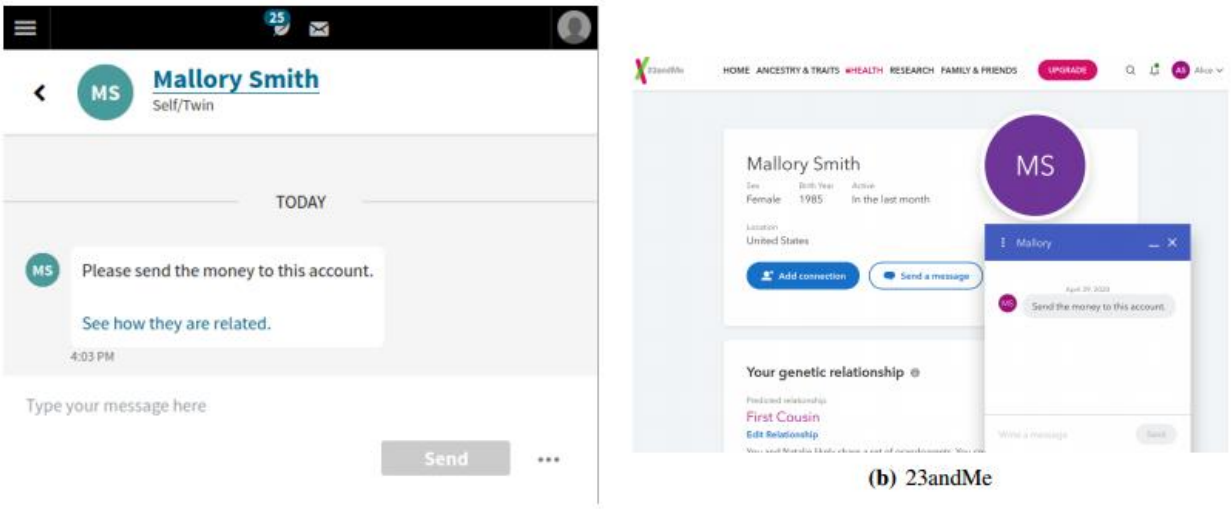
Figure 36 Examples of match lists showing the first five matches. In practice genetic genealogy services

Relative matching has become an integral component of genetic genealogy services. To find relatives, a service generates and then compares a customer's genotype to the genotypes already in their database. These genotypes are composed of hundreds of thousands of single-nucleotide polymorphisms (SNPs), and together, this genotype is unique and identifying to an individual. A SNP is defined as a location in the genome where some proportion of the population has variation. The DTC service searches for other individuals who share long runs of matching SNPs (Gusev et al. 2009; C. A. Ball,

n.d.). Genetic searches leverage the fact that individuals who are more closely related will, on average, share more genetic material. For example, a child will share 50% of their SNPs with their parent. They will also share, on average, 25% of SNPs with a grandparent and about 12.5% with a first cousin.

The matches are displayed to users as a list of individuals, see Figure 36 for an example. Users can explore each other's family trees and profile information. Through links displayed on user profiles users can message each other, as seen in **Figure 37**. In this way, DTC services are also social media services.

While matching may return thousands of relatives, often it is the closer relatives who are of interest. Erlich et al estimated that under reasonable assumptions, in a database with 2% of the population, 99% of individuals of European descent will find at least a third cousin match and 65% of individuals of European descent will find at least a second cousin match (Erlich et al. 2018).



(a) Ancestry.com

(b) 23andMe

Figure 37 Examples of messages on two genetic genealogy platforms. Both platforms display the genetic relationship between two individuals. The Ancestry interface indicates that the messages is between users who are twins or is a duplicate account. The 23andMe interfaces indicates that the users are first cousins.

Recently, law enforcement has used the genetic matching capabilities of some services, namely GEDMatch and FamilyTreeDNA, to perform a new technique called long-range familial searches, or genealogical triangulation (Greytak, Moore, and Armentrout 2019; Lillis 2018). Genealogical triangulation allows law enforcement to identify the "donor" of genetic material.

To perform triangulation on a donor sample, law enforcement will first genotype the sample, then upload the donor genotype to a cooperating service. The service searches for the donor's relatives in their genetic database. The service then provides law enforcement with the list of relatives produced, along with each relative's estimated relationship to the donor (e.g. first cousin, second cousin, etc.). This is much like the list of relatives the donor might see if they were a customer of the service (**Figure 36**). Law enforcement takes this list of relatives and explores each individual's family tree, which is constructed through a combination of public records and police records, to search for individuals common to each family tree. In this way law enforcement "triangulates" on the individual in the intersection of these family trees. Law enforcement may use other data such as age, location, and sex to narrow the list of suspects. In this way, triangulation enables law enforcement to identify the donor.

Triangulation effectively creates an implicit or de-facto population-wide DNA database, enabling identification of individuals outside of any genetic databases. This is in contrast to traditional forensic searches which, due to the lower resolution of data in forensic databases, can only identify DNA from individuals who are already present in a genetic database. Recent work has shown that approximately 60% of individuals with European ancestry can be identified using genealogical triangulation (Erlich et al. 2018). Law enforcement has utilized these databases to solve numerous cases, one notable case being the Golden State Killer (Lillis 2018). The robustness

of genealogical triangulation methods enable its use as a new tool for law enforcement, but this ease of use also opens the door to misuse.

GEDMatch and FamilyTreeDNA have been the choice for law enforcement because they do not require warrants. The services we explore in this paper currently require warrants for all law enforcement requests. GEDMatch is also different from other providers in that they do not provide DTC kits. Instead, GEDMatch users upload genotype results from other providers. Previous research on GEDMatch has found that upload-based services expose private genetic information to attackers who are able to upload artificially crafted genotypes (Ney, Ceze, and Kohno 2020; Edge and Coop 2020). In these works the authors craft artificial genotypes which, when uploaded, cause the services to output match information that reveals sensitive genetic information about the service's customers.

While it is possible to upload a genotype to GEDMatch in order to perform triangulation as law enforcement does, GEDMatch's database is much smaller than those run by other providers. In June 2020 GEDMatch had 1.3 million profiles (“GedMatch” 2020) whereas Ancestry.com has over 16 million (“Ancestry Company Facts,” 2020). This means many fewer close matches, which can be crucial for identification, would be expected from GEDMatch. Despite the limited size and limitations of GEDMatch, the service is still able to solve a number of cases (Murphy 2019a, 2019b). Triangulation performed against the matches from larger services such as Ancestry.com or 23andMe would have a higher probability of success.

One critical step for DNA genotyping is DNA extraction. Human DNA is comprised of 23 pairs of chromosomes, one from each parent, and is stored inside the nucleus and mitochondria of the cell. To properly extract and then genotype DNA, it must be separated from the rest of the

cellular components. DTC services collect saliva from individuals with pre-assembled kits and then extract the DNA from these kits using proprietary methods. A classic method of DNA extraction is a phenol:chloroform:isoamyl alcohol method that takes advantage of DNA's solubility properties in water and ethanol to separate it from the other cellular components. There are other methods that utilize the negative charge or binding properties of DNA to purify samples. From promotional videos, DTC providers likely use a magnetic bead purification method ("23andMe DNA Processing Lab Video" 2017). One property of all DNA extraction protocols is that they can be repeated, intentionally or unintentionally, with minimal damage to the DNA in a way that would impact genotyping.

Once extracted, the DNA is then processed for genotyping. DNA is amplified and labeled using PCR-based methods, and this DNA is then affixed to custom "MicroArray" chips in a process called hybridization. These chips are developed by the company Illumina, and can rapidly detect SNPs in over half a million regions of the genome. 23andMe uses the Illumina Infinium Global Screening Array, and the protocols, procedures and equipment needed for turning extracted DNA into a genotype is provided in detail on the Illumina website ("Infinium Global Screening Array-24 Support Resources," 2020).

In Aldhous et al. the authors attempted to spoof extracted DNA to deCODE genetics and 23andMe but were unable to get a genotype (Aldhous and Reilly 2009). They were able to successfully receive genotype data from deCODE using an unprocessed semen sample, but they were missing proper control samples for comparison. Our project builds on this work, presenting a method for successfully processing extracted DNA from human samples. Our analysis also uses

the raw data provided from each DTC service to verify identity between samples, confirming that successful processing was not due to foreign DNA or amplification error.

Threat model

Examples of actors who would be interested in executing saliva spoofing might be individuals impersonating family members for financial gain, individuals (such as reporters or paparazzi) performing paternity or health tests surreptitiously for a news story, private detectives looking to identify someone for their client or governments/institutions who wish to avail themselves of triangulation but do not have the legal authority to do so. See Section **Potential harms** for discussion.

We assume an attacker has access to biological material containing DNA from the target individual whom they wish to genotype and identify or impersonate. Example source material may include skin, blood, or saliva found on discarded gum, tissues, or bandages. In this paper we restrict our analysis to DNA extracted from saliva, for technical and ethical reasons.

The spoofer need not know the identity of the person they are trying to impersonate. For example, in 2011 a private detective identified the sender of a piece of mail by testing sneeze residue on the mail using traditional DNA forensics (Gardner 2011). Because the detective used traditional DNA typing, he had to test other individuals until he found a match. With spoofing, a detective could use a DTC service to enumerate the close relatives of biological source material, enabling identification through triangulation as discussed in Section **Background and Related Work**.

For our study, we assume that an attacker has access to supplies commonly found in microbiology labs. This includes pipettes, centrifuges, and reagents such as a phenol:cholorform solution or pre-made DNA extraction kits. These resources might be found in commercial labs, government research labs, and some makerspaces. Alternatively, an attacker may purchase pre-made extraction kits, or use an existing extraction and whole genome amplification services that serve scientific and forensic communities, which cost about $100 per sample ("Purified DNA Whole Genome Amplification Service," 2020). The amount of monetary resources, technological resources, and expert knowledge all directly impact the sophistication of the attack. We believe a state-level adversary could extract and amplify sufficient quantities of DNA from a diverse assortment of difficult samples, evading even the most advanced mitigations.

Materials and methods

Our goal is to create a "spoofed" sample using only pure, extracted DNA that processes successfully with DTC services. DNA can be extracted using various methods that are at a range of price points and equipment availability. The choice of extraction method is dependent on the source of the DNA as well as the downstream applications. For this work, we wanted to simplify this process as much as possible in order to highlight the accessibility of this attack to the general public. We chose to extract DNA from saliva and buccal cells, which could be found in discarded items such as gum, disposable cups, or utensils. This choice was also a practical one, because higher DNA concentrations come from fresh cells as opposed to dead cells found in hair or skin. We also used two different methods to extract this DNA — pre-made kits from scientific suppliers as well as a traditional method using phenol:chloroform. We found that these kits are cheap

(between $100 and $200), required minimal equipment and are user-friendly, however these kits resulted in too little DNA to be successfully processed.

Given this, we processed the rest of our samples using phenol:chloroform. The phenol:chloroform method requires more expertise and equipment but is a well-tested and reliable method used for DNA extraction of a large range of cell sources (Ghatak, Muthukumaran, and Nachimuthu 2013). We used equipment and reagents that are available in a standard biology lab; however anyone seeking to build a home lab can replicate this process tens of times with only ~$1000 investment.

Sample preparation

Two testers, both drawn from the authors, provided DNA samples for these experiments. We purchased two kits from both Ancestry.com and 23andMe, one for each person. We prepared a standard saliva sample in the first kit as a control and prepared the second "spoofed" kit using extracted DNA. We then mailed both kits back to the respective services for processing. This process is demonstrated in **Figure 38**.

Following the guidelines as the DTC kits recommend, the testers did not eat or drink for at least 30 minutes prior to collecting saliva or DNA samples. We prepared control samples using saliva and the standard instructions for each kit. Between the two services, the only difference was the volume of saliva required, with 23andMe requiring 2 mL instead of 1 mL for Ancestry.com. We used the data from these control samples to compare against the results of the extracted DNA samples.

For the spoofed kit, we developed a protocol to collect DNA from cells using buccal swabs. This process collects saliva and buccal cells from a tester's mouth and prepares them for DNA extraction using an adapted, traditional phenol:cholorform:isoamyl alcohol DNA extraction; see **Figure 38**a-c. We collected saliva and buccal cells by swabbing each side of the mouth for 15 seconds with Gum Proxabrush Go-Betweens Interdental Brushes, repeated five times. We suspended the sample accumulated on each brush into 500 uL of lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris (pH 8), 1 mM EDTA) using a twisting motion in a 1.5 mL

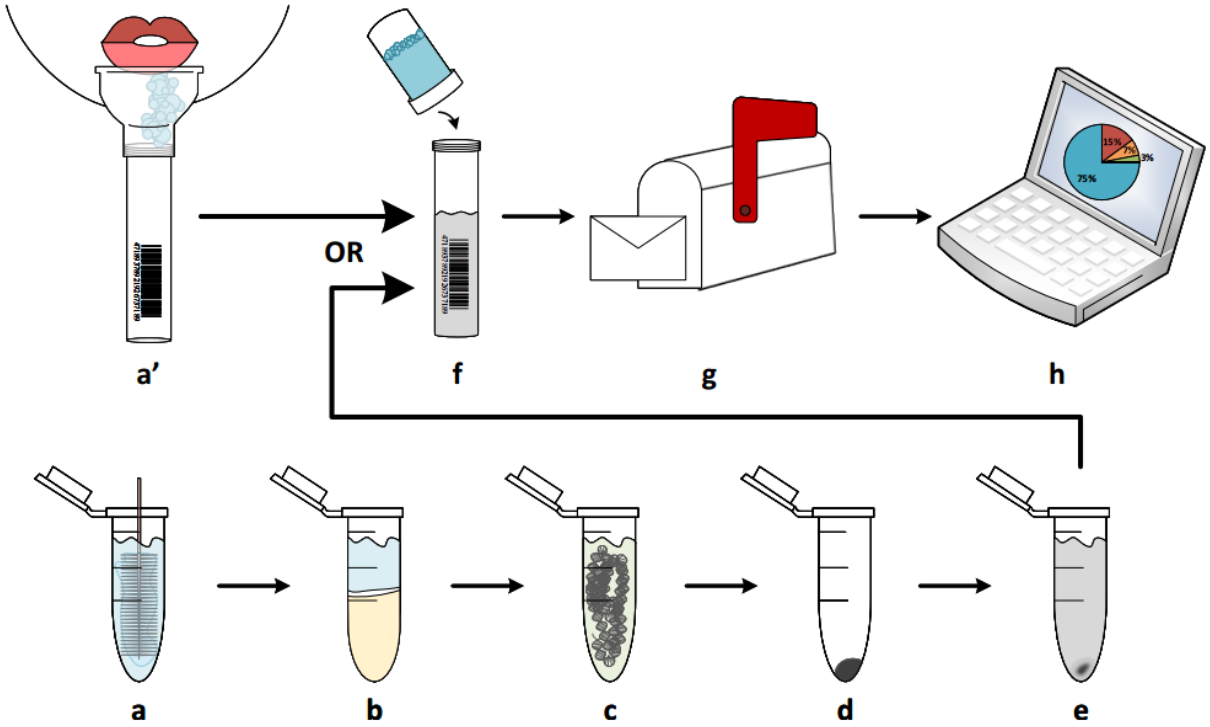


Figure 38 Typical and adapted workflow for DNA sample preparation .} The expected workflow starts at a', where a customer deposits their saliva in a kit purchased from a DTC provider. Our DNA extraction process replaces step a' with steps **(a-e)** as described: **(a)** saliva and buccal cells are suspended in lysis buffer to break the cell membranes and release the DNA **(b)** phenol:chloroform is added to this solution, and this extraction allows DNA to dissolve in the aqueous layer (top, blue), which allows separation from cell debris (middle, white) and organics (bottom, yellow) **(c)** the top layer (blue) containing solubilized DNA is removed and mixed with salt and ethanol to precipitate the DNA **(d)** DNA is collected by centrifugation into a solid pellet and the ethanol/water mixture is removed **(e)** clean DNA is dissolved in water to a total volume of 1 or 2 mL and used in place of **(a')** saliva sample **(f)** saliva or extracted DNA is mixed with stabilizing solution provided in kit and then **(g)** mailed to the DTC provider and **(h)** analyzed.

microcentrifuge tube. This buffer disrupts the lipid membrane of cells, enabling DNA release from inside the nucleus. Phenol:chloroform allows separation of the DNA from the hydrophobic lipid membrane and insoluble proteins and also further aids in cell lysis. The rest of this procedure followed Ghatak et al. without the RNAse incubation step and by substituting ethanol for isopropyl alcohol.

We then developed a method to remove the solid DNA and prepare it for our spoofed kit; see **Figure 38**d-f . We precipitated the DNA at -20 °C overnight. This reduces the solubility of the DNA, enabling us to remove more of the DNA from the liquid. We then pelleted the solid DNA using an Eppendorf 5425 centrifuge for 5 mins at 14,000 rcf, separating it from solution. We aspirated off the remaining 70% ethanol solution and dried the DNA pellet for 8 hrs in a laminar flow hood until the DNA becomes translucent. This drying process ensures no residual ethanol or water that could contaminate our spoofed sample remains. We suspended the DNA pellet in either 1 or 2 mL deionized water (the volume of saliva requested was 1 mL for Ancestry.com and 2 mL for 23andMe) and once dissolved, quantified the concentration of DNA using a ThermoFisher Nanodrop. If the concentration is below 15 ng/µl, we repeated the swabbing and extraction process and suspended the new DNA pellet in the previously extracted DNA solution. We transferred this DNA solution into the collection tube provided in each DTC kit and shipped the samples. All samples in this study had final concentrations between 15 and 25 ng/µl.

We also tested pre-made DNA extraction kits from two different suppliers, the Zymo Research Quick-DNA Miniprep Plus Kit and the Quiagen QIAamp DNA Mini Kit. Following the instructions provided in each kit, we treated gum and skin samples. Unfortunately, we were unable to collect enough DNA to quantify reliably using the Nanodrop. This could be from a range

of factors, including low initial DNA in original sample, high levels of contamination or another factor.

SNP and relative analysis

We set up Ancestry.com and 23andMe accounts using false names and birthdates, one for each kit, for a total of eight accounts across all services. Each tester used the same false identity across services. Services include duplicate detection as a part of their process; that is, they check for samples with different genders or different ages that have matching DNA. If two samples match when they should not, the service will withhold the genotype and match data. To avoid duplicate detection, we mailed kits at staggered times. We downloaded all the SNP and family tree data as soon as it became available, and then we requested deletion of the data.

For each tester and company, we compare the SNPs identified in the extracted DNA kits versus the control saliva kits. From this data, we can determine the total number of locations that are different between the two genotypes. Of these differences, some are due to missing data (no SNP was read or could be determined for that location), whereas some reveal actual differences in the SNPs.

In order to verify that genotype error does not significantly impact genetic matching results, we measure the difference in relationship matches between samples. To collect this data, 23andMe allows users to download a CSV of match percentage. To extract the information from Ancestry.com, we scraped the data from the match webpage using a custom python script.

Ancestry.com reports the number of centimorgans (cM), a unit of genetic distance, shared between an individual and their relative's genomes, while 23andMe reports a percentage of total

shared genome. Ancestry.com uses 3475 cM as the total genomic length for all individuals ("What Does the Match Confidence Score Mean?" 2020). We use this value to convert Ancestry.com cM to shared genomic percentages. For each relative, we compute the difference in SNP match percentage between the control and extracted DNA samples. We then summarize this difference over all relatives identified in each database (n between 1,300 to 90,000) and report those statistics in **Table 29**. The difference in relatives found on each service might be correlated with the matching thresholds. Ancestry.com has a lower threshold for relative matching, displaying all individuals sharing at least one contiguous series of SNPs of 6 cM (C. A. Ball, 2016), whereas 23andMe shares only those with a series 7 cM in length or larger ("DNA Relatives: Detecting Relatives and Predicting Relationships," 2020). Ancestry.com also has ~25% larger user base, which may also contribute to the difference in relative match count. In the case where a relative is identified in one sample and not the other, the difference percentage was given as 100%, therefore our statistics reported are likely an overestimation of difference.

Ethical considerations

We chose only to test extracted saliva as our source material as opposed to blood or skin, which carries different pathogens and risks that might not be addressed from a saliva processing method. This choice was made in part to protect employees at genetic genealogy companies and those transporting our samples to the facility. In theory, phenol:chloroform extraction destroys all viable human and bacterial cells as well as viral capsids. Once extracted, nonrecombinant DNA has minimal risk of infection or harm to humans or the environment and has a lower risk than regular spit, which is commonly classified as biosafety level (BSL) 1 or 2.

By keeping the number to just two authors, we reduce risks involved in managing a large pool of accounts with personal information, as well as minimizing the number of genotypes we store on disk. Our IRB determined that IRB oversight was unnecessary as the only subjects are the authors.

Both services require users to set their accounts to a “public” setting in order to view or download the list of relatives. While the services have no “real name” policy for display names, in order to minimize harms from a relative discovering unexpected matches, we informed siblings and parents in person before performing experiments. We acknowledge that enabling relative matches may have latent consequences for more distant family members, however collecting family match information enabled us to demonstrate that genotyping errors do not alter relative matching fidelity. Future experiments in this field can further reduce harms by maintaining accounts on “private”. On 23andMe we configured our account privacy settings to only display initials in the account public profile. Additionally, we minimized the time during which our accounts were public to the time necessary to scrape or download the list of relatives. We successfully addressed one relative inquiry on the platform while on the “public” setting.

Results

Table 27 Comparison of extracted samples against control samples. Ancestry.com genotype data includes a total of ~620,000 SNPs

Service	Name	# of different SNPs	% of SNPs different	# of genotyped SNPs different
23andMe	Tester 1	3687	0.58%	13
Ancestry.com		1663	0.25%	26
23andMe	Tester 2	9851	1.54%	144
Ancestry.com		3991	0.59%	139

Table 28 Fraction of SNPs missing from genotype.

Service	Name	Sample type	Percent missing
23andme	Tester 1	Saliva	2.67%
		Extracted	2.44%
	Tester 2	Saliva	4.09%
		Extracted	2.97%
Ancestry.com	Tester 1	Saliva	0.14%
		Extracted	0.12%
	Tester 2	Saliva	0.37%
		Extracted	0.72%

Table 29 Difference in genetic matches found by 23andMe and Ancestry.com. 23andMe reports matches as sharing some percent of genomic material. Ancestry.com measures matches by centimorgans. Ancestry.com uses 3475 cM as the total genomic length for all individuals. The data in this table presents the difference in the values reported for extracted versus control samples over n relatives.

Service	Name	Mean difference ± stdev (%)	n
23andme	Tester 1	0.02±0.1%	1358
	Tester 2	0.01±0.04%	1402
Ancestry.com	Tester 1	0.02±0.05%	95266
	Tester 2	0.11±0.15%	16510

Extracted DNA was successfully processed by 23andMe and Ancestry.com, and we were able to download raw SNP data for all samples, as well as genetic matches. Using the raw genotype data, we found between 0.25-1.54% difference in the control samples relative to the extracted samples (**Table 27)**. However, further analysis reveals that the vast majority of this difference can be accounted for by missing reads as opposed to actual SNP differences (**Table 28)**, with actual SNP differences representing less than 0.02% of all reads. This error rate aligns with the reported error rates of Illumina's Infinium Global Screening Array (<0.03%) ("Infinium Global Screening Array v3.0 Reproducibility and Heritability Report," 2020.), the chip used by 23andMe, and supports the conclusion that saliva and extracted DNA are processed identically.

Individuals identified in the family tree profiles of both the extracted and control DNA samples were verified to be accurate using personal knowledge of relatives, as well as by networking among family members using the "mutual friends" feature on Facebook. For all available relatives, the percentage of DNA shared was compared between the extracted and control samples. For both Ancestry.com and 23andMe, we found sub 1% differences between the two samples (**Table 29**). This demonstrates that the genotyping error seen in **Table 27** does not significantly contribute to differences in relative matching.

In the case of Ancestry.com, one extracted saliva sample was flagged for quality control reasons. Ancestry.com sent an email for the account associated with one of the spoofed kits. In the email they explained that they already had that DNA profile in their system (due to our earlier controls). Using their metrics, spoofed kit DNA appeared to match the SNP and family tree profile of the control kit. This is due to an error on our part, where we forgot to request deletion of the control kit data before the spoofed kit data had begun processing. Ancestry.com's response to this error demonstrates that by Ancestry.com's own metrics our control samples match our spoofed samples. While it is reassuring that Ancestry.com might have some elements of protection for those within their database, those who are missing from the database are still vulnerable to spoofing. For the purposes of data comparison, we were able to resolve the hold on the account with the extracted sample by sending an email from both accounts claiming they were siblings. For other samples we staggered sample arrival to avoid duplicate detection.

In earlier pilot experiments, the services were unable to process our results. In these cases, DNA was extracted using a phenol:chloroform method which resulted in less than 10 ng/µL. We presume low concentration was the cause for failure, but we were not provided with

the failure cause from the service. DNA that was extracted from chewing gum using phenol:choroform also failed to process, possibly due to low concentration of DNA (<8 ng/μL), contamination from the gum itself or both. Subsequent attempts to extract DNA from chewing gum using column-based DNA extraction (Zymo Research, Quiagen) may have eliminated potential gum contamination but resulted in too little DNA to quantify via nanodrop.

Discussion

Potential harms

Using sample spoofing, an attacker can impersonate anyone whose biological material they can acquire. While posing as the individual, the attacker can contact a matched relative through the service, and use the trust in the service to manipulate a victim's close relatives with financial or emotional motives. The FTC found that in 2019, imposter scams accounted for over 20% of all fraud reported in the US, the second most common form of fraud after identity theft ("Consumer Sentinel Network Data Book 2019," 2019.). The details of imposter scams vary with the medium used and the intended victim. For example, a scammer might pose as a family member in the hospital in need of money for critical medical care in a so-called "family emergency scam" (Schifferle 2020).

Others have raised concerns about the use of DTC services for illicit paternity testing, but the focus has been on fathers testing their children (Moray et al. 2017). Testing via DTC service is potentially more powerful than traditional paternity testing, as traditional paternity testing only determines if a pair of individuals is parent/child. DTC services reveal extended families, beyond simple parent/child status, and spoofing enables a broader range of unscrupulous

individuals to perform paternity testing. In 2002, The Guardian surfaced allegations of a plot to steal Prince Harry's hair in order to perform paternity testing ("Prince Harry 'honey Trap' Allegations" 2002). There was enough concern about non-consensual testing that in 2004 the UK passed the Tissues act, which forbids such testing ("Q&A: Human Tissue Act" 2006; Joh 2011).

The revelation that the Golden State Killer was found using genetic genealogy (Lillis 2018) brought with it concerns about law enforcement's use of triangulation to identify suspects (Berkman, Miller, and Grady 2018; Syndercombe Court 2018; Scudder, 2018). In one case, Ancestry.com matched crime scene DNA to one of their customers, Michael Usry. Follow up testing showed that Mr. Usry did not in fact match the crime scene DNA, but in the month it took for the test to exclude the suspect, news articles had already proliferated across search engines with the suspects name (Scudder, 2018). GEDMatch had claimed for a time that law enforcement searches were restricted to homicide, rape or kidnapping, but later, BuzzFeed News found that GEDMatch had been used by law enforcement for an assault case (Aldhous 2019), sparking controversy. GEDMatch adopted an opt-in policy to law enforcement searches (Aldhous, 2019), greatly diminishing the effective size of the database to law enforcement. The forensics behind triangulation are just as susceptible to mix-ups as the forensics involved in traditional forensics in a lab. The court systems in most developed countries have an appeals process that allows for erroneous evidence to be re-examined.

Spoofing allows an attacker to enumerate relatives in the same way that law enforcement enumerates relatives for triangulation. Spoofing would allow a broad range of individuals, governments, and corporations to use triangulation. The concerns about overreach, errors, and lack of transparency raised by law enforcement use of triangulation are amplified when anyone

can perform triangulation through spoofing. Entities involved in spoofing may not be interested or capable of implementing the checks and remedies that exist when law enforcement agencies use triangulation in the public eye.

Mitigations

We believe there are potential mitigation strategies that genetic genealogy companies could employ. The strategies fit into two categories. The first consists of authenticating the sample itself, by examining properties of the sample to test authenticity. The second strategy consists of authenticating the customer, by asking them to prove their identity. While the mitigation strategies we propose may increase the cost of an attack, they are not silver bullets. These trade-offs in usability, cost or sample processing speed may be necessary to develop a secure and reliable verification method.

Sample authentication

DTC providers can prevent misuse of their genealogy services by implementing additional sample analysis steps. Verification that the sample provided is indeed saliva could significantly reduce the success of extracted DNA attacks. Human saliva has a unique viscosity, turbidity, chemical, and enzymatic composition that could be used to verify a sample. Simple tests with a viscometer or microscope could be used to analyze the composition of the sample. Further tests that determine enzyme or chemical composition of the sample using western blots or HPLC could also aid in authentication of the saliva.

It would be possible for an advanced attacker to bypass nearly all of these checks. For example, saliva proteins and enzymes could be synthesized and then added to an extracted DNA

sample. Small molecules can be purchased or synthesized and added in appropriate quantities to fool chemical composition analysis. However, a robust method that checks the sample for live cells could impede an advanced attack, as it would prevent DNA amplification from low-concentration samples. DNA amplification requires DNA extraction, a process that destroys live cells. It would be difficult to introduce “artificial” cells into a sample that would look indistinguishable from buccal cells. Using foreign cells to pass cell checks would likely contaminate a sample with foreign DNA, causing significant error in genotyping. Though effective, sample analysis of this magnitude would require significant equipment investment, time, and specialization that may be cost-prohibitive.

Customer authentication

Ideally, services would authenticate customers in two ways. First by using identity verification, that is, confirming that the user is the person who they say they are. Second by confirming that the provided sample came from that person. Verifying the identity of the customer may be difficult but is standard practice in some industries. On the other hand, confirming that the sample came from the customer is presently impossible.

While identity verification does not directly prevent spoofing, though it may act as a deterrent, it can be useful as a reactive measure that enables law enforcement action when an attack is detected. User authentication would also enable DTC services to implement UI elements similar to Twitter's “blue checkmark”, which indicates the person is using the name provided on their identification. This would help users spot and avoid impersonation scams.

Identity verification is common for compliance with “know your customer” policies or laws, similar to those seen in the financial sector. This could be implemented through stringent identification requirements of government IDs, background checks, utilities statements or social security numbers. Less onerous mitigations might follow rules established by companies such as AirBnB, which require customers to send in a photo of their driver’s license or passport along with a current photo of themselves in order to verify their identity.

Requesting identification is not without drawbacks. User verification imposes additional costs on companies and added complexity for the user. Some users might also be uncomfortable providing even more personal information, such as address and photos, to services.

Alternatively, a “blacklist” approach might be possible. Genetic genealogy services could offer individuals who wish to keep their information private the ability to register their genotypes. Services could then flag incoming samples that match already registered genotypes and halt processing of flagged samples. Such a system would likely rely on a trusted third party to manage the registered genotypes so that they can be shared between multiple services. This has the drawback of centralizing very sensitive information with a single party who needs to be trusted by both the genetic genealogy services and customers. Individuals who do not wish to give their information to genetic genealogy services will likely be just as reluctant to give their information to a third party. A blacklist approach would be hard to scale, and would miss individuals who do not register but may benefit from registration.

Limitations and future work

While we focused our experiments on genetic genealogy services, many sequencing-only services have similarly lax authentication policies. Using such services, it is possible to obtain genotypes from samples for ~$400 which can then be used on websites such as GEDMatch. Sequencing services such as Genewiz also provide full genome sequencing for ~$1300 per sample. While some larger companies may adopt authentication procedures, without strict worldwide regulation, it is not realistic to expect all genotyping services to properly authenticate their customers or the samples.

In previous experiments, we attempted DNA extraction from gum and skin using extraction kits from Zymo Research and Qiagen. These kits are typically used for extracting DNA from blood, skin, hair, and other samples ("See User-Developed Protocols," 2020; "Quick-DNA™ Miniprep Plus Kit," 2020). While other researchers have had success in extracting DNA from gum samples (Eychner, Schott, and Elkins 2017), we found that the low amount of DNA available in gum as well as the potential contamination from the gum itself prevented yields sufficient enough for quantification. However, with more advanced methods, small amounts of DNA can be extracted from disposed items such as gum. Small amounts of DNA can be amplified in a method called whole genome amplification. Increasing the amount of DNA involves a method called polymerase chain reaction (PCR), where a DNA polymerase enzyme can be used to copy input DNA material and improve concentration by a million-fold. Once copied to generate a sufficient amount of DNA, this DNA could be sent for genotyping. Therefore, even low-concentration environmental sources of DNA samples such as hair, blood or skin may become

viable for spoofing with additional molecular processing steps. Previous research shows that DNA transfer in the nanogram range is common (Meakin and Jamieson 2013). For this reason, permitting access to genetic genealogy services with DNA as the sole authenticator puts individuals at risk.

Our technique explores only a small portion of the potential sample spoofing attack space. For example, it may be possible to spoof samples with altered or synthetic DNA. Using services such as Integrated DNA Technologies (IDT) or another custom nucleic acid provider, regions of interest — for example SNPs that provide markers for cystic fibrosis, located within the cystic fibrosis transmembrane conductance regulator gene — can be synthesized. Given genetic modification tools like CRISPR-Cas9, it becomes feasible to use this synthesized DNA to replace regions of the original DNA, forming an altered genome that appears to have cystic fibrosis. These synthetic samples would make it possible to execute the genotype inference attack introduced by (Ney, Ceze, and Kohno 2020; Edge and Coop 2020) and allow probing of the database and exposure of individuals that also contain cystic fibrosis markers. Synthetic attacks of this nature are not limited to individuals within the database, as triangulation-based inference techniques from close relatives can also apply.

Conclusion

We have demonstrated that extracted DNA can be effectively used as a substitute for saliva in direct to consumer genetic testing kits, leaving individuals vulnerable. In order to prevent this, we provide some solutions that consumer genetic genealogy companies should implement to authenticate both the customers and the samples to prevent identity-based attacks and fraud.

References

"23andMe DNA Processing Lab Video." 2017. https://www.youtube.com/watch?v=Enn3BHE3v4s (visited on 05/31/2020).

"About Us - 23andMe." 2020. https://mediacenter.23andme.com/company/about-us/.

Aldhous, Peter. 2019 "This Genealogy Database Helped Solve Dozens Of Crimes. But Its New Privacy Rules Will Restrict Access By Cops." https://www.buzzfeednews.com/article/%0Apeteraldhous/this- genealogy- database-helped- solvedozens-of-crimes-but.

Aldhous, Peter. 2019. "The Arrest Of A Teen On An Assault Charge Has Sparked New Privacy Fears About DNA Sleuthing." BuzzFeed News. 2019. https:%0A//www.buzzfeednews.com/article/peteraldhous/geneticgenealogy-parabon-gedmatch-assault.%0A.

Aldhous, Peter, and Michael Reilly. 2009. "How My Genome Was Hacked." New Scientist. https://doi.org/10.1016/s0262-4079(09)60829-0.

"Ancestry Company Facts." 2020. https://www.ancestry.com/corporate/about-ancestry/company-facts.

Ball, Catherine A. "Ancestry DNA Matching White Paper. Discovering Genetic Matches across a Massive, Expanding Genetic Database." Tech. Rep. 2016. https://www.ancestry.com/dna/resource/whitePaper/%0AAncestryDNA-Matching-White-Paper.pdf.

Berkman, Benjamin E., Wynter K. Miller, and Christine Grady. 2018. “Is It Ethical to Use Genealogy Data to Solve Crimes?” Annals of Internal Medicine. https://doi.org/10.7326/M18-1348.

Carpenter, Guy H. 2013. “The Secretion, Components, and Properties of Saliva.” Annual Review of Food Science and Technology. https://doi.org/10.1146/annurev-food-030212-182700.

“Consumer Sentinel Network Data Book 2019.” 2019. https://www.ftc.gov/system/files/documents/%0Areports/consumer-sentinel-network-data-book-2019/%0Aconsumer sentinel network data book 2019.pdf.

“DNA Relatives: Detecting Relatives and Predicting Relationships.” 2020.

Edge, Michael D., and Graham Coop. 2020. “Attacks on Genetic Privacy via Uploads to Genealogical Databases.” ELife. https://doi.org/10.7554/eLife.51810.

Erlich, Yaniv, Tal Shor, Itsik Pe’er, and Shai Carmi. 2018. “Identity Inference of Genomic Data Using Long-Range Familial Searches.” Science. https://doi.org/10.1126/science.aau4832.

Eychner, Alison M., Kelly M. Schott, and Kelly M. Elkins. 2017. “Assessing DNA Recovery from Chewing Gum.” Medicine, Science and the Law. https://doi.org/10.1177/0025802416676413.

Gardner, Eriq. 2011. “Gene Swipe: Few DNA Labs Know Whether Chromosomes Are Yours or If You Stole Them.” 2011. https://www.abajournal.com/magazine/%0Aarticle / gene swipe few dna labs know whether%0Achromosomes are yours or if you stole.

"GedMatch." 2020. https://www.gedmatch.com/curt msg.htm

Ghatak, Souvik, Rajendra Bose Muthukumaran, and Senthil Kumar Nachimuthu. 2013. "A Simple Method of Genomic DNA Extraction from Human Samples for PCR-RFLP Analysis." Journal of Biomolecular Techniques. https://doi.org/10.7171/jbt.13-2404-001.

Greytak, Ellen M., Ce Ce Moore, and Steven L. Armentrout. 2019. "Genetic Genealogy for Cold Case and Active Investigations." Forensic Science International. https://doi.org/10.1016/j.forsciint.2019.03.039.

Gusev, Alexander, Jennifer K. Lowe, Markus Stoffel, Mark J. Daly, David Altshuler, Jan L. Breslow, Jeffrey M. Friedman, and Itsik Pe'Er. 2009. "Whole Population, Genome-Wide Mapping of Hidden Relatedness." Genome Research. https://doi.org/10.1101/gr.081398.108.

"Infinium Global Screening Array-24 Support Resources." 2020.

"Infinium Global Screening Array v3.0 Reproducibility and Heritability Report." n.d.

Joh, Elizabeth E. 2011. "DNA Theft: Recognizing the Crime of Nonconsensual Genetic Collection and Testing." Boston University Law Review.

Lillis, Ryan. 2018. "Open-Source' Genealogy Site Provided Missing DNA Link to East Area Rapist, Investigator Says." 2018.

Meakin, Georgina, and Allan Jamieson. 2013. “DNA Transfer: Review and Implications for Casework.” Forensic Science International: Genetics. https://doi.org/10.1016/j.fsigen.2013.03.013.

Moray, Nathalie, Katherina E. Pink, Pascal Borry, and Maarten H.D. Larmuseau. 2017. “Paternity Testing under the Cloak of Recreational Genetics.” European Journal of Human Genetics. https://doi.org/10.1038/ejhg.2017.31.

Murphy, Heather. 2019a. “Genealogy Sites Have Helped Identify Suspects. Now They’ve Helped Convict One. A New Forensic Technique Sailed through Its First Test in Court, Leading to a Guilty Verdict. But beyond the Courtroom, a Battle over Privacy Is Intensifying.” NY Times, 2019. https://www.nytimes.com/2019/07/01/us/dnagenetic-genealogy-trial.html.

Murphy, Heather. 2019b. “How Volunteer Sleuths Identified a Hiker and Her Killer After 36 Years. What Does It Actually Take to Identify a Person through Genetic Genealogy? Wading through Infidelities and Pornography.,” 2019. https://www.nytimes.com/2019/%0A05/11/us/cold-case-genealogy-dna.html.

Ney, Peter, Luis Ceze, and Tadayoshi Kohno. 2020. “Genotype Extraction and False Relative Attacks: Security Risks to Third-Party Genetic Genealogy Services Beyond Identity Inference.” In . https://doi.org/10.14722/ndss.2020.23049.

“Prince Harry ’honey Trap’ Allegations.” 2002. In: BBC (Dec. 2002). URL: http://news.bbc.co.uk/2/hi/uk news/2577539.stm.

"Purified DNA Whole Genome Amplification Service." https://www.expedeon.com/purified-dna-wholegenome-amplification-service/ .

"Q&A: Human Tissue Act." 2006. 2006. http://news.bbc.co.uk/2/hi/health/4944018.stm.

"Quick-DNATM Miniprep Plus Kit." https://files.zymoresearch.com/protocols/_d4068_d4069 quick-dna_miniprep_plus kit.pdf (visited on 05/31/2020).

Saunders, A. M., W. J. Strittmatter, D. Schmechel, P. H. St. George-Hyslop, M. A. Pericak-Vance, S. H. Joo, B. L. Rosi, et al. 1993. "Association of Apolipoprotein E Allele E4 with Late-Onset Familial and Sporadic Alzheimer's Disease." Neurology. https://doi.org/10.1212/wnl.43.8.1467.

Schifferle, Lisa Weintraub. 2020. "Grandparent Scams in the Age of Coronavirus." 2020.

Scudder, Nathan. n.d. "Crowdsourced and Crowdfunded: The Future of Forensic DNA?" Australian Journal of Forensic Sciences 2018 (July): 235–41. https://doi.org/10.1080/00450618.2018.1486456.

"See User-Developed Protocols." https://www.qiagen.com/us/products/top-sellers/qiaamp-dna-minikit/#resources (visited on 05/31/2020).

Syndercombe Court, Denise. 2018. "Forensic Genealogy: Some Serious Concerns." Forensic Science International: Genetics. https://doi.org/10.1016/j.fsigen.2018.07.011.

"What Does the Match Confidence Score Mean?" 2020. 2020. https://www.ancestry.com/cs/dna- help/%0Amatches/confidence.